EXAMTOPICS

-Expert Verified, Online, **Free**.



CERTIFICATION TEST

-CertificationTest.net - Cheap & Quality Resources With Best Support

Question #1 *Topic 1*

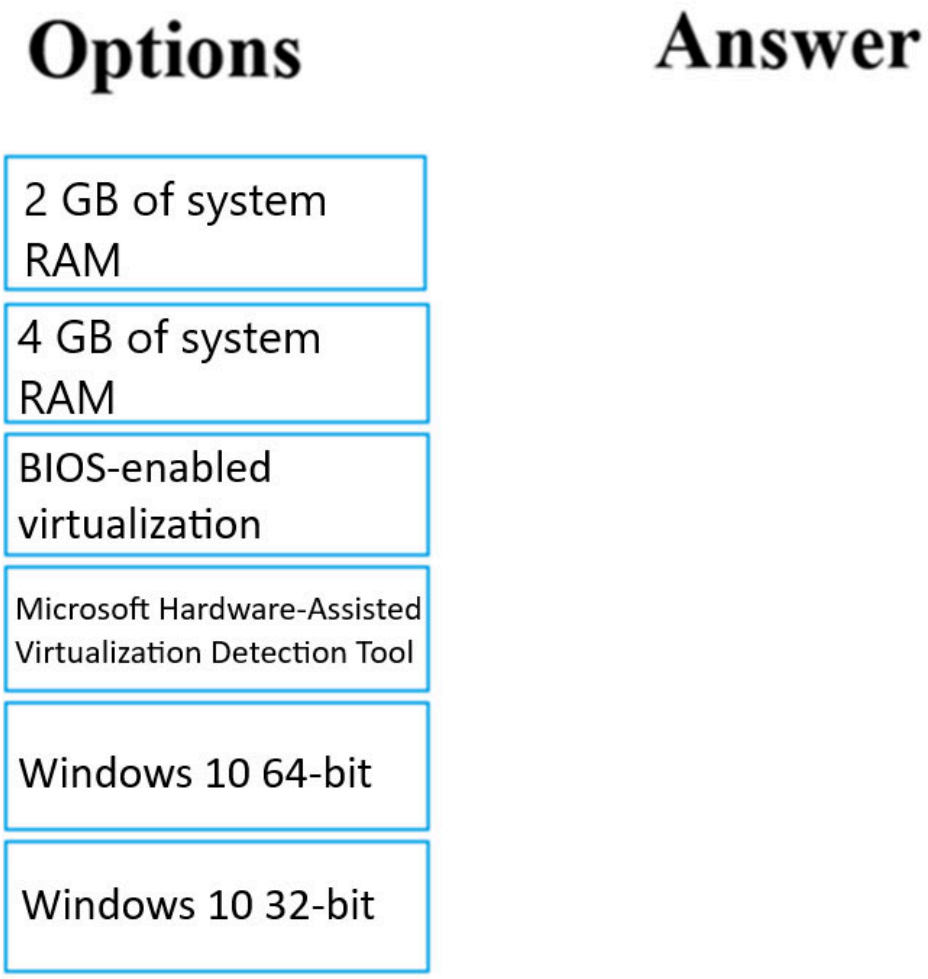
DRAG DROP -

You are planning to host practical training to acquaint staff with Docker for Windows.

Staff devices must support the installation of Docker.

Which of the following are requirements for this installation? Answer by dragging the correct options from the list to the answer area.

Select and Place:

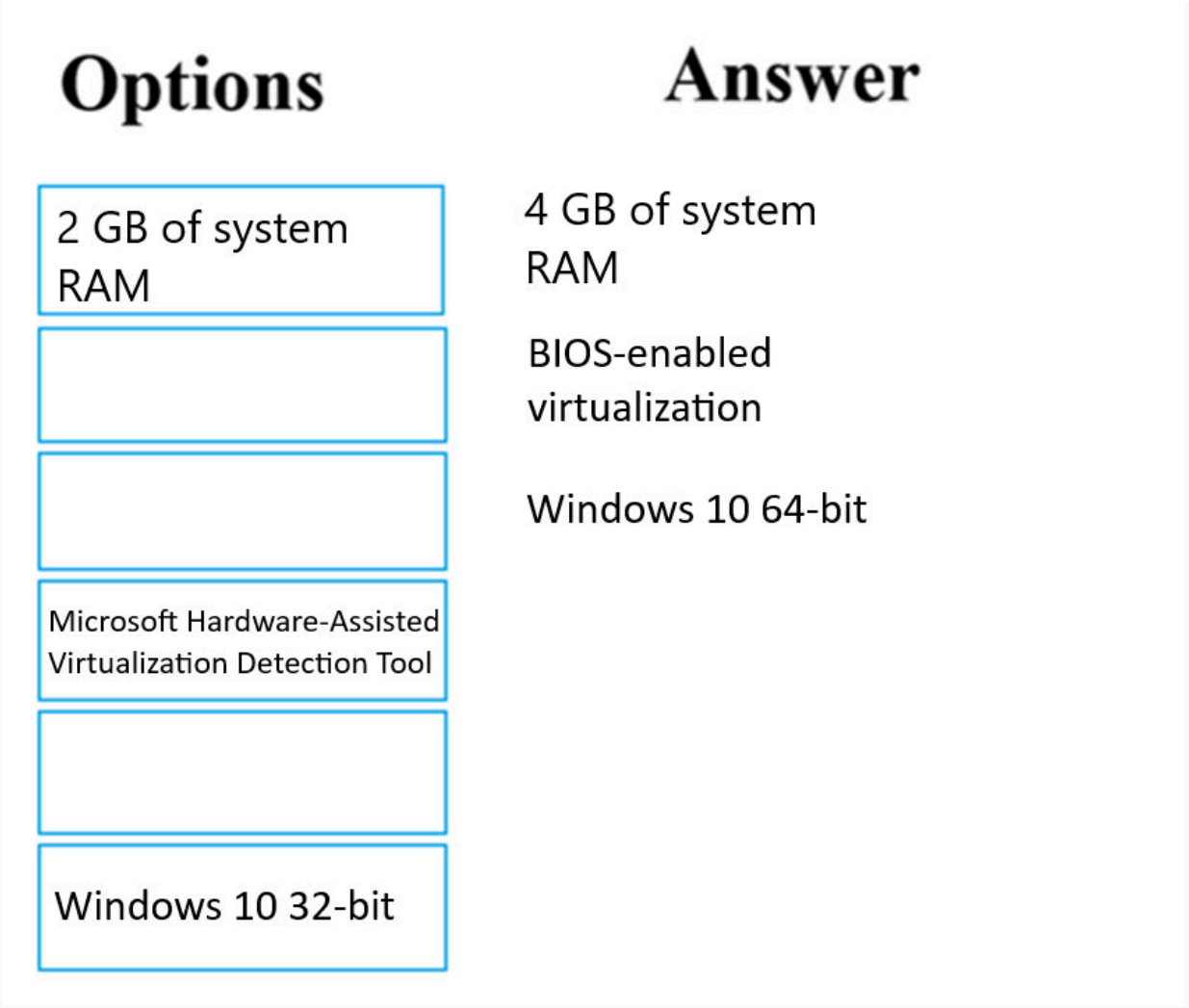


Suggested Answer:

Options	Answer
2 GB of system RAM	4 GB of system RAM
	BIOS-enabled virtualization
	Windows 10 64-bit
Microsoft Hardware-Assisted Virtualization Detection Tool	
Windows 10 32-bit	

WSL 2 backend

- Windows 10 64-bit: Home or Pro 2004 (build 19041) or higher, or Enterprise or Education 1909 (build 18363) or higher.
- Enable the WSL 2 feature on Windows. For detailed instructions, refer to the Microsoft documentation.
- The following hardware prerequisites are required to successfully run WSL 2 on Windows 10:
  - 64-bit processor with Second Level Address Translation (SLAT)
  - 4GB system RAM
  - BIOS-level hardware virtualization support must be enabled in the BIOS settings. For more information, see Virtualization.
- Download and install the Linux kernel update package.



Reference: https://docs.docker.com/toolbox/toolbox_install_windows/ https://blogs.technet.microsoft.com/canitpro/2015/09/08/step-by-step-enabling-hyper-v-for-use-on-windows-10/ https://docs.docker.com/docker-for-windows/install/

emmanuelodenyire Highly Voted 2 years, 6 months ago

Here are the requirements for installing Docker on Windows: Windows 10 Pro, Enterprise, or Education (64-bit) Hyper-V and Containers Windows features must be enabled. The CPU must support Hardware Virtualization. At least 4GB of RAM. Windows 10 Anniversary Update (Build 1607+) or Windows Server 2016. A 64-bit processor. These requirements should be checked before hosting the practical training to ensure that all staff devices meet the necessary specifications. So the answer is correct

upvoted 16 times

james2033 Most Recent 1 year, 10 months ago

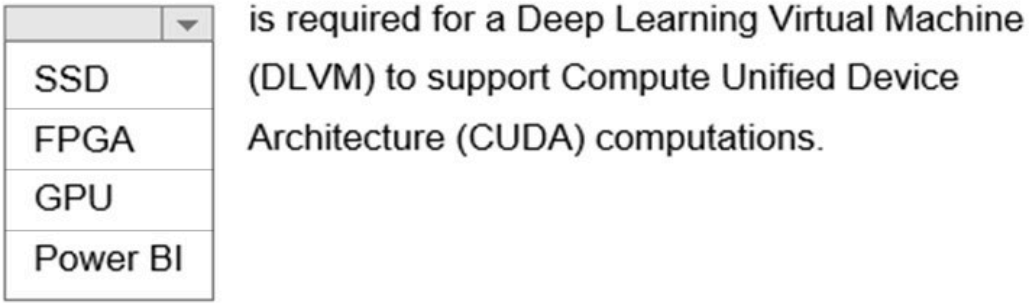
1) 4 GB of system RAM 2) BIOS-enabled virtualization 3) Windows 10 64-bit

upvoted 3 times

Question #2 *Topic 1*

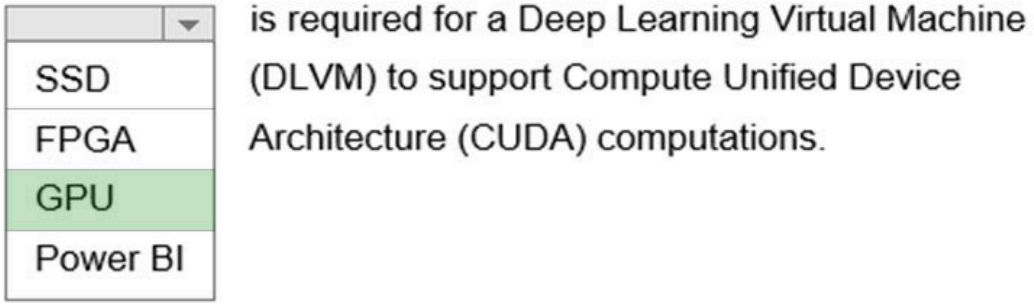
HOTSPOT -
Complete the sentence by selecting the correct option in the answer area.
Hot Area:

Answer Area

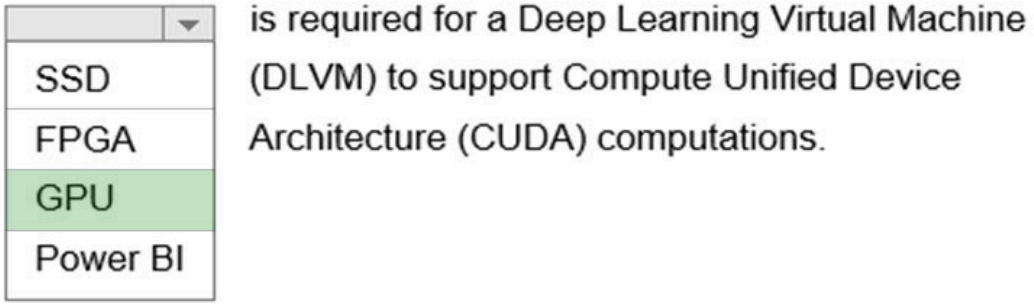


Suggested Answer:

Answer Area



Answer Area



A Deep Learning Virtual Machine is a pre-configured environment for deep learning using GPU instances.

james2033 Most Recent 1 year, 10 months ago
GPU is required for a Deep Learning Virtual Machine (DLVM) to support Compute Unified Device Architecture (CUDA) computations.
upvoted 2 times

PopeyeDS 2 years, 1 month ago
By using DLVM, developers and researchers can focus on their deep learning tasks without spending time on manual software setup, ensuring a streamlined and optimized environment for deep learning model training and inference using CUDA-based acceleration. The DLVM image is typically used on cloud platforms or local machines equipped with NVIDIA GPUs.
upvoted 1 times

Yoshizn 2 years, 6 months ago

Correct. CUDA (Compute Unified Device Architecture) is a parallel computing platform and application programming interface model created by Nvidia, crucial for supporting the various deep learning applications.

upvoted 1 times

emmanuelodenyire 2 years, 6 months ago

A Deep Learning Virtual Machine is a pre-configured environment for deep learning that is optimized for use with GPU instances, providing the necessary hardware resources for training and running large deep learning models. These virtual machines typically come pre-installed with popular deep learning frameworks and tools, making it easy to get started with deep learning projects. So for this question, the answer is correct

upvoted 1 times

Question #3 *Topic 1*

You need to implement a Data Science Virtual Machine (DSVM) that supports the Caffe2 deep learning framework.
Which of the following DSVM should you create?

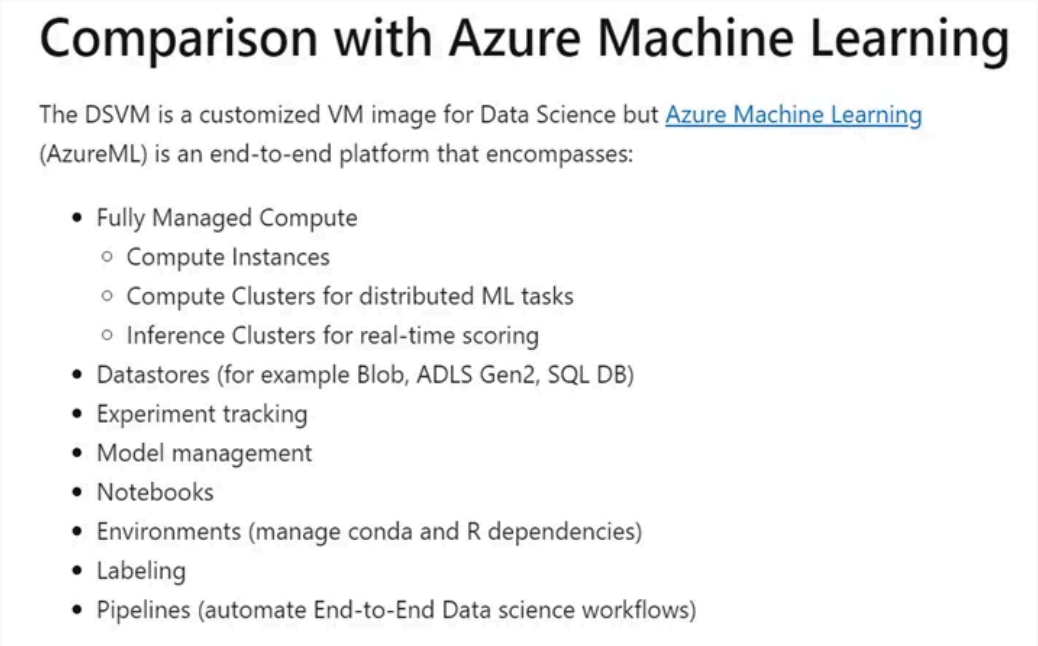
A. Windows Server 2012 DSVM

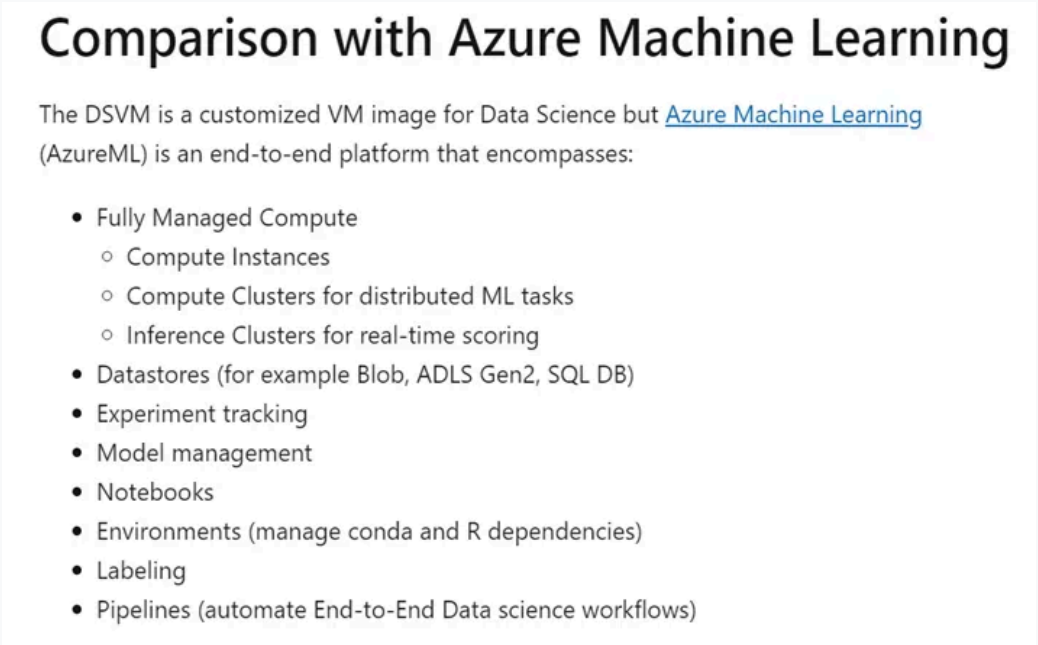
B. Windows Server 2016 DSVM

C. Ubuntu 16.04 DSVM

D. CentOS 7.4 DSVM

Suggested Answer: *C*





Caffe2 is supported by Data Science Virtual Machine for Linux. Microsoft offers Linux editions of the DSVM on Ubuntu 16.04 LTS and CentOS 7.4. However, only the DSVM on Ubuntu is preconfigured for Caffe2. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview

Community vote distribution

C (100%)

evangelist Most Recent 2 years ago

Data Science Deep Learning has to be carried out on linux-based OS

upvoted 2 times

james2033 2 years, 4 months ago

You need to implement a Data Science Virtual Machine (DSVM) that support Caffe2 deep learning framework. You should create DSVM "Ubuntu 16.04 DSVM" (Answer: C).

upvoted 1 times

Ratz 2 years, 6 months ago

Recent Update: The DSVM is available on: Windows Server 2019 Windows Server 2022 (Preview) Ubuntu 20.04 LTS

upvoted 1 times

PopeyeDS 2 years, 7 months ago

Ubuntu 16.04 is generally preferred for its strong support in the deep learning community and its compatibility with Caffe2.

upvoted 1 times

emmanuelodenyire 3 years ago

The correct option would be C. Ubuntu 16.04 DSVM. The Caffe2 deep learning framework is supported on Ubuntu 16.04, so creating a Data Science Virtual Machine based on that operating system would provide the necessary environment for running the framework. Other options may also work, but this one is specifically mentioned as being supported by the Caffe2 framework.

upvoted 1 times

Question #4 *Topic 1*

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You have been tasked with employing a machine learning model, which makes use of a PostgreSQL database and needs GPU processing, to forecast prices.
You are preparing to create a virtual machine that has the necessary tools built into it.
You need to make use of the correct virtual machine type.
Recommendation: You make use of a Geo AI Data Science Virtual Machine (Geo-DSVM) Windows edition.
Will the requirements be satisfied?

A. Yes

B. No

Suggested Answer: *B*
The Azure Geo AI Data Science VM (Geo-DSVM) delivers geospatial analytics capabilities from Microsoft's Data Science VM. Specifically, this VM extends the AI and data science toolkits in the Data Science VM by adding ESRI's market-leading ArcGIS Pro Geographic Information System. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview

Community vote distribution
B (100%)

nick234987 Highly Voted 4 years, 4 months ago
Because a Data Science Virtual Machine (DSVM) Windows edition also has GPU and you do not need Geo capabilities
upvoted 10 times

evangelist Most Recent 2 years ago
Geo AI Data Science Virtual Machine (Geo-DSVM) has nothing to do with the answer. WRONG
upvoted 1 times

Matt2000 2 years, 3 months ago
Is Azure Geo AI Data Science VM (Geo-DSVM) currently available on Azure? I only find old references. That is usually an indicator that a service has been deprecated.
upvoted 2 times

Vikyyy 2 years, 3 months ago
Geo capabilites why would that be used in price prediction unless we know its a real estate prices !!
upvoted 1 times

james2033 2 years, 4 months ago
You have been tasked with employing a machine learning model, which make use of + a PostgreSQL database + GPU processing to forecast prices. You are preparing to create a virtual machine that has the NECESSARY TOOLS built into it. --> Should not use Geo AI Data Science VM (Geo-DSVM) Windows edition. --> Choose answer B - No.
upvoted 1 times

PopeyeDS 2 years, 7 months ago
A more appropriate recommendation would be to use a Data Science Virtual Machine (DSVM) with GPU support. This would ensure that the virtual machine has the necessary tools for machine learning, including PostgreSQL for database management, and also provides GPU capabilities for accelerated processing required by the model.
upvoted 1 times

orionduo 3 years ago
No is right
upvoted 1 times

emmanuelodenyire 3 years ago
B. No. While a Geo AI Data Science Virtual Machine (Geo-DSVM) Windows edition might be a good choice for geographic data processing and analysis, it does not necessarily meet the requirement for GPU processing for the machine learning model and a PostgreSQL database. A different virtual machine type with GPU support and PostgreSQL capabilities might be a better choice for this specific use case. It's recommended to check the specific virtual machine offerings and their specifications to ensure that the necessary hardware and software resources are available for the task at hand.

upvoted 2 times

Question #5 *Topic 1*

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You have been tasked with employing a machine learning model, which makes use of a PostgreSQL database and needs GPU processing, to forecast prices.
You are preparing to create a virtual machine that has the necessary tools built into it.
You need to make use of the correct virtual machine type.
Recommendation: You make use of a Deep Learning Virtual Machine (DLVM) Windows edition.
Will the requirements be satisfied?

A. Yes

B. No

Suggested Answer: *B*
DLVM is a template on top of DSVM image. In terms of the packages, GPU drivers etc are all there in the DSVM image. Mostly it is for convenience during creation where we only allow DLVM to be created on GPU VM instances on Azure. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview

Community vote distribution
A (53%) B (47%)

Shyam_kishor Highly Voted 3 years, 10 months ago
Answer is correct
upvoted 7 times

Guru_1 Most Recent 8 months, 1 week ago
* Linux DLVMs/DSVMs typically come pre-configured with PostgreSQL. * Windows DLVMs/DSVMs generally do not have PostgreSQL pre-installed, though PostgreSQL can be installed on Windows systems manually
upvoted 1 times

KeiNek 1 year, 6 months ago
B PostgreSQL : Ubuntu https://learn.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/tools-included?view=azureml-api-2
upvoted 3 times

ManuelHenriques 1 year, 6 months ago
No. Need Ubuntu for PostgreSQL
upvoted 1 times

sanctafrax 1 year, 6 months ago
DLVM is discontinued. This question is therefore outdated. If you'd go back in time when it was still relevant. answer would still be NO. This because the question states "built into it". DLVM does have GPU support, no problem. it does NOT have PostgreSQL natively installed. Not even in the current version of DSVM (windows). ONLY the ubuntu version has it preinstalled. https://learn.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/tools-included?view=azureml-api-2 Therefore PostgreSQL questions and DLVM/DSVM answers IF "built into it" is asked, are by definition. no. You can however install PostgreSQL anyways, it will work. BUT that is not the question.
upvoted 2 times

emmanuelodenyire 1 year, 10 months ago
I will also stand with a Yes. A Deep Learning Virtual Machine (DLVM) is a pre-configured virtual machine image on Microsoft Azure that is optimized for training deep learning models and includes popular tools such as TensorFlow, PyTorch, and Caffe2. The DLVM Windows edition includes support for GPU processing, making it suitable for the task of running a machine learning model that requires GPU processing to forecast prices. Additionally, the DLVM can be configured to use PostgreSQL as the database, satisfying the requirement for a PostgreSQL database. Therefore, the recommendation to use a DLVM Windows edition will satisfy the requirements.
upvoted 3 times

Truman 1 year, 10 months ago
. The Deep Learning Virtual Machine (DLVM) Windows edition comes with NVIDIA GPU drivers pre-installed, which means that it can support GPU processing out of the box. Additionally, the DLVM comes with a variety of pre-installed tools and libraries that can be useful for machine learning workloads, including support for popular deep learning frameworks such as TensorFlow and PyTorch. Therefore, the DLVM Windows edition can be a good choice for a machine learning workload that requires both a PostgreSQL database and GPU processing.

upvoted 2 times

Sa_Msa 1 year, 10 months ago

The DLVM is a specially configured variant of the Data Science Virtual Machine (DSVM) that comes with pre-configured environments for developing and deploying deep learning models on GPU instances. Regarding PostgreSQL, the DLVM (Deep Learning Virtual Machine) is a specially configured variant of the DSVM (Data Science Virtual Machine) that is custom made to help users jump-start deep learning on Azure GPU VMs2. The DLVM uses the same underlying VM images of the DSVM and hence comes with the same set of data science tools and deep learning frameworks as the base VM. So, you can have access to PostgreSQL .

upvoted 1 times

prabhjot 2 years, 6 months ago

NO is Ans - as per chat GPT- No, the recommendation to use a Deep Learning Virtual Machine (DLVM) Windows edition may not fully satisfy the requirements mentioned. PostgreSQL is typically used on Linux-based systems, and GPU processing for machine learning is often better supported on Linux. The DLVM Windows edition may not be the most suitable choice for this scenario. Instead, you might want to consider a virtual machine with a Linux operating system that supports GPU acceleration and provides compatibility with PostgreSQL for your machine learning model.

upvoted 1 times

Vikyyy 2 years, 9 months ago

The Windows editions of the DSVM come preinstalled with GPU drivers, frameworks, and GPU versions of deep learning frameworks. On the Linux editions, deep learning on GPUs is enabled on the Ubuntu DSVMs. You can also deploy the Ubuntu or Windows editions of the DSVM to an Azure virtual machine that isn't based on GPUs. In this case, all the deep learning frameworks falls back to the CPU mode.

upvoted 1 times

james2033 2 years, 10 months ago

Question is obsoleted https://learn.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview?view=azureml-api-2#whats-included-in-the-data-science-vm DSVM https://azure.microsoft.com/en-us/products/virtual-machines/data-science-virtual-machines

upvoted 1 times

eternaleclipse 2 years, 10 months ago

This wasn't in the exam study material

upvoted 2 times

PradhanManva 2 years, 10 months ago

This is the answer.

upvoted 1 times

chen7777 2 years, 11 months ago

B. No. Not all configurations of the Deep Learning Virtual Machine (DLVM) come with GPU hardware as a default or built-in feature. DLVMs are designed to cater to a wide range of machine learning and deep learning workloads, and some configurations may be GPU-enabled, while others may not. If you require GPU acceleration for tasks like deep learning models or other GPU-intensive computations, you would need to choose a specific DLVM configuration that includes GPU resources or add GPU hardware to the virtual machine separately. This ensures that your virtual machine has the necessary GPU processing power to handle such workloads efficiently.

upvoted 1 times

phydev 3 years, 1 month ago

Yes, because DLVM is a modified DSVM.

upvoted 2 times

phdykd 3 years, 5 months ago

B. No. The recommendation to use a Deep Learning Virtual Machine (DLVM) Windows edition is not suitable for the given requirements because it is designed for Windows-based deep learning tasks and does not support GPU processing for PostgreSQL databases. Instead, a suitable virtual machine for this task would be one that supports GPU processing and has the necessary tools and libraries for machine learning and PostgreSQL database management. One such option is the Data Science Virtual Machine (DSVM), which is available for both Windows and Linux operating systems and has support for GPU processing and machine learning tools, including TensorFlow, PyTorch, and scikit-learn.

upvoted 2 times

mamau 3 years, 6 months ago

No is the answer ,it would be DSVM

upvoted 1 times

Question #6 *Topic 1*

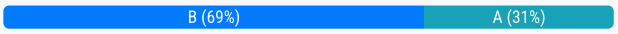
This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You have been tasked with employing a machine learning model, which makes use of a PostgreSQL database and needs GPU processing, to forecast prices.
You are preparing to create a virtual machine that has the necessary tools built into it.
You need to make use of the correct virtual machine type.
Recommendation: You make use of a Data Science Virtual Machine (DSVM) Windows edition.
Will the requirements be satisfied?

A. Yes

B. No

Suggested Answer: *A*
In the DSVM, your training models can use deep learning algorithms on hardware that's based on graphics processing units (GPUs). PostgreSQL is available for the following operating systems: Linux (all recent distributions), 64-bit installers available for macOS (OS X) version 10.6 and newer ع€" Windows (with installers available for 64-bit version; tested on latest versions and back to Windows 2012 R2. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview

Community vote distribution



Tin_Tin Highly Voted 2 years, 6 months ago
DSVM Windows does not support PostgreSQL.
upvoted 8 times

rishi_ram Highly Voted 3 years, 2 months ago
Deep Learning Virtual Machine (DLVM) and Data Science Virtual Machine (DSVM) can be used for machine learning workloads and come with pre-installed tools and frameworks. In this case, you can choose either the DLVM or DSVM based on your preference and specific requirements. Both virtual machine types can support the use of a PostgreSQL database and GPU processing for your machine learning model. To summarize, both DLVM and DSVM can satisfy the requirements of using a PostgreSQL database and GPU processing for your machine learning model. You can choose the one that best suits your needs and preferences. Answer should be A I am updating the answer
upvoted 6 times

sar77 Most Recent 1 year ago
The Data Science Virtual Machine (DSVM) Windows edition is a solid choice for this scenario. It supports GPU-based processing when deployed on an appropriate N-series VM SKU, and while PostgreSQL is not preinstalled, it can be easily added manually — which is acceptable in this context since the recommendation implies extensibility. Why this recommendation satisfies the requirements: DSVM includes deep learning frameworks (e.g., TensorFlow, PyTorch) and GPU drivers. It can be deployed on GPU-enabled hardware for high-performance training. PostgreSQL can be installed on Windows Server 2019, which is the base OS for DSVM. DSVM is designed for flexible data science workloads, including custom setups. So yes — this VM type meets both the GPU and PostgreSQL requirements when configured properly.
upvoted 2 times

sanctafrax 1 year, 6 months ago
Question states "built into it". DSVM windows does support PostgreSQL, but its not natively installed. therefore an out of the box DSVM windows install will not "run" PostgreSQL. Only DSVM ubuntu has it preinstalled. You can however install PostgreSQL, and it will work. But that is NOT what the question states. https://learn.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/tools-included?view=azureml-api-2
upvoted 2 times

astone42 1 year, 7 months ago
The question specifically mentions "necessary tools built into it", DSVM windows edition does not have PostgreSQL.
upvoted 1 times

Xsesi 2 years ago
I would stand for A since: firstly, DSVM Windows has pre-configured GPU Option; secondly, PostgreSQL can be installed and run on various operating systems, including Windows. And it's possible to connect to a PostgreSQL server from the DSVM Windows edition.
upvoted 1 times

james2033 2 years, 10 months ago

https://learn.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/tools-included?view=azureml-api-2#build-deep-learning-and-machine-learning-solutions

upvoted 2 times

eternaleclipse 2 years, 10 months ago

This wasn't in the exam study material

upvoted 3 times

daviduzo 3 years, 1 month ago

Option A is correct/ DSVM windows edition certifies the requirements

upvoted 1 times

rishi_ram 3 years, 2 months ago

https://learn.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/tools-included?view=azureml-api-2 Postgress is not available for windows edition, Please check the link . Answer should be B

upvoted 4 times

Truman 3 years, 4 months ago

The Windows edition of the Data Science Virtual Machine (DSVM) does not come with GPU support by default. If you need GPU processing for your machine learning workload, you should choose the Linux edition of the DSVM, which comes with NVIDIA GPU drivers pre-installed. The Linux edition of the DSVM is designed for data science workloads that require GPU processing and includes a variety of pre-installed tools and libraries that can be useful for machine learning workloads."The Windows editions of the DSVM come preinstalled with GPU drivers, frameworks, and GPU versions of deep learning frameworks. On the Linux editions, deep learning on GPUs is enabled on the Ubuntu DSVMs." https://learn.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview?view=azureml-api-2

upvoted 1 times

Truman 3 years, 4 months ago

Selected Answer: B

The Windows edition of the Data Science Virtual Machine (DSVM) does not come with GPU support by default. If you need GPU processing for your machine learning workload, you should choose the Linux edition of the DSVM, which comes with NVIDIA GPU drivers pre-installed. The Linux edition of the DSVM is designed for data science workloads that require GPU processing and includes a variety of pre-installed tools and libraries that can be useful for machine learning workloads.

upvoted 1 times

rashjan 3 years, 3 months ago

"The Windows editions of the DSVM come preinstalled with GPU drivers, frameworks, and GPU versions of deep learning frameworks. On the Linux editions, deep learning on GPUs is enabled on the Ubuntu DSVMs." https://learn.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview?view=azureml-api-2

upvoted 1 times

emmanuelodenyire 3 years, 6 months ago

I know most of the people will choose B, but for me I will stand with a Yes. My choice: A A Data Science Virtual Machine (DSVM) is a pre-configured virtual machine image on Microsoft Azure that is equipped with a comprehensive set of tools that are commonly used in data science and deep learning, including a GPU-enabled version. The Windows edition of the DSVM includes PostgreSQL, making it a suitable choice for the task of running a machine learning model that requires a PostgreSQL database and GPU processing to forecast prices. Therefore, the recommendation to use a DSVM Windows edition will satisfy the requirements.

upvoted 4 times

victorafb 3 years, 10 months ago

I'ts corret guys, if as our colleague said that PostgreeSQL is avaliable in linux, this schema shows that is also avaliable in Windows DSVM. https://azure.microsoft.com/en-us/products/virtual-machines/data-science-virtual-machines/#product-overview

upvoted 3 times

ning 4 years, 2 months ago

Selected Answer: B

PostgreSQL only mentioned in Linux VM, at least per documentation ... It should be B

upvoted 4 times

ning 4 years, 2 months ago

PostgreSQL only mentioned in Linux VM, at least per documentation ... It should be Bnope The Data Science Virtual Machine (DSVM) is a customized VM image on the Azure cloud platform built specifically for doing data science. It has many popular data science tools preinstalled and pre-configured to jump-start building intelligent applications for advanced analytics. The DSVM is available on: Windows Server 2019 Ubuntu 18.04 LTS Ubuntu 20.04 LTS https://learn.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overviewSure it has many popular data science tools but does it have anything that can access a PostgreSQL database preinstalled? The linux one does.

upvoted 4 times

exnaniantwort 3 years, 11 months ago

nope The Data Science Virtual Machine (DSVM) is a customized VM image on the Azure cloud platform built specifically for doing data science. It has many popular data science tools preinstalled and pre-configured to jump-start building intelligent applications for advanced analytics. The DSVM is available on: Windows Server 2019 Ubuntu 18.04 LTS Ubuntu 20.04 LTS https://learn.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overviewSure it has many popular data science tools but does it have anything that can access a PostgreSQL database preinstalled? The linux one does.

upvoted 3 times

exnaniantwort 3 years, 11 months ago

nope

The Data Science Virtual Machine (DSVM) is a customized VM image on the Azure cloud platform built specifically for doing data science. It has many popular data science tools preinstalled and pre-configured to jump-start building intelligent applications for advanced analytics.

The DSVM is available on:

Windows Server 2019 Ubuntu 18.04 LTS Ubuntu 20.04 LTS

https://learn.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview

upvoted 3 times

SweetChilliPhilly 3 years, 9 months ago

Sure it has many popular data science tools but does it have anything that can access a PostgreSQL database preinstalled? The linux one does.

upvoted 3 times

peppertool 4 years, 2 months ago

Should be deep learning virtual machine linux edition

upvoted 3 times

pancman 4 years, 4 months ago

Answer is correct. Refer to: https://azuremarketplace.microsoft.com/en-ca/marketplace/apps/microsoft-dsvm.dsvm-win-2019?tab=Overview

upvoted 1 times

ranjsi01 4 years, 6 months ago

B. It should be DSVM Linux edition

upvoted 4 times

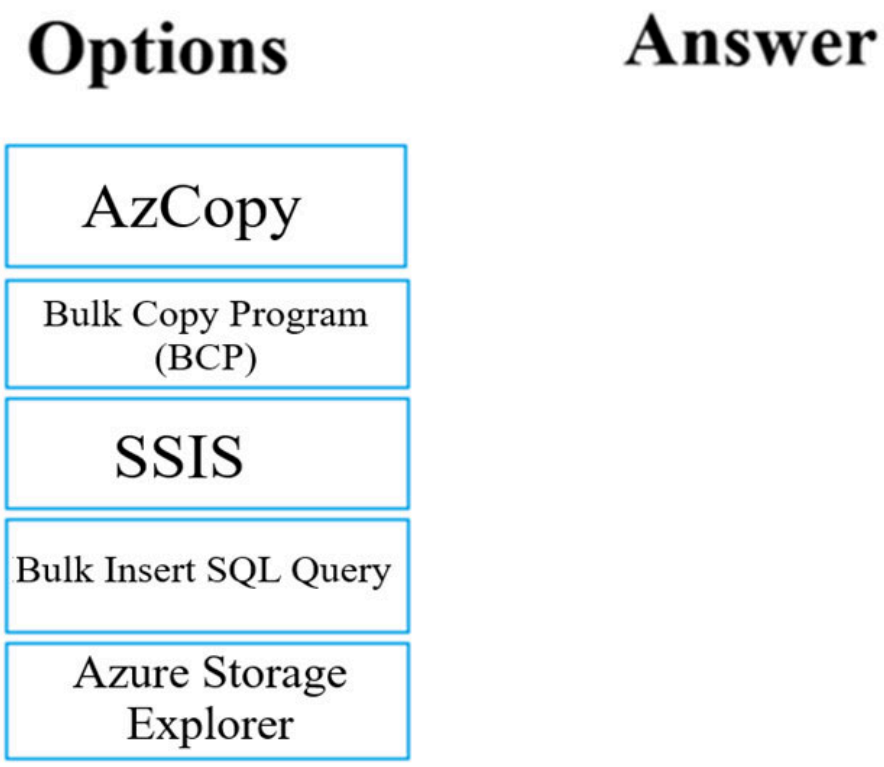
Question #7 *Topic 1*

DRAG DROP -

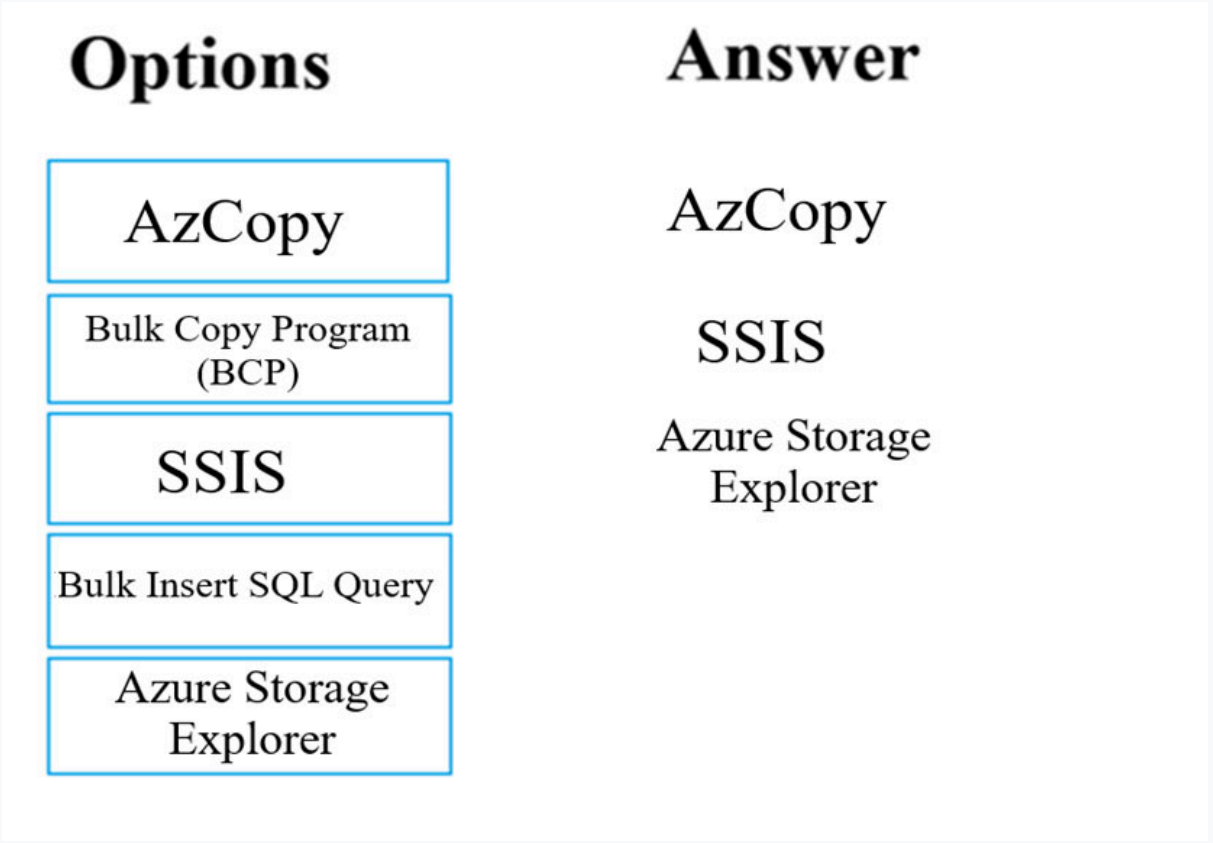
You have been tasked with moving data into Azure Blob Storage for the purpose of supporting Azure Machine Learning.

Which of the following can be used to complete your task? Answer by dragging the correct options from the list to the answer area.

Select and Place:



Suggested Answer:

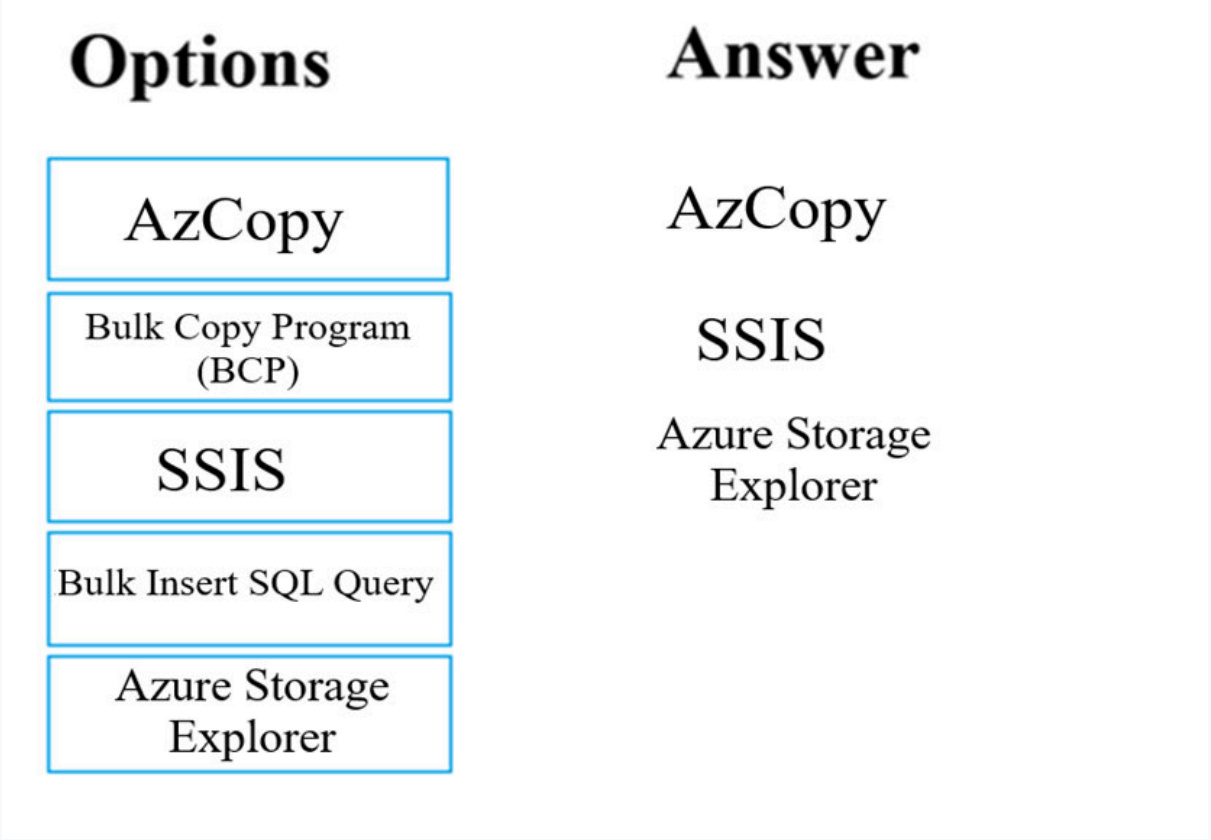


Different technologies for moving data

The following articles describe how to move data to and from Azure Blob storage using different technologies.

- Azure Storage-Explorer
- AzCopy
- Python
- SSIS

Which method is best for you depends on your scenario. The Scenarios for advanced analytics in Azure Machine Learning article helps you determine the resources you need for a variety of data science workflows used in the advanced analytics process.



You can move data to and from Azure Blob storage using different technologies: ➪ Azure Storage-Explorer ➪ AzCopy ➪ Python ➪ SSIS
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/move-azure-blob

mamau Highly Voted 2 years, 6 months ago
Given is correct,also can use python script
upvoted 6 times

james2033 Most Recent 1 year, 10 months ago
SSIS stands for 'SQL Server Integration Services'
upvoted 3 times

emmanuelodenyire 2 years, 6 months ago
I think the given is correct
upvoted 1 times

ning 3 years, 2 months ago
Correctcorrectcorrect
upvoted 4 times

ning 3 years, 2 months ago
Correct
upvoted 4 times

Gabonia 3 years ago
correct
upvoted 1 times

Gabonia 3 years ago
correctcorrect
upvoted 1 times

kay000001 2 years ago
correct
upvoted 1 times

Question #8 *Topic 1*

HOTSPOT -

Complete the sentence by selecting the correct option in the answer area.

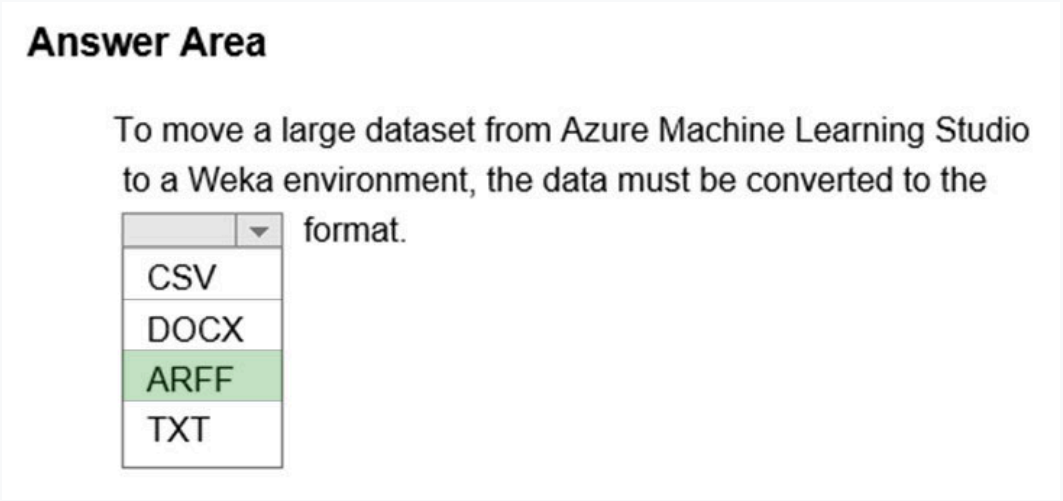
Hot Area:

Answer Area

To move a large dataset from Azure Machine Learning Studio to a Weka environment, the data must be converted to the [ ] format.

- CSV
- DOCX
- ARFF
- TXT

Suggested Answer:



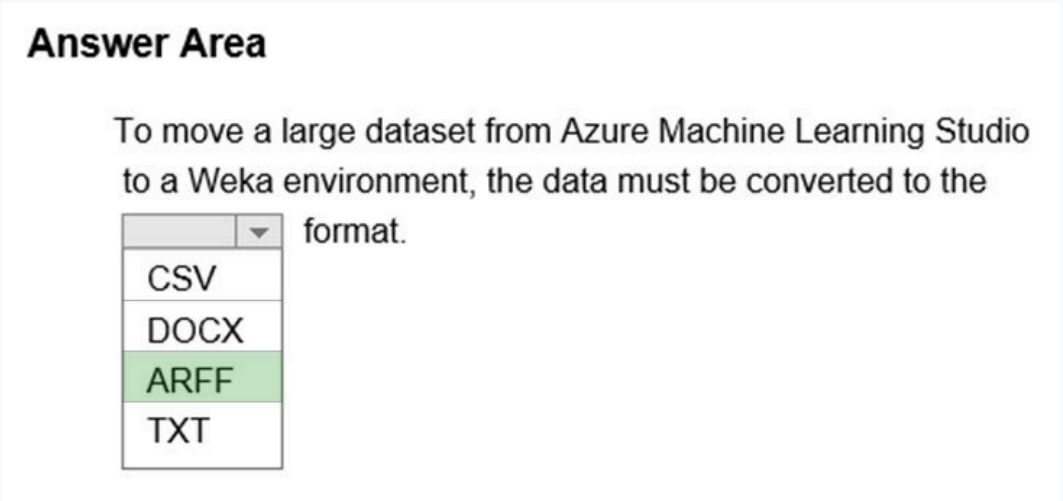
Module overview

This article describes how to use the Convert to ARFF module in Machine Learning Studio (classic), to convert datasets and results the attribute-relation file format used by the Weka toolset. This format is known as ARFF.

The ARFF data specification for Weka supports multiple machine learning tasks, including data preprocessing, classification, and feature selection. In this format, data is organized by entites and their attributes, and is contained in a single text file. You can find details of the Weka file format in the Technical Notes section.

In general, conversion to the Weka file format is required only if you want to use both Machine Learning and Weka, and intend to move your training data back and forth between them.

For more information about the Weka toolset, see this Wikipedia article: Weka (machine learning)



Use the Convert to ARFF module in Azure Machine Learning Studio, to convert datasets and results in Azure Machine Learning to the attribute-relation file format used by the Weka toolset. This format is known as ARFF. The ARFF data specification for Weka supports multiple machine learning tasks, including data preprocessing, classification, and feature selection. In this format, data is organized by entities and their attributes, and is contained in a single text file. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/convert-to-arff

eternaleclipse Highly Voted 2 years, 4 months ago

This wasn't in the exam study material

upvoted 8 times

oliveirafernandaz Most Recent 1 year, 10 months ago

I didn't see anything about Weka on Azure study material

upvoted 4 times

james2033 2 years, 4 months ago

ARFF: Attribute-relation file format used by Weka.

upvoted 2 times

PopeyeDS 2 years, 7 months ago

Weka, a popular machine learning and data mining software, supports the ARFF (Attribute-Relation File Format) as its standard data format. The ARFF format is a plain text file format that allows you to define the attributes and data instances in a structured manner.

upvoted 2 times

Manlogs 3 years, 4 months ago

Correct

upvoted 1 times

Gabonia 3 years, 6 months ago

correct

upvoted 1 times

Question #9 *Topic 1*

You have been tasked with designing a deep learning model, which accommodates the most recent edition of Python, to recognize language.
You have to include a suitable deep learning framework in the Data Science Virtual Machine (DSVM).
Which of the following actions should you take?

A. You should consider including Rattle.

B. You should consider including TensorFlow.

C. You should consider including Theano.

D. You should consider including Chainer.

Suggested Answer: *B*

Machine learning is a complex discipline. But implementing machine learning models is far less daunting and difficult than it used to be, thanks to machine learning frameworks—such as Google's TensorFlow—that ease the process of acquiring data, training models, serving predictions, and refining future results.

Created by the Google Brain team, TensorFlow is an open source library for numerical computation and large-scale machine learning. TensorFlow bundles together a slew of machine learning and deep learning (aka neural networking) models and algorithms and makes them useful by way of a common metaphor. It uses Python to provide a convenient front-end API for building applications with the framework, while executing those applications in high-performance C++.

Machine learning is a complex discipline. But implementing machine learning models is far less daunting and difficult than it used to be, thanks to machine learning frameworks—such as Google's TensorFlow—that ease the process of acquiring data, training models, serving predictions, and refining future results.

Created by the Google Brain team, TensorFlow is an open source library for numerical computation and large-scale machine learning. TensorFlow bundles together a slew of machine learning and deep learning (aka neural networking) models and algorithms and makes them useful by way of a common metaphor. It uses Python to provide a convenient front-end API for building applications with the framework, while executing those applications in high-performance C++.

Reference: https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html

Community vote distribution

B (100%)

Gabonia Highly Voted 4 years ago

Correct

upvoted 6 times

Pradiksha Most Recent 1 year, 9 months ago

The task is to do language recognition which is an NLP task and as given we need to accomadate python. 1. Rattle is a free, open-source graphical user interface (GUI) for the R software that helps beginners learn data mining, machine learning, and predictive analytics. not for python. 2. Though Theano is a foundation for early deep learning research, it lacks active support and new features. so It's not suitable for modern NLP tasks. 3. Chainer, while capable, has less support and community resources compared to TensorFlow and is no longer actively developed, making it less suitable for current and future NLP tasks. 4. So for a language recognition task on the DSVM, TensorFlow is more suitable than others

upvoted 1 times

emmanuelodenyire 1 year, 10 months ago

I will stand with option B. TensorFlow is an open-source software library for data flow and differentiable programming across a range of tasks. It was developed by the Google Brain team and is used for building and training machine learning models, particularly neural networks. TensorFlow provides a flexible and efficient platform for implementing machine learning algorithms, and supports a variety of programming languages including Python, C++, and Java. The library provides a comprehensive set of tools and functionality, including visualization and debugging tools, and it has a strong community of developers and users contributing to its ongoing development. TensorFlow is widely used in both academia and industry for various applications such as image classification, natural language processing, and reinforcement learning.

upvoted 1 times

james2033 2 years, 10 months ago

PyTorch

upvoted 3 times

AsuraX 2 years, 10 months ago

Correct

upvoted 1 times

orionduo 2 years, 12 months ago

Correct

upvoted 1 times

orionduo 3 years, 6 months ago

The answer is correct

upvoted 1 times

Mirjalol 3 years, 6 months ago

Tensorflow is correct

upvoted 1 times

Question #10 *Topic 1*

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You have been tasked with evaluating your model on a partial data sample via k-fold cross-validation.
You have already configured a k parameter as the number of splits. You now have to configure the k parameter for the cross-validation with the usual value choice.
Recommendation: You configure the use of the value k=3.
Will the requirements be satisfied?

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Community vote distribution
B (100%)

saurabh288 Highly Voted 5 years ago
Usual choice is key word here and usual choice is K=5 or 10. So answer is B.
upvoted 13 times

gaint Highly Voted 5 years, 1 month ago
Answer should be AAgree. Why can't be 3? 3 is not “the usual value choice”
upvoted 7 times

gaint 5 years, 1 month ago
Answer should be A
upvoted 7 times

YipingRuan 5 years ago
Agree. Why can't be 3? 3 is not “the usual value choice”
upvoted 3 times

emmanuelodenyire Most Recent 1 year, 10 months ago
I will go with B. The answer to the question is B. No. The recommendation to use k=3 is a common practice in k-fold cross-validation, but it may not necessarily satisfy the requirements in every case. It depends on the specific requirements of the task and the characteristics of the data sample. For example, if the data sample is small, using k=3 may not provide enough training data for the model to learn from, resulting in a high variance in the evaluation metric. In this case, a larger value of k may be more appropriate. On the other hand, if the data sample is very large, using k=3 may result in a low bias but high variance, in which case a smaller value of k may be more appropriate. Therefore, it's important to consider the specific requirements and characteristics of the task and data sample when choosing the value of k for k-fold cross-validation. In general, the recommendation to use k=3 is a good starting point, but it may not always be the best choice.
upvoted 3 times

evangelist 2 years, 6 months ago
The usual choice for k in k-fold cross-validation, especially in the context of evaluating machine learning models, is typically k=5 or k=10
upvoted 1 times

james2033 2 years, 10 months ago
Default value k = 10
upvoted 1 times

lookaaaa 3 years, 9 months ago
Usually we would choose k = 10
upvoted 1 times

Edriv 3 years, 9 months ago
k-fold involves many more procedures, not only k configuration
upvoted 1 times

ning 4 years, 2 months ago

I have no idea for this one, usual value = 10 default, and in designer you cannot change that, but you can add a partition and sample step to set K=3, so what is the question asking for???

upvoted 1 times

Thornehead 4 years, 4 months ago

Usually K = 5 or 10. So B is the correct answer.

upvoted 1 times

synapse 4 years, 5 months ago

Ok.. it says "Usual value choice" -- I guess then the answer is No... it should be 5 or 10.

upvoted 1 times

synapse 4 years, 5 months ago

I don't think anyone restricts you from selecting 3.

upvoted 2 times

Tsardoz 4 years, 7 months ago

You can use as many splits as you want. It all depends on the data. Train/test/validate is basically 3 splits that are just swapped around. 3 is perfectly fine.

upvoted 3 times

dija123 4 years, 8 months ago

3 is not the usual value choice

upvoted 1 times

sim39 4 years, 11 months ago

k can absolutely be 3

upvoted 2 times

snsnsnsn 4 years, 11 months ago

answer is correct

upvoted 2 times

dijaa 4 years, 11 months ago

answer is correct, the default is 10

upvoted 4 times

dijaa 4 years, 11 months ago

answer is correct, the default is 10 Not only 10, sometime 5 is commonly used

upvoted 4 times

dushmantha 4 years, 11 months ago

Not only 10, sometime 5 is commonly used

upvoted 3 times

Question #11 *Topic 1*

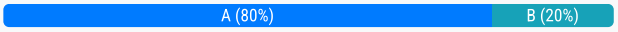
This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You have been tasked with evaluating your model on a partial data sample via k-fold cross-validation.
You have already configured a k parameter as the number of splits. You now have to configure the k parameter for the cross-validation with the usual value choice.
Recommendation: You configure the use of the value k=10.
Will the requirements be satisfied?

A. Yes

B. No

Suggested Answer: *A*
Leave One Out (LOO) cross-validation Setting K = n (the number of observations) yields n-fold and is called leave-one out cross-validation (LOO), a special case of the K-fold approach. LOO CV is sometimes useful but typically doesn't shake up the data enough. The estimates from each fold are highly correlated and hence their average can have high variance. This is why the usual choice is K=5 or 10. It provides a good compromise for the bias-variance tradeoff.

Community vote distribution



pancman Highly Voted 4 years, 4 months ago
I highly doubt that anyone would ask you for what the usual value for k is, in an actual exam. I would skip this question.
upvoted 5 times

lianaliam Most Recent 1 year, 2 months ago
k muss be 5 or 10
upvoted 1 times

emmanuelodenyire 1 year, 10 months ago
The answer is A. Yes, because k=10 is a common choice for the number of folds in k-fold cross-validation, and it can satisfy the requirement to evaluate the model on a partial data sample. However, the appropriate value for k may depend on the size of the dataset and the desired level of accuracy. In some cases, a value other than 10 might be more appropriate.
upvoted 1 times

evangelist 2 years, 6 months ago
The usual choice for k in k-fold cross-validation, especially in the context of evaluating machine learning models, is typically k=5 or k=10
upvoted 2 times

james2033 2 years, 10 months ago
k = 10 . Quote 'The algorithm defaults to 10 folds if you have not previously partitioned the dataset. ' at https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/cross-validate-model?view=azureml-api-2#how-cross-validation-works
upvoted 2 times

PradhanManva 2 years, 10 months ago
This is the answer.
upvoted 1 times

Skychaser 4 years, 3 months ago
5 or 10
upvoted 2 times

ranjsi01 4 years, 6 months ago
5 or 10
upvoted 1 times

dija123 4 years, 8 months ago
Selected Answer: A
10 is the usual value choice
upvoted 1 times

dija123 4 years, 8 months ago

10 is the usual value choiceThe usual value choice for the k parameter in k-fold cross-validation is 10. This value is commonly used because it provides a good balance between having a high number of folds (which results in a lower bias but a higher variance) and a low number of folds (which results in a higher bias but a lower variance)

upvoted 1 times

CSLK 3 years, 6 months ago

The usual value choice for the k parameter in k-fold cross-validation is 10. This value is commonly used because it provides a good balance between having a high number of folds (which results in a lower bias but a higher variance) and a low number of folds (which results in a higher bias but a lower variance)

upvoted 1 times

Question #12 *Topic 1*

You construct a machine learning experiment via Azure Machine Learning Studio.
You would like to split data into two separate datasets.
Which of the following actions should you take?

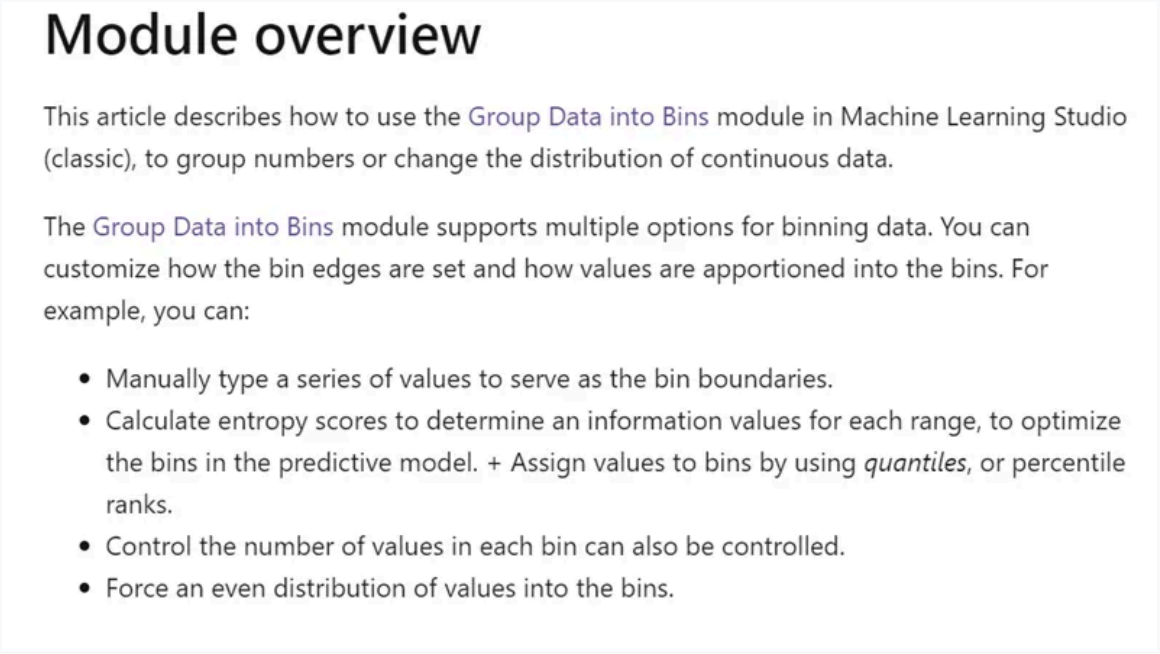
A. You should make use of the Split Data module.

B. You should make use of the Group Categorical Values module.

C. You should make use of the Clip Values module.

D. You should make use of the Group Data into Bins module.

Suggested Answer: *D*



Module overview

This article describes how to use the Group Data into Bins module in Machine Learning Studio (classic), to group numbers or change the distribution of continuous data.

The Group Data into Bins module supports multiple options for binning data. You can customize how the bin edges are set and how values are apportioned into the bins. For example, you can:

- Manually type a series of values to serve as the bin boundaries.
- Calculate entropy scores to determine an information values for each range, to optimize the bins in the predictive model. + Assign values to bins by using *quantiles*, or percentile ranks.
- Control the number of values in each bin can also be controlled.
- Force an even distribution of values into the bins.

Module overview

This article describes how to use the Group Data into Bins module in Machine Learning Studio (classic), to group numbers or change the distribution of continuous data.

The Group Data into Bins module supports multiple options for binning data. You can customize how the bin edges are set and how values are apportioned into the bins. For example, you can:

- Manually type a series of values to serve as the bin boundaries.
- Calculate entropy scores to determine an information values for each range, to optimize the bins in the predictive model. + Assign values to bins by using *quantiles*, or percentile ranks.
- Control the number of values in each bin can also be controlled.
- Force an even distribution of values into the bins.

The Group Data into Bins module supports multiple options for binning data. You can customize how the bin edges are set and how values are apportioned into the bins. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

Community vote distribution

A (100%)

dp100uber Highly Voted 4 years, 7 months ago

Split data is correct, not D.I agree

upvoted 34 times

dp100uber Highly Voted 4 years, 7 months ago

Selected Answer: A

Split data is correct, not D.

upvoted 34 times

Gabonia 4 years ago

I agree

upvoted 1 times

lianaliam Most Recent 1 year, 2 months ago

test_split to two part

upvoted 1 times

PopeyeDS 1 year, 10 months ago

The Split Data module is specifically designed for dividing a dataset into multiple parts. It allows you to specify the ratio or proportion of data to allocate to each resulting dataset. By configuring the Split Data module, you can split your data into two separate datasets based on your desired split ratio, such as 70% for training and 30% for testing.

upvoted 2 times

evangelist 1 year, 10 months ago

A. You should make use of the Split Data module. The Split Data module in Azure Machine Learning Studio is specifically designed for dividing datasets into two or more parts based on specified criteria. This is useful for creating training and testing datasets, which is a common requirement in machine learning experiments to evaluate the performance of models. The other modules mentioned serve different purposes unrelated to splitting datasets for training and testing.

upvoted 2 times

eternaleclipse 2 years, 10 months ago

A. This material is not updated.

upvoted 1 times

james2033 2 years, 10 months ago

Split Data module.

upvoted 1 times

PradhanManva 2 years, 10 months ago

This is the answer.

upvoted 1 times

endeesa 3 years, 2 months ago

Split data module is the correct answer

upvoted 1 times

frogman_8654 3 years, 2 months ago

https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/split-data?view=azureml-api-2

upvoted 2 times

musuai 3 years, 4 months ago

Know-it-all Bot: A. You should make use of the Split Data module. Me: Why not D? Know-it-all Bot: The Group Data into Bins module is used to group continuous data into discrete bins or intervals. It is not appropriate for splitting data into separate datasets. In contrast, the Split Data module is specifically designed for splitting data into training and testing datasets based on a specified percentage or other criteria, and would be the appropriate choice in this scenario.

upvoted 1 times

Truman 3 years, 4 months ago

To split data into two separate datasets in Azure Machine Learning Studio, you can use the "Split Data" module, which can be found in the "Data Transformation" category in the module palette.

upvoted 2 times

varunhebbar 3 years, 4 months ago

Option A

upvoted 1 times

Ammy_b 3 years, 6 months ago

Answer : A - Split the Data Module will able to split the dataset into two like training and testing

upvoted 1 times

Mirjalol 3 years, 6 months ago

D is for 'old' Machine Learning studio, now Azure ML studio has Split option

upvoted 2 times

ZIMARAKI 3 years, 6 months ago

A is the correct answer

upvoted 1 times

jais7508 3 years, 9 months ago

https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/split-data A is the answer , the ans selected as B is selected based on old Machine Learning Studio but now azure machine learning is used to perform any task and according to that A is right.
https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/cross-validate-model

upvoted 1 times

JTWang 3 years, 10 months ago

A is the answer!

upvoted 2 times

Question #13 *Topic 1*

You have been tasked with creating a new Azure pipeline via the Machine Learning designer.
You have to makes sure that the pipeline trains a model using data in a comma-separated values (CSV) file that is published on a website. A dataset for the file for this file does not exist.
Data from the CSV file must be ingested into the designer pipeline with the least amount of administrative effort as possible.
Which of the following actions should you take?

A. You should make use of the Convert to TXT module.

B. You should add the Copy Data object to the pipeline.

C. You should add the Import Data object to the pipeline.

D. You should add the Dataset object to the pipeline.

Suggested Answer: *D*
The preferred way to provide data to a pipeline is a Dataset object. The Dataset object points to data that lives in or is accessible from a datastore or at a Web URL. The Dataset class is abstract, so you will create an instance of either a FileDataset (referring to one or more files) or a TabularDataset that's created by from one or more files with delimited columns of data. Example: from azureml.core import Dataset iris_tabular_dataset = Dataset.Tabular.from_delimited_files([(def_blob_store, 'train-dataset/iris.csv')]) Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-your-first-pipeline

Community vote distribution
C (100%)

JTWang Highly Voted 4 years, 4 months ago
The Import Data component support read data from following sources: URL via HTTP Azure cloud storages through Datastores) Azure Blob Container Azure File Share Azure Data Lake Azure Data Lake Gen2 Azure SQL Database Azure PostgreSQL
upvoted 15 times

lianaliam Most Recent 1 year, 2 months ago
import
upvoted 1 times

shahid.azad 2 years ago
why the revealed tell wrong answer
upvoted 1 times

evangelist 2 years, 6 months ago
C. You should add the Import Data object to the pipeline. Adding the Import Data object to the pipeline in Azure Machine Learning designer is the most efficient way to ingest data from a CSV file published on a website. This action allows you to directly access and import the data into the pipeline with minimal administrative effort, eliminating the need for manual data conversion or copying. The Import Data object supports various data sources, including web URLs, making it ideal for this task.
upvoted 1 times

Matt2000 2 years, 6 months ago
The instruction explicitly says that there is no dataset registered for the csv file. D (add dataset object) presupposes that such a dataset exists and thus is False. C (import data module) remains as the only viable option.
upvoted 1 times

Charoon 2 years, 8 months ago
According to MS in reference to the Import Data component: "All functionality provided by this component can be done by datastore and datasets in the workspace landing page. We recommend you use datastore and dataset which includes additional features like data monitoring" https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/import-data?view=azureml-api-2 Hence DataSet is the preferred method, i.e. answer D.
upvoted 1 times

james2033 2 years, 10 months ago
Let's see screenshot, it is CSV https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/media/module/import-data-path.png?view=azureml-api-2 https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/import-data?view=azureml-api-2#how-to-configure-import-data C is the correct answer: 'Import Data' in pipeline.

upvoted 1 times

PopeyeDS 3 years, 1 month ago

the correct answer is C. You should add the Import Data object to the pipeline. However, if you prefer to create the dataset manually, then you could use the Dataset object.

upvoted 1 times

Sa_Msa 3 years, 1 month ago

The Import Data module in Azure Machine Learning designer allows you to read data from various data sources, including web URLs, and import it directly into your pipeline. By configuring the Import Data object with the URL of the CSV file, you can easily bring the data into the pipeline for further processing.

upvoted 2 times

Mirjalol 3 years, 6 months ago

Dataset for this file does not exist.... and proceeded with answer D? My question is how you add dataset object to pipeline if it does not exist? Like literally how? Answer is C obviously

upvoted 2 times

KingTN 3 years, 6 months ago

There is a note in the link, regarding import-data "https://docs.microsoft.com/en-us/azure/machine-learning/component-reference/import-data", which recommends to use dataset... I am not sure when this note was added.. This is the note: "All functionality provided by this component can be done by datastore and datasets in the workspace landing page. We recommend you use datastore and dataset which includes additional features like data monitoring. To learn more, see How to Access Data and How to Register Datasets article. After you register a dataset, you can find it in the Datasets -> My Datasets category in designer interface. This component is reserved for Studio(classic) users to for a familiar experience." So I think as of Jan 2023, the answer should be registering a dataset (similar to D, but may need to be rephrased).

upvoted 2 times

Vinit9 3 years, 6 months ago

B. You should add the Copy Data object to the pipeline. To ingest data from a CSV file published on a website into the Azure Machine Learning designer pipeline with the least amount of administrative effort possible, you should add the Copy Data object to the pipeline. The Copy Data object is used to copy data from a source to a destination, and it supports a variety of sources, including web URLs. By using the Copy Data object, you can quickly and easily ingest the data from the CSV file into the pipeline, without having to create a dataset or perform any additional preprocessing. This will minimize administrative effort and allow you to quickly get started with training a model using the data.

upvoted 2 times

Mirjalol 3 years, 6 months ago

I have a question for yall fellas: The question mentions that there is no dataset for this file... Can we import data object if there is no dataset of this file?

upvoted 3 times

KIshor1212 3 years, 8 months ago

There are two ways you can import data into the designer: Azure Machine Learning datasets - Register datasets in Azure Machine Learning to enable advanced features that help you manage your data. Import Data component - Use the Import Data component to directly access data from online data sources

upvoted 1 times

pancman 4 years, 4 months ago

The correct answer is C. Import data component is used to import data from data sources such as web URLs with minimum effort. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/component-reference/import-dataThe Import Data component support read data from following sources: URL via HTTP Azure cloud storages through Datastores) Azure Blob Container Azure File Share Azure Data Lake Azure Data Lake Gen2 Azure SQL Database Azure PostgreSQL

upvoted 3 times

pancman 4 years, 4 months ago

Selected Answer: C

The correct answer is C. Import data component is used to import data from data sources such as web URLs with minimum effort. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/component-reference/import-data

upvoted 3 times

Thornehead 4 years, 4 months ago

In Azure Designer, there is no such object as "IMPORT DATA". So the answer is "insert Dataset".No, there is a component called import data and it is in fact the correct answer to this question. Check here: https://docs.microsoft.com/en-us/azure/machine-learning/component-reference/import-dataGod it is tricky

upvoted 3 times

Thornehead 4 years, 4 months ago

In Azure Designer, there is no such object as "IMPORT DATA". So the answer is "insert Dataset".

upvoted 3 times

pancman 4 years, 4 months ago

No, there is a component called import data and it is in fact the correct answer to this question. Check here: https://docs.microsoft.com/en-us/azure/machine-learning/component-reference/import-data

upvoted 4 times

windy610 2 years, 9 months ago

God it is tricky

upvoted 1 times

TheCyanideLancer 4 years, 7 months ago

shouldn't the answer be C ?I think so too, it should be C, as it mentions less administrative effort

upvoted 2 times

TheCyanideLancer 4 years, 7 months ago

shouldn't the answer be C ?

upvoted 2 times

RAHULsingla 4 years, 7 months ago

I think so too, it should be C, as it mentions less administrative effort

upvoted 2 times

Question #14 *Topic 1*

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You are in the process of creating a machine learning model. Your dataset includes rows with null and missing values.
You plan to make use of the Clean Missing Data module in Azure Machine Learning Studio to detect and fix the null and missing values in the dataset.
Recommendation: You make use of the Replace with median option.
Will the requirements be satisfied?

A. Yes

B. No

Suggested Answer: *B*
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

Community vote distribution
A (74%) B (26%)

synapse Highly Voted 4 years, 5 months ago
This is an incomplete question. We don't know what type of data it is. Continuous or categorical. If it's continuous then it's A else its B
upvoted 13 times

Xsytt419 Highly Voted 4 years, 7 months ago
should the answer be YES?
upvoted 6 times

lianaliam Most Recent 1 year, 2 months ago
replace with mean for replace missing values
upvoted 1 times

Vinit9 1 year, 10 months ago
Using the Clean Missing Data module in Azure Machine Learning Studio with the Replace with median option can help detect and fix null and missing values in your dataset. The Replace with median option replaces missing values in a dataset with the median value of the corresponding column. This method of imputing missing values can provide a good balance between preserving the overall distribution of the data and avoiding the introduction of extreme values. By using the Clean Missing Data module with the Replace with median option, you can help ensure that your dataset is cleaned and ready for use in creating a machine learning model, satisfying the requirements.
upvoted 1 times

james2033 1 year, 10 months ago
https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/clean-missing-data#bkmk_ReplaceMissing - Replace using MICE (Multivariate Imputation using Chained Equations) - Custom substitution value - Replace with mean - Replace with median: Calculates the column median value, and uses the median value as the replacement for any missing value in the column. (*) - Replace with mode - Remove entire row - Replace using Probabilistic PCA 'rows with null and missing values', so no removing entire row. I choose YES (*), it is A.
upvoted 2 times

Xsesi 2 years ago
Since we do not know the type of data. Replace with mode would be prefer if the data is categorical.
upvoted 1 times

deyoz 2 years, 5 months ago
Answer is no, median is not appropriate to replace missing values of categorical columns.
upvoted 2 times

evangelist 2 years, 6 months ago
A is correct
upvoted 1 times

Ratz 2 years, 8 months ago
Answer B: All the replace missing value options apply to the column. The question talks about randomly missing values in the row. Hence removing the row will be ideal.
upvoted 2 times

eternaleclipse 2 years, 10 months ago

No. because we don't know the TYPE of data it is. To simply replace with median may not work. What if it was text instead of numerical?

upvoted 2 times

rakeshmk 2 years, 10 months ago

Missing value replacement depends on the nature of your data..median is robust to outliers. Also one can go for mean..

upvoted 1 times

PradhanManva 2 years, 10 months ago

This is the answer.

upvoted 1 times

mefor 3 years ago

Yes, using the "Replace with median" option in the Clean Missing Data module in Azure Machine Learning Studio can help satisfy the requirements of dealing with null and missing values in your machine learning dataset. The median is a suitable option for replacing missing values in numerical features because it's less sensitive to outliers compared to the mean. By choosing this option, the module will identify columns with missing values and replace those missing values with the median value of each respective column. This can help maintain the integrity of your dataset and ensure that your machine learning model receives meaningful input data. However, keep in mind that the choice of replacement strategy can also depend on the nature of your data and the specific requirements of your machine learning problem. It's always a good practice to assess the impact of different imputation methods on your model's performance to find the best strategy for your particular case.

upvoted 1 times

endeesa 3 years, 2 months ago

We simply dont have enough information about the dataset to know if Median substitution will work, so the answer is No

upvoted 1 times

ManuelHenriques 3 years, 5 months ago

You should not assume that it is correct to use median if you don't know if data is continuous or not so B

upvoted 4 times

Obhee 3 years, 6 months ago

Replace with median: Calculates the column median value, and uses the median value as the replacement for any missing value in the column. Applies only to columns that have Integer or Double data types. See the Technical notes section for more information.
https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/clean-missing-data

upvoted 1 times

KIshor1212 3 years, 8 months ago

eplace with median: Calculates the column median value, and uses the median value as the replacement for any missing value in the column. Applies only to columns that have Integer or Double data types. https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/clean-missing-data** Replace with median

upvoted 1 times

KIshor1212 3 years, 8 months ago

Selected Answer: A

eplace with median: Calculates the column median value, and uses the median value as the replacement for any missing value in the column.

Applies only to columns that have Integer or Double data types.

https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/clean-missing-data

upvoted 1 times

KIshor1212 3 years, 8 months ago

** Replace with median

upvoted 1 times

Question #15 *Topic 1*

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You are in the process of creating a machine learning model. Your dataset includes rows with null and missing values.
You plan to make use of the Clean Missing Data module in Azure Machine Learning Studio to detect and fix the null and missing values in the dataset.
Recommendation: You make use of the Custom substitution value option.
Will the requirements be satisfied?

A. Yes

B. No

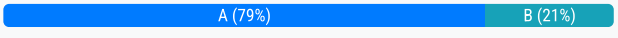
Suggested Answer: *B*





Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

Community vote distribution



endeesa Most Recent 1 year, 10 months ago

As there is a lack of information regarding the dataset in question, utilizing Custom Substitution would be advantageous. This allows for greater control in selecting appropriate substitutions that align with the data, resulting in a favourable outcome. Therefore, the answer is yes.

upvoted 1 times

james2033 1 year, 10 months ago

See latest reference document at here https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/clean-missing-data?view=azureml-api-2#:~:text=Custom%20substitution%20value%3A%20Use%20this%20option%20to%20specify%20a%20placeholder%20value Quote 'For Cleaning Mode, select one of the following options for replacing or removing missing values: Custom substitution value: Use this option to specify a placeholder value (such as a 0 or NA) that applies to all missing values. The value that you specify as a replacement must be compatible with the data type of the column.' Yes. We can use 'Custom substitution value'.

upvoted 3 times

james2033 1 year, 10 months ago

https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/clean-missing-data#bkmk_ReplaceMissing - Replace using MICE (Multivariate Imputation using Chained Equations) - Custom substitution value (*) - Replace with mean - Replace with median: Calculates the column median value, and uses the median value as the replacement for any missing value in the column. - Replace with mode - Remove entire row - Replace using Probabilistic PCA 'rows with null and missing values', so no removing entire row. I choose YES (*)

upvoted 3 times

evangelist 2 years, 6 months ago

If the requirement is to effectively handle missing values in a way that preserves the integrity of the dataset for machine learning purposes, you might need to consider other options as well, such as imputation methods (mean, median, mode, MICE) that can maintain statistical properties of the dataset. The best method depends on the nature of your data and the specific requirements of your machine learning model.

upvoted 1 times

Ratz 2 years, 8 months ago

Answer B: All the replace missing value options apply to the column. The question talks about randomly missing values in the row. Hence removing the row will be ideal.

upvoted 1 times

mhmichiel 3 years, 5 months ago

You dont know the dataset and therefore you cant assume A is correct. This means B is correct in this case. The only option that could be correct withoud seing the dataset is to remove the rows when missing values exist.

upvoted 2 times

Gary_Chambers 3 years, 6 months ago

I think the tricky part is the null values. Clean missing data lets you input for that but doesn't address the null or NaN values.

upvoted 1 times

Mirjalol 3 years, 6 months ago

Answer is 'A', here is the reference for those who have doubts: https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/clean-missing-data

upvoted 1 times

meysa 3 years, 6 months ago

If the columns have different data types, we can not set one custom substitution for all the missing values. The only strategy that can apply to all data types is the mode.

upvoted 1 times

roncil 3 years, 7 months ago

yes the answer is A.

upvoted 1 times

KIshor1212 3 years, 8 months ago

Custom substitution value: Use this option to specify a placeholder value (such as a 0 or NA) that applies to all missing values. The value that you specify as a replacement must be compatible with the data type of the column. https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/clean-missing-data

upvoted 2 times

FlexingD 3 years, 9 months ago

Should be A

upvoted 2 times

dinhhungitsoft 3 years, 10 months ago

A is correct, Clean Missing Data module also provides "Custom substitution value" cleaning mode

upvoted 2 times

synapse 4 years, 5 months ago

Since it says custom, the answer would be Yes. AWhat a ridiculous answer you are giving? The question did not mention 'custom' only suggested answer is custom option... If you see word 'custom', do you always choose it as correct answer?

upvoted 3 times

synapse 4 years, 5 months ago

Selected Answer: A

Since it says custom, the answer would be Yes. A

upvoted 3 times

Mirjalol 3 years, 6 months ago

What a ridiculous answer you are giving? The question did not mention 'custom' only suggested answer is custom option... If you see word 'custom', do you always choose it as correct answer?

upvoted 4 times

ranjsi01 4 years, 6 months ago

answer should be YES, because we can use '0' for numeric and 'na' for text columnsIf you use a 0 for numeric, it is a value that you add to that particular row/cell. And this will impact your predictions. It should be a NO.

upvoted 1 times

ranjsi01 4 years, 6 months ago

answer should be YES, because we can use '0' for numeric and 'na' for text columns

upvoted 1 times

zaidurfshahr 4 years, 4 months ago

If you use a 0 for numeric, it is a value that you add to that particular row/cell. And this will impact your predictions. It should be a NO.

upvoted 3 times

Question #16 *Topic 1*

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You are in the process of creating a machine learning model. Your dataset includes rows with null and missing values.
You plan to make use of the Clean Missing Data module in Azure Machine Learning Studio to detect and fix the null and missing values in the dataset.
Recommendation: You make use of the Remove entire row option.
Will the requirements be satisfied?

A. Yes

B. No

Suggested Answer: *A*
Remove entire row: Completely removes any row in the dataset that has one or more missing values. This is useful if the missing value can be considered randomly missing. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

Community vote distribution



klowqw Highly Voted 3 years, 11 months ago
not really if you have just a null in a one cell, you don't have to delete the whole rowGood point!Answer is correct. Here is the reference: https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/clean-missing-data
upvoted 8 times

klowqw Highly Voted 3 years, 11 months ago
not really if you have just a null in a one cell, you don't have to delete the whole row
upvoted 8 times

Edriv 3 years, 8 months ago
Good point!
upvoted 1 times

Mirjalol 3 years, 6 months ago
Answer is correct. Here is the reference: https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/clean-missing-data
upvoted 3 times

ranjsi01 4 years, 6 months ago
correct
upvoted 6 times

ranjsi01 4 years, 6 months ago
correctI agree
upvoted 6 times

Gabonia 4 years ago
I agree
upvoted 1 times

vkm_97 Most Recent 1 year, 5 months ago
Azure Machine Learning Studio provides multiple ways to handle missing data, including mean, median, mode, and custom substitution. Avoid deleting rows unless absolutely necessary, as it can introduce bias and reduce the dataset size. The best approach depends on the context, but Microsoft emphasizes data preservation wherever possible.
upvoted 2 times

ZIMARAKI 1 year, 10 months ago
Selected Answer: A
From Microsoft Documentation: This component supports multiple types of operations for "cleaning" missing values, including:

Replacing missing values with a placeholder, mean, or other value Completely removing rows and columns that have missing values Inferring values based on statistical methods

upvoted 1 times

ZIMARAKI 1 year, 10 months ago

From Microsoft Documentation: This component supports multiple types of operations for "cleaning" missing values, including: Replacing missing values with a placeholder, mean, or other value Completely removing rows and columns that have missing values Inferring values based on statistical methodsThat's fine, but that is an option that the module allows you, however, it is not advisable to delete an entire row for a null value when there are relevant columns for the model and it has records in the dataset. I think that from that point of view, the correct answer is B.

upvoted 1 times

fcoguerrero88 2 years, 11 months ago

That's fine, but that is an option that the module allows you, however, it is not advisable to delete an entire row for a null value when there are relevant columns for the model and it has records in the dataset. I think that from that point of view, the correct answer is B.

upvoted 3 times

james2033 1 year, 10 months ago

CLEANING MODE: REMOVE ENTIRE ROW Quote 'Remove entire row: Completely removes any row in the dataset that has one or more missing values. This is useful if the missing value can be considered randomly missing.'

upvoted 2 times

james2033 1 year, 10 months ago

https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/clean-missing-data#bkmk_ReplaceMissing - Replace using MICE (Multivariate Imputation using Chained Equations) - Custom substitution value - Replace with mean - Replace with median: Calculates the column median value, and uses the median value as the replacement for any missing value in the column. - Replace with mode - Remove entire row (*) - Replace using Probabilistic PCA 'rows with null and missing values', we can removing entire row. I choose YES (*)

upvoted 2 times

evangelist 1 year, 10 months ago

allowed options: The Clean Missing Data module offers several options for handling missing data, including: Remove entire row: Removes any row with missing values. Replace using mean/median/mode: Replaces missing values with the mean, median, or mode of the column. Replace with custom value: Replaces missing values with a custom value specified by the user. Replace using MICE: Uses Multiple Imputation by Chained Equations to replace missing values.

upvoted 1 times

mhmichiel 3 years, 5 months ago

correct

upvoted 1 times

ManuelHenriques 3 years, 5 months ago

What if you dataset is not very large and it gets compromised with removing rows? As we do not have that information I think it is not suitable to go for remove rows, so B.

upvoted 1 times

Cococo 3 years, 6 months ago

Read this again - "to detect and fix the null and missing values" FIX, not remove them, not delete them, not get rid of them but FIX (replace with median is the answer I think). Remove entire row: Completely removes any row in the dataset with one or more missing values.

upvoted 2 times

Question #17 *Topic 1*

You need to consider the underlined segment to establish whether it is accurate.
To transform a categorical feature into a binary indicator, you should make use of the Clean Missing Data module.
Select `No adjustment required` if the underlined segment is accurate. If the underlined segment is inaccurate, select the accurate option.

A. No adjustment required.

B. Convert to Indicator Values

C. Apply SQL Transformation

D. Group Categorical Values

Suggested Answer: *B*
Use the Convert to Indicator Values module in Azure Machine Learning Studio. The purpose of this module is to convert columns that contain categorical values into a series of binary indicator columns that can more easily be used as features in a machine learning model. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/convert-to-indicator-values

Community vote distribution
B (93%)

Peeking Highly Voted 3 years, 1 month ago
This is equivalent to one-hot encoding of a categorical column
upvoted 6 times

Sjefen Highly Voted 3 years, 11 months ago
Correct
upvoted 5 times

lianaliam Most Recent 1 year, 2 months ago
OHC for categirical column
upvoted 1 times

Lion007 2 years, 1 month ago
The answer is correct. B "....the Convert to Indicator Values module in Machine Learning Studio (classic). The purpose of this module is to convert columns that contain categorical values into a series of binary indicator columns that can more easily be used as features in a machine learning model." See https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/convert-to-indicator-valuesUsage tips >> If the column contains missing values, a separate indicator column is created for the missing category, with this name: - Missing See https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/convert-to-indicator-values#usage-tips
upvoted 2 times

Lion007 2 years, 1 month ago
Selected Answer: B
The answer is correct. B

"....the Convert to Indicator Values module in Machine Learning Studio (classic). The purpose of this module is to convert columns that contain categorical values into a series of binary indicator columns that can more easily be used as features in a machine learning model." See https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/convert-to-indicator-values
upvoted 2 times

Lion007 2 years, 1 month ago
Usage tips >> If the column contains missing values, a separate indicator column is created for the missing category, with this name: - Missing See https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/convert-to-indicator-values#usage-tips
upvoted 1 times

amiria7 3 years ago
The answer is correct
upvoted 1 times

JTWang 3 years, 4 months ago
Correct
upvoted 3 times

synapse 3 years, 11 months ago

correct

upvoted 2 times

Question #18 *Topic 1*

You need to consider the underlined segment to establish whether it is accurate.
To improve the amount of low incidence cases in a dataset, you should make use of the SMOTE module.
Select `No adjustment required` if the underlined segment is accurate. If the underlined segment is inaccurate, select the accurate option.

A. No adjustment required.

B. Remove Duplicate Rows

C. Join Data

D. Edit Metadata

Suggested Answer: *A*
Use the SMOTE module in Azure Machine Learning Studio to increase the number of underrepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote

Community vote distribution
A (100%)

c025dd6 Most Recent 7 months, 3 weeks ago
It's a shame that this is the answer, in reality smote rarely works. You are much better off just doing stratification or undersampling.
upvoted 1 times

PradhanManva 1 year, 10 months ago
This is the answer.
upvoted 3 times

BakYac 2 years, 6 months ago
SMOTE first selects a minority class instance a at random and finds its k nearest minority class neighbors. The synthetic instance is then created by choosing one of the k nearest neighbors b at random and connecting a and b to form a line segment in the feature space. The synthetic instances are generated as a convex combination of the two chosen instances a and b.
upvoted 3 times

exnaniantwort 2 years, 11 months ago
This article describes how to use the SMOTE component in Azure Machine Learning designer to increase the number of underrepresented cases in a dataset that's used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases. You connect the SMOTE component to a dataset that's imbalanced. There are many reasons why a dataset might be imbalanced. For example, the category you're targeting might be rare in the population, or the data might be difficult to collect. Typically, you use SMOTE when the class that you want to analyze is underrepresented. The component returns a dataset that contains the original samples. It also returns a number of synthetic minority samples, depending on the percentage that you specify. https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/smote
upvoted 4 times

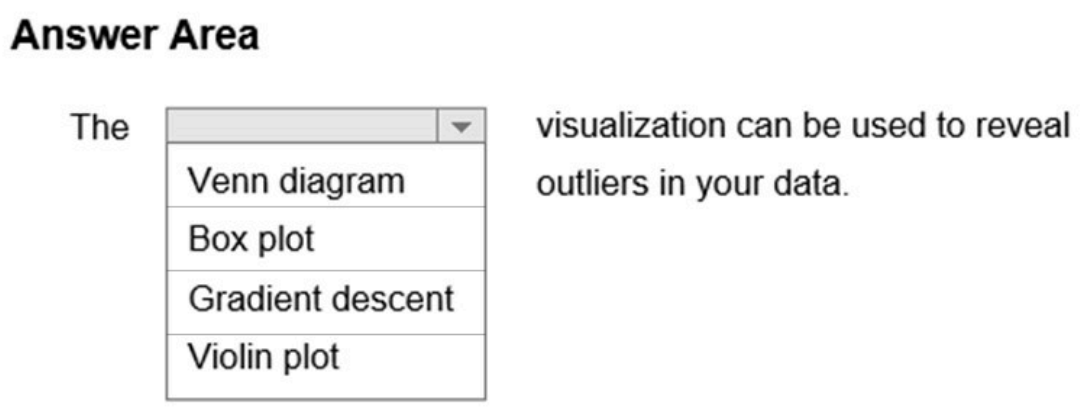
pancman 3 years, 4 months ago
Correct.
upvoted 1 times

Question #19 *Topic 1*

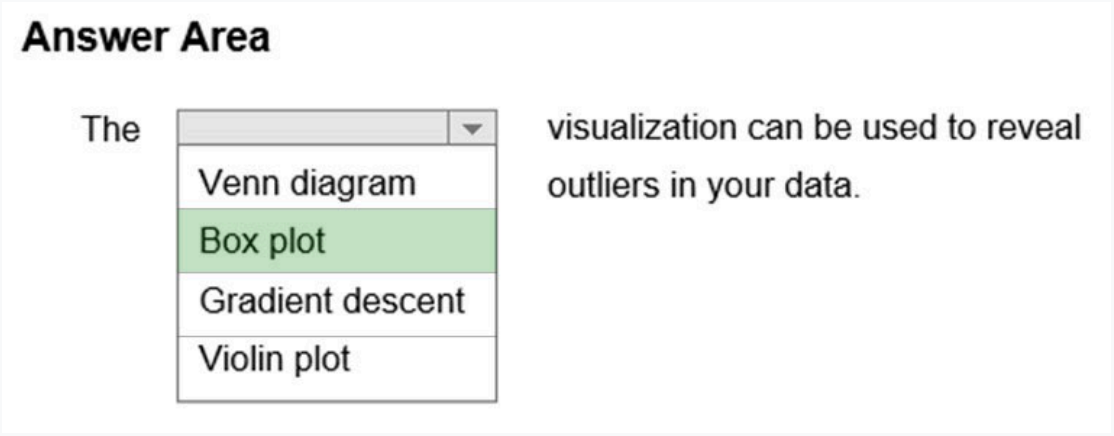
HOTSPOT -

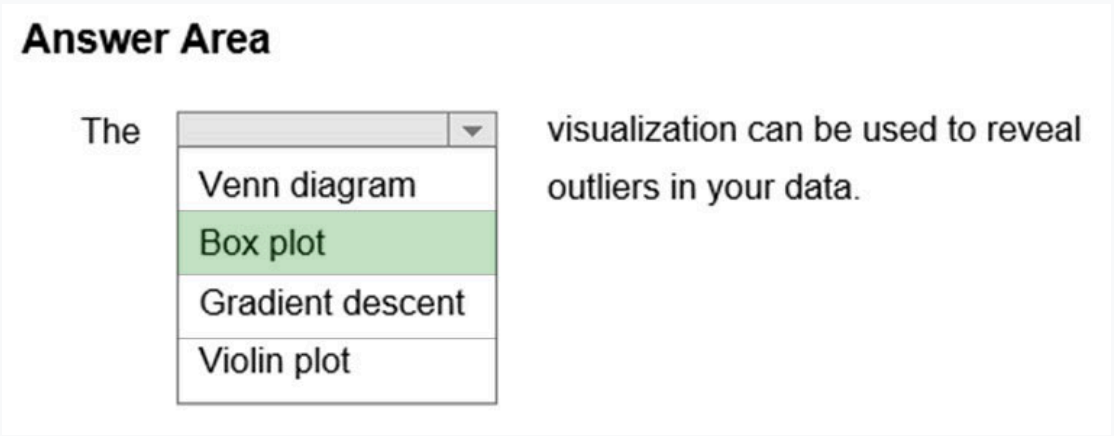
You need to consider the underlined segment to establish whether it is accurate.

Hot Area:



Suggested Answer:





The box-plot algorithm can be used to display outliers. Reference: https://medium.com/analytics-vidhya/what-is-an-outliers-how-to-detect-and-remove-them-which-algorithm-are-sensitive-towards-outliers-2d501993d59

Learner_Thunder2 Highly Voted 4 years, 1 month ago

Box Plot is correctvery trueBox plots/whisker plots are used for finding outliers.

upvoted 7 times

Learner_Thunder2 Highly Voted 4 years, 1 month ago

Box Plot is correct

upvoted 7 times

Gabonia 3 years, 6 months ago

very true

upvoted 2 times

sar77 Most Recent 1 year, 1 month ago

Box plots/whisker plots are used for finding outliers.

upvoted 1 times

Tin_Tin 2 years, 1 month ago

https://learn.microsoft.com/en-us/azure/databricks/visualizations/visualization-types

upvoted 2 times

PradhanManva 2 years, 4 months ago

Box plot

upvoted 1 times

clark88 3 years, 1 month ago

The dots in a boxplot visualize the outliers. So boxplot is correct.

upvoted 1 times

Question #20 *Topic 1*

You are planning to host practical training to acquaint learners with data visualization creation using Python. Learner devices are able to connect to the internet.
Learner devices are currently NOT configured for Python development. Also, learners are unable to install software on their devices as they lack administrator permissions. Furthermore, they are unable to access Azure subscriptions.
It is imperative that learners are able to execute Python-based data visualization code.
Which of the following actions should you take?

A. You should consider configuring the use of Azure Container Instance.

B. You should consider configuring the use of Azure BatchAI.

C. You should consider configuring the use of Azure Notebooks.

D. You should consider configuring the use of Azure Kubernetes Service.

Suggested Answer: *C*
Reference: https://notebooks.azure.com/

Community vote distribution
C (100%)

pancman Highly Voted 4 years, 4 months ago
The answer is correct. Azure Notebooks are accessible by any device that can connect to the internet. The other options require installation of additional software, which is not permitted in this scenario.
upvoted 10 times

Learner_Thunder2 Highly Voted 4 years, 7 months ago
Azure Notebooks is correct
upvoted 5 times

sar77 Most Recent 1 year, 1 month ago
Azure Notebooks is correct
upvoted 1 times

raidenstrike1945 1 year, 8 months ago
Why C? there's no Internet, no Azure Subscription. How ppl can use Azure Notebook without these two?
upvoted 1 times

james2033 1 year, 10 months ago
Learner devices NOT configured for Python development --> Azure Notebooks. Choose C. Quote 'Since you don't share compute instances, other users who run your notebook will do so on their own compute instance.' at https://learn.microsoft.com/en-us/azure/machine-learning/how-to-run-jupyter-notebooks?view=azureml-api-2#share-a-notebook
upvoted 1 times

evangelist 1 year, 10 months ago
The correct answer is C. You should consider configuring the use of Azure Notebooks. Explanation: Azure Notebooks is a free service that provides Jupyter notebooks that are hosted in the cloud. The service allows users to write and execute Python (among other languages) code in their browsers without the need for any configuration or installation of software on the local machine. This makes it an ideal solution for scenarios where learners do not have administrator permissions to install software and cannot access Azure subscriptions directly. Azure Notebooks support data visualization libraries in Python, such as Matplotlib, Seaborn, or Plotly, making it suitable for the practical training aimed at acquainting learners with data visualization creation.
upvoted 2 times

sl_mslconsulting 2 years, 2 months ago
Seems Azure Notebooks is no longer available.
upvoted 1 times

kay1101 2 years, 3 months ago
I understood C may be the best choice here, however, how could you use azure notebooks without subscription? notebook prerequisites: https://learn.microsoft.com/en-us/azure/machine-learning/how-to-run-jupyter-notebooks?view=azureml-api-2
upvoted 2 times

MarinaMijailovic 3 years, 1 month ago

I dont understand why does it say they are unable to access Azure subscriptions but they can use some Azure services like Azure notebooks. Can someone please explain?

upvoted 3 times

daviduzo 3 years, 1 month ago

Other options require you to install additional software which the no admin access disallows

upvoted 1 times

synapse 4 years, 5 months ago

Answer is correct

upvoted 3 times

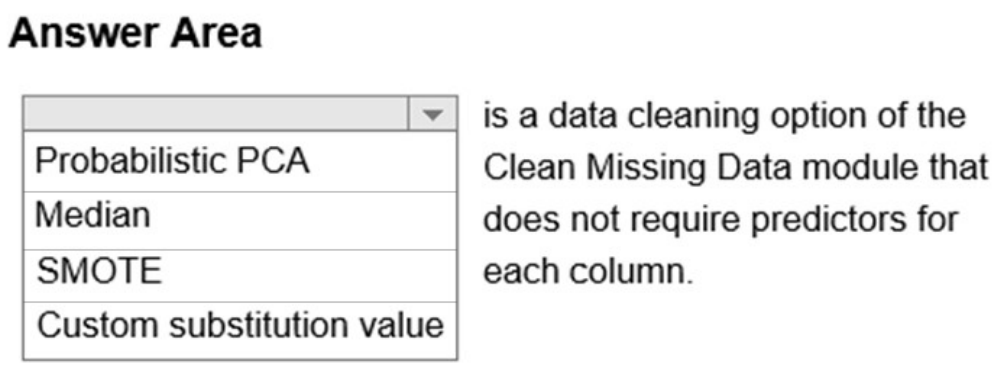
txu002 4 years, 5 months ago

Configuring the use of Azure Kubernetes Service

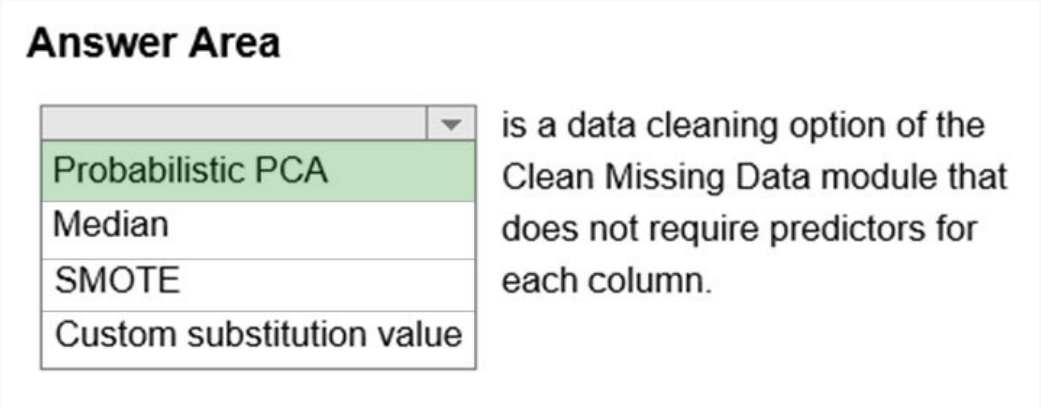
upvoted 1 times

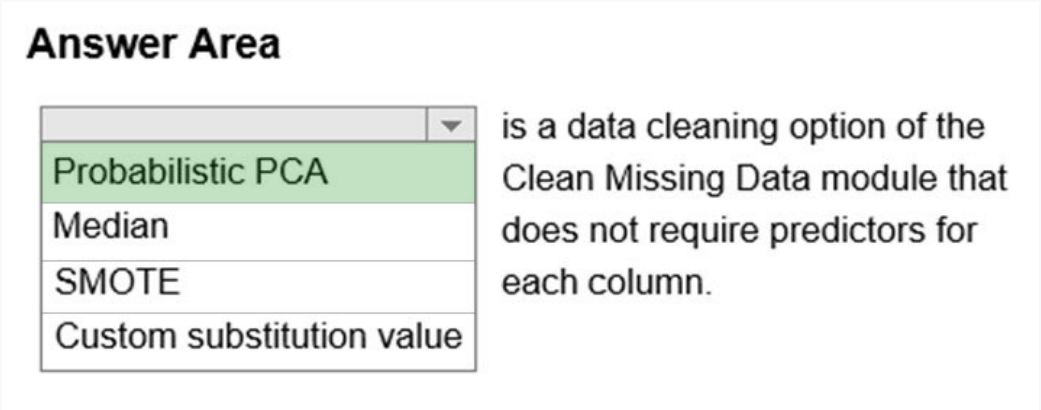
Question #21 *Topic 1*

HOTSPOT -
Complete the sentence by selecting the correct option in the answer area.
Hot Area:



Suggested Answer:





Replace using Probabilistic PCA: Compared to other options, such as Multiple Imputation using Chained Equations (MICE), this option has the advantage of not requiring the application of predictors for each column. Instead, it approximates the covariance for the full dataset. Therefore, it might offer better performance for datasets that have missing values in many columns. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

pancman Highly Voted 4 years, 4 months ago
I don't think that this is a real exam question. Median and custom substitution techniques don't require a predictor either.
upvoted 20 times

rishi_ram Highly Voted 3 years, 2 months ago
Replace using Probabilistic PCA: Replaces the missing values by using a linear model that analyzes the correlations between the columns and estimates a low-dimensional approximation of the data, from which the full data is reconstructed. The underlying dimensionality reduction is a

probabilistic form of Principal Component Analysis (PCA), and it implements a variant of the model proposed in the Journal of the Royal Statistical Society, Series B 21(3), 611–622 by Tipping and Bishop. Compared to other options, such as Multiple Imputation using Chained Equations (MICE), this option has the advantage of not requiring the application of predictors for each column. Instead, it approximates the covariance for the full dataset. Therefore, it might offer better performance for datasets that have missing values in many columns. https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/clean-missing-data

upvoted 5 times

Guru_1 Most Recent 8 months, 1 week ago

The are the only valid options (Up-to-date) *Median *Custom substitution value But, "Probabilistic PCA" which is outdated along with Machine Learning Studio (classic) which also the Support ended on 31 August 2024 by the Microsoft. In Azure Machine Learning Studio also is not showing the option "Probabilistic PCA". But, Still this option is opted as answer. Either is this valid questions and really a part of the exam ???

upvoted 2 times

geethavkr 2 years ago

correct.. Outdated but in previous versions it says Replace using Probabilistic PCA: Replaces the missing values by using a linear model that analyzes the correlations between the columns and estimates a low-dimensional approximation of the data, from which the full data is reconstructed. The underlying dimensionality reduction is a probabilistic form of Principal Component Analysis (PCA), and it implements a variant of the model proposed in the Journal of the Royal Statistical Society, Series B 21(3), 611–622 by Tipping and Bishop. Compared to other options, such as Multiple Imputation using Chained Equations (MICE), this option has the advantage of not requiring the application of predictors for each column. Instead, it approximates the covariance for the full dataset. Therefore, it might offer better performance for datasets that have missing values in many columns.

upvoted 1 times

kay1101 2 years, 2 months ago

I think this is an outdated question. as of may 2024, PCA is no longer in the clean missing data module. reference: https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/clean-missing-data?view=azureml-api-2 however, in the past, PCA did in the clean missing data module. reference:https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/clean-missing-data at the time of the question was created, PCA may be correct. but now, i thick is either median or custom substitution value.

upvoted 1 times

InversaRadice 2 years, 8 months ago

answer is 100% correct ... Replace using Probabilistic PCA: ... Compared to other options, such as Multiple Imputation using Chained Equations (MICE), this option has the advantage of not requiring the application of predictors for each column. https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/clean-missing-data

upvoted 2 times

eternaleclipse 2 years, 10 months ago

What pancman said. outdated question

upvoted 1 times

IvanTT 2 years, 10 months ago

It can't be "A. Probabilistic PCA" because it isn't an option for the Clean Missing Data module. Here is the reference: https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/clean-missing-data?view=azureml-api-2 It could be "D. Custom Substitution Value". The option "B. Median" isn't the exact option for the module which it's "Replace with median".

upvoted 1 times

james2033 2 years, 10 months ago

Qutote "Replace using Probabilistic PCA: Replaces the missing values by using a linear model that analyzes the correlations between the columns and estimates a low-dimensional approximation of the data, from which the full data is reconstructed. The underlying dimensionality reduction is a probabilistic form of Principal Component Analysis (PCA), and it implements a variant of the model proposed in the Journal of the Royal Statistical Society, Series B 21(3), 611–622 by Tipping and Bishop. Compared to other options, such as Multiple Imputation using Chained Equations (MICE), this option has the advantage of not requiring the application of predictors for each column." Reference https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/clean-missing-data#:~:text=this%20option%20has%20the%20advantage%20of%20not%20requiring%20the%20application%20of%20predictors%20for%20each%20column.

upvoted 1 times

rakeshmk 2 years, 10 months ago

PCA is a dimensionality reduction technique.. Median can be the answer

upvoted 3 times

PradhanManva 2 years, 10 months ago

PCA -This is the answer.

upvoted 1 times

MarinaMijailovic 3 years, 3 months ago

Correct answer is medain - it only calulates the medain from the given column, no other columns required pca - needs predictors to calculate the probabilities smote - needs predictors to generate synthetic samples for the minority class csv - doesn't really need predictors per se, but still requires some knoweldge about the data to pick the right value

upvoted 3 times

Truman 3 years, 4 months ago

One data cleaning option that does not require predictors for each column in the Clean Missing Data module is the "Replace with mean" option. This option replaces missing values in a column with the mean of the available values in that column All these options are false

upvoted 1 times

Vic9 3 years, 4 months ago

A https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/clean-missing-data "Replace using Probabilistic PCA: Replaces the missing values by using a linear model that analyzes the correlations between the columns and estimates a low-dimensional approximation of the data, from which the full data is reconstructed. The underlying dimensionality reduction is a probabilistic form of Principal Component Analysis (PCA), and it implements a variant of the model proposed in the Journal of the Royal Statistical Society, Series B 21(3), 611–622 by Tipping and Bishop. Compared to other options, such as Multiple Imputation using Chained Equations (MICE), this option has the advantage of not requiring the application of predictors for each column. Instead, it approximates the covariance for the full dataset. Therefore, it might offer better performance for datasets that have missing values in many columns."

upvoted 2 times

phdykd 3 years, 5 months ago

A) Probabilistic PCA and C) SMOTE are not data cleaning options in the clean missing data module. Probabilistic PCA is a technique used for dimensionality reduction and feature extraction in machine learning, and it is not specifically designed to handle missing data. SMOTE (Synthetic Minority Over-sampling Technique) is a technique used for dealing with imbalanced datasets in machine learning, and it is not designed to handle missing data. Therefore, the correct answer to the question "..... is a data cleaning option of the clean missing data module that does not require predictors for each column" is either B) Median or D) Custom substitution value.

upvoted 2 times

Peeking 3 years, 6 months ago

PCA is wrong.

upvoted 2 times

ranjsi01 4 years, 6 months ago

correct

upvoted 3 times

Question #22 *Topic 1*

You have recently concluded the construction of a binary classification machine learning model.
You are currently assessing the model. You want to make use of a visualization that allows for precision to be used as the measurement for the assessment.
Which of the following actions should you take?

A. You should consider using Venn diagram visualization.

B. You should consider using Receiver Operating Characteristic (ROC) curve visualization.

C. You should consider using Box plot visualization.

D. You should consider using the Binary classification confusion matrix visualization.

Suggested Answer: *D*
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-understand-automated-ml#confusion-matrix

Community vote distribution
D (100%)

spaceykacey Highly Voted 4 years, 3 months ago
you cannot visualize precision with ROC. True Positive Rate(on ROC's y axis) = Recall. Not precision. PR curve is used to visualize precision. I think I would go with Confusion matrix even though it requires further computations
upvoted 16 times

evangelist Most Recent 1 year, 12 months ago
it's important to note that while ROC is immensely useful for evaluating and comparing models, it doesn't directly display precision.
upvoted 2 times

jdada 2 years, 3 months ago
D. You should consider using the Binary classification confusion matrix visualization.
upvoted 1 times

james2033 2 years, 4 months ago
Based on https://learn.microsoft.com/en-us/azure/machine-learning/how-to-understand-automated-ml?view=azureml-api-2#binary-vs-multiclass-classification-metrics . Visualization for Classification: Not has Venn diagram; box plot diagram. Consider ROC (Receiver Operating Characteristic) and Binary classification confusion matrix. 'The receiver operating characteristic (ROC) curve plots the relationship between true positive rate (TPR) and false positive rate (FPR) as the decision threshold changes.' https://learn.microsoft.com/en-us/azure/machine-learning/how-to-understand-automated-ml?view=azureml-api-2#confusion-matrix Confusion matrix has 'precision to be used as the measurement', but Receiver operating characteristic has not.
upvoted 1 times

PradhanManva 2 years, 4 months ago
This is the answer.
upvoted 1 times

SoftAI 2 years, 10 months ago
classification confusion matrix is the best accuracy measure
upvoted 2 times

ZIMARAKI 2 years, 11 months ago
D for precision
upvoted 2 times

KarthikKumarK 3 years ago
Correct. https://builtin.com/data-science/precision-and-recall
upvoted 2 times

clark88 3 years, 1 month ago
precision, recall, f1-score. Are part of the confusion matrix, I agree that this answer is correct.
upvoted 3 times

VJPrakash 4 years, 6 months ago

Shouldnt the answer be "B". ROC is a graph against TPR and FPR.. precision could be clearly visualized.we can plot confusion matrix as gridYou can get the precision number without any further calculations in a confusion matrix. ROC curve shows True Positive vs. False Positive. But it doesn't show precision.

upvoted 4 times

VJPrakash 4 years, 6 months ago

Shouldnt the answer be "B". ROC is a graph against TPR and FPR.. precision could be clearly visualized.

upvoted 4 times

dijaa Highly Voted 4 years, 5 months ago

we can plot confusion matrix as grid

upvoted 5 times

pancman 3 years, 10 months ago

You can get the precision number without any further calculations in a confusion matrix. ROC curve shows True Positive vs. False Positive. But it doesn't show precision.

upvoted 4 times

Question #23 *Topic 1*

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You have been tasked with evaluating your model on a partial data sample via k-fold cross-validation.
You have already configured a k parameter as the number of splits. You now have to configure the k parameter for the cross-validation with the usual value choice.
Recommendation: You configure the use of the value k=1.
Will the requirements be satisfied?

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Community vote distribution
B (100%)

astralyt Highly Voted 4 years, 6 months ago
k can be 5 or 10
upvoted 9 times

astralyt Highly Voted 4 years, 6 months ago
k can be 5 or 10More specifically, k MUST be within: 1 < k <= N
upvoted 9 times

sim39 4 years, 5 months ago
More specifically, k MUST be within: 1 < k <= N
upvoted 3 times

evangelist Most Recent 1 year, 12 months ago
The usual value choice for k in k-fold cross-validation typically ranges from 5 to 10, with k=10 being a common default. These values are chosen because they provide a good balance between training on enough data points and having a sufficient number of iterations to get a reliable estimate of the model's performance.
upvoted 2 times

james2033 2 years, 4 months ago
Default value k = 10.
upvoted 1 times

pratham1612 2 years, 8 months ago
No is correct
upvoted 1 times

Peeking 3 years, 1 month ago
'Usual k' = 5 or 10.
upvoted 2 times

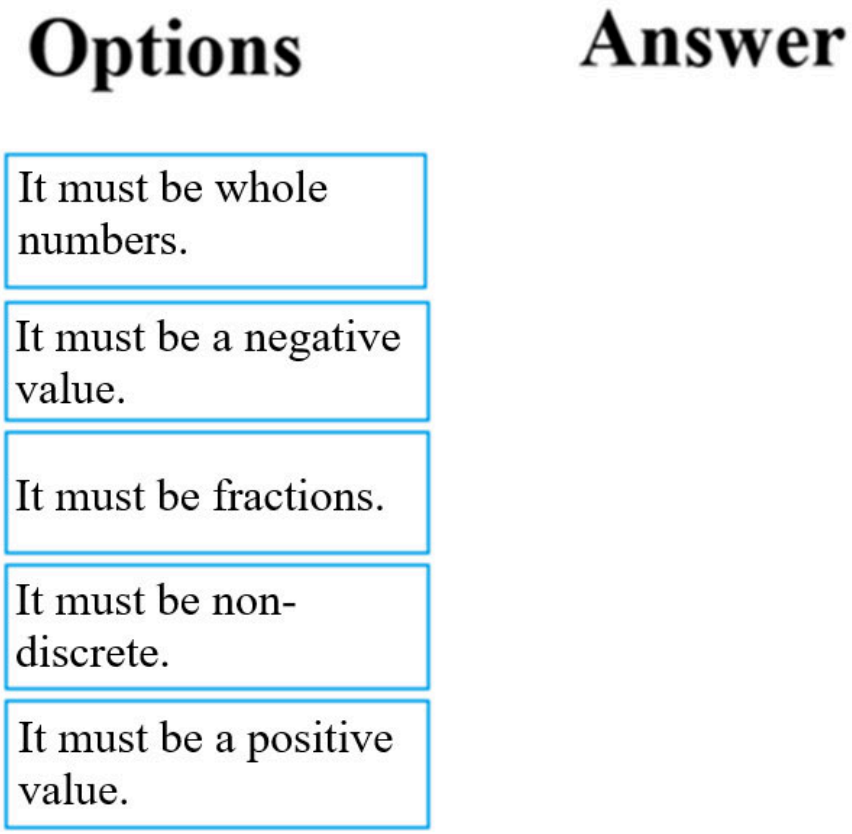
DRAG DROP -

You are in the process of constructing a regression model.

You would like to make it a Poisson regression model. To achieve your goal, the feature values need to meet certain conditions.

Which of the following are relevant conditions with regards to the label data? Answer by dragging the correct options from the list to the answer area.

Select and Place:



Suggested Answer:

Options	Answer
It must be whole numbers.	It must be whole numbers.
It must be a negative value.	It must be a positive value.
It must be fractions.	
It must be non-discrete.	
It must be a positive value.	

Options	Answer
It must be whole numbers.	It must be whole numbers.
It must be a negative value.	It must be a positive value.
It must be fractions.	
It must be non-discrete.	
It must be a positive value.	

Poisson regression is intended for use in regression models that are used to predict numeric values, typically counts. Therefore, you should use this module to create your regression model only if the values you are trying to predict fit the following conditions: ✑ The response variable has a Poisson distribution. ✑ Counts cannot be negative. The method will fail outright if you attempt to use it with negative labels. ✑ A Poisson

distribution is a discrete distribution; therefore, it is not meaningful to use this method with non-whole numbers. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/poisson-regression

pancman Highly Voted 3 years, 10 months ago

Here are the requirements of the Poisson regression: The response variable has a Poisson distribution. Counts cannot be negative. The method will fail outright if you attempt to use it with negative labels. A Poisson distribution is a discrete distribution; therefore, it is not meaningful to use this method with non-whole numbers.

upvoted 13 times

pancman Highly Voted 3 years, 10 months ago

Answer is correct

upvoted 7 times

evangelist Most Recent 1 year, 12 months ago

For constructing a Poisson regression model, the label (dependent variable) must meet specific conditions due to the nature of Poisson distribution. Poisson regression is used for modeling count data and the events occurrence in a fixed interval of time or space. The correct conditions for the label data in a Poisson regression model are: Explanation: It must be whole numbers: Poisson regression models the count of occurrences of an event, thus the dependent variable must indeed be whole numbers (integers) since you cannot have a fraction of an occurrence. It must be a positive value: More precisely, the values must be non-negative integers (including zero). Poisson regression is suitable for counts which cannot be negative.

upvoted 1 times

james2033 2 years, 4 months ago

* It must be whole numbers. * It must be a positive value.

upvoted 3 times

phydev 2 years, 7 months ago

Yep, Poisson regression do be like that.

upvoted 3 times

Question #25 *Topic 1*

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You are in the process of carrying out feature engineering on a dataset.
You want to add a feature to the dataset and fill the column value.
Recommendation: You must make use of the Group Categorical Values Azure Machine Learning Studio module.
Will the requirements be satisfied?

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Community vote distribution



ppchar Highly Voted 4 years, 9 months ago
The typical use for grouping categorical values is to merge multiple string values into a single new level.
upvoted 18 times

Xsesi 2 years ago
Group Categorical Values Module is used to group similar categorical values into a single category, which is useful for reducing the number of categories or combining similar categories instead of adding new features or filling column values with specific data.
upvoted 1 times

microscape 2 years, 2 months ago
Selected Answer: A
For Output mode, indicate whether you want to output just the new levels, or append the changes to see the original column, with the replacements side by side.

The default, ResultOnly, shows only the new values. The Inplace option replaces the existing column values with the new levels.
https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/group-categorical-values?redirectedfrom=MSDN
upvoted 1 times

microscape 2 years, 2 months ago
For Output mode, indicate whether you want to output just the new levels, or append the changes to see the original column, with the replacements side by side. The default, ResultOnly, shows only the new values. The Inplace option replaces the existing column values with the new levels.
https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/group-categorical-values?redirectedfrom=MSDNbased on this link the answer is B. No. 'Group Categorical Values' can be used to create new columns by combining multiple string values to a new value/column; however it cannot be used for columns of numeric type or columns designated as labels or FEATURES.
upvoted 1 times

Ben999 Most Recent 1 year, 7 months ago
based on this link the answer is B. No. 'Group Categorical Values' can be used to create new columns by combining multiple string values to a new value/column; however it cannot be used for columns of numeric type or columns designated as labels or FEATURES.
upvoted 1 times

evangelist 2 years, 5 months ago
The "Group Categorical Values" module in Azure Machine Learning Studio is designed to reduce the granularity of categorical data. It groups together categorical values that appear infrequently in the data, which can be beneficial for models that may not handle high-cardinality categorical data well. This module is useful for managing and simplifying categorical features but does not directly address the task of adding a new feature and filling its values.
upvoted 4 times

GHill1982 2 years, 6 months ago
The answer is no, the recommendation is incorrect. The Group Categorical Values Azure Machine Learning Studio module is designed to group a set of categorical values of a categorical feature into a smaller number of groups. It is not suitable for adding a new feature to the dataset or filling the

column value.

upvoted 2 times

james2033 2 years, 10 months ago

The question is Out of date. Don't need care about this question. https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/group-categorical-values?redirectedfrom=MSDN

upvoted 3 times

PradhanManva 2 years, 10 months ago

This is the answer.

upvoted 1 times

PopeyeDS 3 years, 1 month ago

Extract from website: Let's assume that you want to group cars in the Automobile price dataset by engine size, using the number of cylinders. Rather than lots of different engine sizes, you will create the new levels, "big", "small", and "other" as follows: Big engines: six cylinders or larger Small engines: two or four cylinders Other: anything else I think A is correct

upvoted 1 times

Ammy_b 3 years, 6 months ago

Changing the string values into numerical by group categorial is also one of the feature engineering but too many columns might be added while processing, leading to an error. A is Correct

upvoted 2 times

Edriv 3 years, 8 months ago

https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/group-categorical-values?redirectedfrom=MSDN

upvoted 1 times

Question #26 *Topic 1*

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You are in the process of carrying out feature engineering on a dataset.
You want to add a feature to the dataset and fill the column value.
Recommendation: You must make use of the Join Data Azure Machine Learning Studio module.
Will the requirements be satisfied?

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Community vote distribution



gnsu Highly Voted 4 years, 3 months ago
No is Correct answer. Add Columns need to be used. Join data is needed only for database style joins https://docs.microsoft.com/bs-cyrl-ba/azure/machine-learning/component-reference/add-columns https://docs.microsoft.com/bs-cyrl-ba/azure/machine-learning/component-reference/join-data
upvoted 14 times

evangelist Most Recent 1 year, 12 months ago
The "Join Data" module in Azure Machine Learning Studio allows you to merge two datasets based on a common key or keys. This is particularly useful when you want to augment your existing dataset with additional information from another source. For example, if you have a dataset of sales transactions and another dataset with product information, you could use the "Join Data" module to add product details (a new feature) to the sales transactions dataset by joining on a product ID column present in both datasets.
upvoted 2 times

james2033 2 years, 4 months ago
The 'Join Data' module in Microsoft Azure Machine Learning Studio (classic) is used to merge two datasets using a database-style join operation. If you want to add a feature to the dataset and FILL COLUMN VALUE --> No. Join dataset https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/join-data?view=azureml-api-2
upvoted 1 times

PradhanManva 2 years, 4 months ago
This is the answer.
upvoted 1 times

Sa_Msa 2 years, 7 months ago
Answer is correct. I tested it multiple times and got error regarding the size. You need to use "add columns" if you want to add two dataset together.
upvoted 2 times

SoftAI 2 years, 10 months ago
Join Data Azure Machine Learning Studio is a data transformation tool
upvoted 1 times

mhmichiel 2 years, 11 months ago
If you want to add a colom with a new feature you will need data. This data could be in a secondery dataset which needs to be joined to the original dataset via a join. A could be correct. Concidering that it would be B, how would I fill the column? Manualy? That sounds like a unlogical option concidering that the datasets are usualy pretty big and not made up.
upvoted 2 times

michaelmorar 3 years ago
JOIN implies a query, not inserting a column to a database.
upvoted 3 times

Sibajene 3 years, 1 month ago
B. No The Join Data module in Azure Machine Learning Studio is used to combine two datasets by matching values in key columns. It is not a general-purpose tool for adding features or filling column values in a single dataset.

upvoted 2 times

Edriv 3 years, 2 months ago

Option B

upvoted 1 times

KIshor1212 3 years, 2 months ago

To perform a join on two datasets, they should be related by a key column. Composite keys using multiple columns are also supported. Add the datasets you want to combine, and then drag the Join Data component into your pipeline.The question says 'add a feature to the dataset ' ... there are no 2 datasets here. You are not joining them... You are adding a damn single feature to the dataset. Why you guys like to mislead people?You are right, however, how do you want to add the feature to the dataset withoud data? Concidering that you want to add a feature to a dataset and that you dont want to make it up you need to ad a coulomn representing the new feature. To add this colomn you need a secondary dataset which you join to the main dataset. This means that A is not incorect. The question is incorrect and the answer is not nesseserely b.

upvoted 1 times

KIshor1212 3 years, 2 months ago

Selected Answer: A

To perform a join on two datasets, they should be related by a key column. Composite keys using multiple columns are also supported. Add the datasets you want to combine, and then drag the Join Data component into your pipeline.

upvoted 1 times

Mirjalol Highly Voted 3 years ago

The question says 'add a feature to the dataset ' ... there are no 2 datasets here. You are not joining them... You are adding a damn single feature to the dataset. Why you guys like to mislead people?

upvoted 5 times

Mirjalol 3 years ago

The question says 'add a feature to the dataset ' ... there are no 2 datasets here. You are not joining them... You are adding a damn single feature to the dataset. Why you guys like to mislead people?You are right, however, how do you want to add the feature to the dataset withoud data? Concidering that you want to add a feature to a dataset and that you dont want to make it up you need to ad a coulomn representing the new feature. To add this colomn you need a secondary dataset which you join to the main dataset. This means that A is not incorect. The question is incorrect and the answer is not nesseserely b.

upvoted 5 times

mhmichiel 2 years, 11 months ago

You are right, however, how do you want to add the feature to the dataset withoud data? Concidering that you want to add a feature to a dataset and that you dont want to make it up you need to ad a coulomn representing the new feature. To add this colomn you need a secondary dataset which you join to the main dataset. This means that A is not incorect. The question is incorrect and the answer is not nesseserely b.

upvoted 1 times

dynamic_person91 3 years, 5 months ago

join is used for datasets not for columns b is correct

upvoted 1 times

ning 3 years, 8 months ago

Selected Answer: A

This is possible

upvoted 1 times

ning 3 years, 8 months ago

This is possibleThere is an assumption that column is in another dataset with a key to map to the base dataset

upvoted 1 times

ning 3 years, 8 months ago

There is an assumption that column is in another dataset with a key to map to the base dataset

upvoted 1 times

synapse 3 years, 11 months ago

Selected Answer: B

Don't use Join Data Use Add Columns

upvoted 2 times

synapse 3 years, 11 months ago

Don't use Join Data Use Add ColumnsAdd columns is definitely incorrect! 1. it is cross join, not link data with key column 2. it will add all columns from both data set Join is the correct answer, you can control which key column to join and which columns are from left dataset and which columns are from right data set

upvoted 2 times

ning 3 years, 8 months ago

Add columns is definitely incorrect! 1. it is cross join, not link data with key column 2. it will add all columns from both data set Join is the correct answer, you can control which key column to join and which columns are from left dataset and which columns are from right data set

upvoted 2 times

Tsardoz 4 years, 1 month ago

I think Yes is correct. The second dataset might be a key which is joined with a second column being the extra column to add to the dataset.

upvoted 4 times

Question #27 *Topic 1*

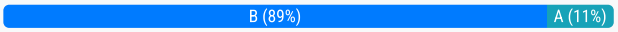
This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You are in the process of carrying out feature engineering on a dataset.
You want to add a feature to the dataset and fill the column value.
Recommendation: You must make use of the Edit Metadata Azure Machine Learning Studio module.
Will the requirements be satisfied?

A. Yes

B. No

Suggested Answer: *A*
Typical metadata changes might include marking columns as features. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/edit-metadata https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/join-data https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-categorical-values

Community vote distribution



gaint Highly Voted 4 years, 7 months ago
Edit meta data cannot add a new column, it can change properties of existing columns
upvoted 30 times

gaint Highly Voted 4 years, 7 months ago
Edit meta data cannot add a new column, it can change properties of existing columnsAgree. New column can be added through SQL transformation or python scriptExactly!question says fill column, not add"You want to add a feature to the dataset and fill the column value."
upvoted 30 times

azure1000 4 years, 6 months ago
Agree. New column can be added through SQL transformation or python script
upvoted 4 times

azure1000 4 years, 6 months ago
Agree. New column can be added through SQL transformation or python scriptExactly!
upvoted 4 times

pancman 3 years, 10 months ago
Exactly!
upvoted 1 times

Gabonia 3 years, 6 months ago
question says fill column, not add"You want to add a feature to the dataset and fill the column value."
upvoted 1 times

Gabonia 3 years, 6 months ago
question says fill column, not add
upvoted 1 times

lcgcastro96 2 years, 8 months ago
"You want to add a feature to the dataset and fill the column value."
upvoted 1 times

evangelist Most Recent 1 year, 12 months ago
filling value has nothing to do with Edit metadata
upvoted 1 times

PopeyeDS 2 years, 7 months ago
Edit metadata helps to change column type to categorical or any other.. not necessary to add a feature
upvoted 1 times

Sa_Msa 2 years, 7 months ago

"Edit Meta Data" as the name is saying, is for editing the meta data, you can select the columns that you want to change, you can change their type e.g. numerical to categorial, then you can choose if they are features or labels etc.. you can also change the column name here too. Below is the options you have: Column Edit column [Column names: test2] Data type [Categorical] Fields [Label] New column names -

upvoted 1 times

Nadine_nm 2 years, 8 months ago

The answer is correct, the module allows you to : Mark columns as features. https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/edit-metadata?view=azureml-api-2

upvoted 1 times

eloyinaay 2 years, 11 months ago

To add a column the only way is with python scripts.

upvoted 1 times

mamau 3 years ago

YES. The recommendation to use the Edit Metadata Azure Machine Learning Studio module would satisfy the requirements of adding a feature to the dataset and filling the column value. The Edit Metadata module allows for editing of the metadata of the dataset, including adding columns, modifying column names, and changing data types. It would be an appropriate tool for adding a new feature to the dataset and filling its values.

upvoted 2 times

mamau 3 years ago

YES. The recommendation to use the Edit Metadata Azure Machine Learning Studio module would satisfy the requirements of adding a feature to the dataset and filling the column value. The Edit Metadata module allows for editing of the metadata of the dataset, including adding columns, modifying column names, and changing data types. It would be an appropriate tool for adding a new feature to the dataset and filling its values.mamau, could you bring a proof please? a link...

upvoted 2 times

Yuriy_Ch 2 years, 11 months ago

mamau, could you bring a proof please? a link...

upvoted 2 times

Edriv 3 years, 2 months ago

Keyword => Add

upvoted 1 times

synapse 3 years, 11 months ago

Can't add new columns. Change existing ones

upvoted 2 times

Tsardoz 4 years, 1 month ago

Edit metadata can change existing column names not add new ones

upvoted 3 times

dija123 4 years, 2 months ago

Selected Answer: A

Agree with Edit metadata

upvoted 1 times

dija123 4 years, 2 months ago

Agree with Edit metadataTry it in Azure ML Studio and you will see that you can't add a new column using "Edit Metadata"

upvoted 1 times

pancman 3 years, 10 months ago

Try it in Azure ML Studio and you will see that you can't add a new column using "Edit Metadata"

upvoted 2 times

gnsu 4 years, 3 months ago

Given answer is wrong: Should be No. Add Columns must be right answer. https://docs.microsoft.com/en-us/azure/machine-learning/component-reference/add-columns

upvoted 1 times

bdsrca 4 years, 5 months ago

The answer is correct, you can ADD a new column by editing metadata. ref: https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/edit-metadataNo, For New column names, enter the new name of the selected column or columns. Not for adding new column..I can only find to give a new name to the existing columns. But not adding a completely new columnThanks for this link.Try it in Azure ML Studio yourself and you will see that you can't add a new column using "Edit Metadata". Please don't spread false information unless you're sure!

upvoted 4 times

bdsrca 4 years, 5 months ago

The answer is correct, you can ADD a new column by editing metadata. ref: https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/edit-metadata

upvoted 4 times

snsnsnsn 4 years, 5 months ago

No, For New column names, enter the new name of the selected column or columns. Not for adding new column..

upvoted 2 times

dushmantha 4 years, 5 months ago

I can only find to give a new name to the existing columns. But not adding a completely new column

upvoted 2 times

sapna_tare 2 years, 10 months ago

Thanks for this link.

upvoted 1 times

pancman 3 years, 10 months ago

Try it in Azure ML Studio yourself and you will see that you can't add a new column using "Edit Metadata". Please don't spread false information unless you're sure!

upvoted 3 times

Question #28 *Topic 1*

You have been tasked with ascertaining if two sets of data differ considerably. You will make use of Azure Machine Learning Studio to complete your task.
You plan to perform a paired t-test.
Which of the following are conditions that must apply to use a paired t-test? (Choose all that apply.)

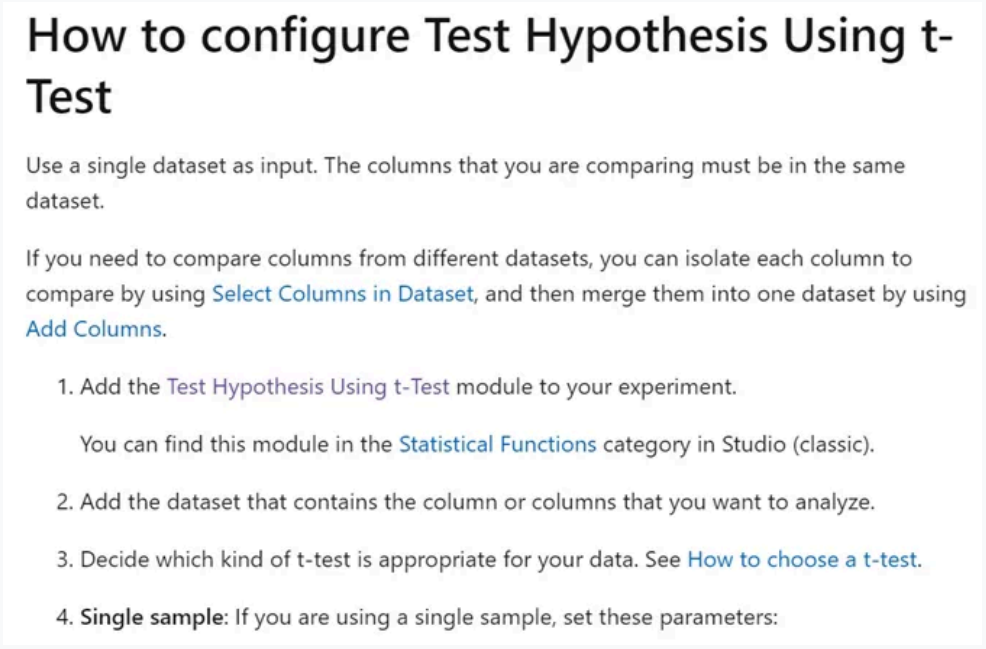
A. All scores are independent from each other.

B. You have a matched pairs of scores.

C. The sampling distribution of d is normal.

D. The sampling distribution of x1- x2 is normal.

Suggested Answer: *B, C*



How to configure Test Hypothesis Using t-Test

Use a single dataset as input. The columns that you are comparing must be in the same dataset.

If you need to compare columns from different datasets, you can isolate each column to compare by using Select Columns in Dataset, and then merge them into one dataset by using Add Columns.

1. Add the Test Hypothesis Using t-Test module to your experiment.

   You can find this module in the Statistical Functions category in Studio (classic).

2. Add the dataset that contains the column or columns that you want to analyze.

3. Decide which kind of t-test is appropriate for your data. See How to choose a t-test.

4. **Single sample**: If you are using a single sample, set these parameters:

How to configure Test Hypothesis Using t-Test

Use a single dataset as input. The columns that you are comparing must be in the same dataset.

If you need to compare columns from different datasets, you can isolate each column to compare by using Select Columns in Dataset, and then merge them into one dataset by using Add Columns.

1. Add the Test Hypothesis Using t-Test module to your experiment.

   You can find this module in the Statistical Functions category in Studio (classic).

2. Add the dataset that contains the column or columns that you want to analyze.

3. Decide which kind of t-test is appropriate for your data. See How to choose a t-test.

4. **Single sample**: If you are using a single sample, set these parameters:

Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/test-hypothesis-using-t-test

Community vote distribution

BC (90%) (5% (5%

lookaaaa Highly Voted 2 years, 8 months ago

A is for single sample t-test, D is for unpaired t-test, BC are for paired t-test

upvoted 10 times

jdada Highly Voted 1 year, 9 months ago

B. You have a matched pairs of scores. C. The sampling distribution of d is normal. D. The sampling distribution of x1- x2 is normal.

upvoted 6 times

amittal09 Most Recent 1 year, 3 months ago

https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/test-hypothesis-using-t-test#bkmk_HowToChoose

upvoted 1 times

MarinaMijailovic 2 years, 1 month ago

C and D refer to the same thing

upvoted 1 times

Ram19830916 2 years, 2 months ago

A<B<C Choose a paired t-test when these conditions apply: You have a matched pairs of scores. For example, you might have two different measures per person, or matched pairs of individuals (such as a husband and wife). Each pair of scores is independent of every other pair. The sampling distribution of d is normal.

upvoted 1 times

MarinaMijailovic 2 years, 5 months ago

Can someone explain what is d and what is x1-x2

upvoted 3 times

WinnieS 2 years, 6 months ago

B and C are correct. A should change to EACH PAIR of scores is independent of other pairs, not ALL scores. D should change to C

upvoted 2 times

meysa 2 years, 7 months ago

B and D are correct.

upvoted 2 times

KIshor1212 2 years, 8 months ago

Choose a paired t-test when these conditions apply: ***---You have a matched pairs of scores. For example, you might have two different measures per person, or matched pairs of individuals (such as a husband and wife). Each pair of scores is independent of every other pair. *** --The sampling distribution of d is normal.

upvoted 4 times

slcheng 2 years, 9 months ago

Choose a paired t-test when these conditions apply: You have a matched pairs of scores. For example, you might have two different measures per person, or matched pairs of individuals (such as a husband and wife). Each pair of scores is independent of every other pair. The sampling distribution of d is normal.

upvoted 1 times

JTWang 2 years, 10 months ago

BC is the correct answer.

upvoted 2 times

Vuuma 2 years, 11 months ago

Choose a paired t-test when these conditions apply: 1. You have a matched pairs of scores. For example, you might have two different measures per person, or matched pairs of individuals (such as a husband and wife). 2. Each pair of scores is independent of every other pair. 3.The sampling distribution of d is normal.

upvoted 2 times

Vuuma 2 years, 11 months ago

Correct answer is BC

upvoted 1 times

Chimman 3 years, 5 months ago

Clearly mentioned that it should be ABC as per the documentation. https://docs.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/test-hypothesis-using-t-test

upvoted 4 times

TheCyanideLancer 3 years, 6 months ago

As per MS documentation ans is ABC https://docs.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/test-hypothesis-using-t-test

upvoted 4 times

Tsardoz 3 years, 7 months ago

Answer should be ABD not sure why it just chose AB

upvoted 2 times

Tsardoz 3 years, 7 months ago

Leaving out c because 'd' is not defined. https://www.jmp.com/en_au/statistics-knowledge-portal/t-test/paired-t-test.html#:~:text=Paired%20t%2Dtest%20assumptions&text=Subjects%20must%20be%20independent.,obtained%20from%20the%20same%20subject.&text=The%20measured%20differences%20are%20normally%20distributed.

upvoted 1 times

Question #29 *Topic 1*

You want to train a classification model using data located in a comma-separated values (CSV) file.
The classification model will be trained via the Automated Machine Learning interface using the Classification task type.
You have been informed that only linear models need to be assessed by the Automated Machine Learning.
Which of the following actions should you take?

A. You should disable deep learning.

B. You should enable automatic featurization.

C. You should disable automatic featurization.

D. You should set the task type to Forecasting.

Suggested Answer: *C*
Reference: https://econml.azurewebsites.net/spec/estimation/dml.html https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-automated-ml-for-ml-models

Community vote distribution
A (100%)

gaint Highly Voted 4 years, 7 months ago
Answer should be ADisabling automatic featurization does not cause the models evaluated to be linear only. Automatic featurization has the following steps listed https://docs.microsoft.com/en-us/azure/machine-learning/how-to-configure-auto-features#automatic-featurization The only way to force linear algorithms to be evaluated is to use the blocked algorithms. list.
upvoted 22 times

gaint Highly Voted 4 years, 7 months ago
Answer should be A
upvoted 22 times

lander_c 4 years, 4 months ago
Disabling automatic featurization does not cause the models evaluated to be linear only. Automatic featurization has the following steps listed https://docs.microsoft.com/en-us/azure/machine-learning/how-to-configure-auto-features#automatic-featurization The only way to force linear algorithms to be evaluated is to use the blocked algorithms. list.
upvoted 6 times

phydev 2 years, 7 months ago
This was on Exam today (20 July 2023). There was another option to Block all the other algorithms except the linear ones, which I chose as my answer.
upvoted 21 times

amittal09 Most Recent 1 year, 3 months ago
The goal is to assess only linear models. Deep learning models are non-linear and complex. By disabling deep learning algorithms in the Automated Machine Learning settings, you ensure that only linear models (like Logistic Regression, Linear SVM, etc.) are considered during the model selection process.
upvoted 2 times

deyoz 1 year, 11 months ago
I go for A
upvoted 1 times

NullVoider_0 2 years, 2 months ago
AutoML tries different models and algorithms during the automation and tuning process. If you want to focus only on linear models, you need to limit the scope of algorithms that AutoML considers.
upvoted 1 times

james2033 2 years, 4 months ago
'Deep learning' for polynomial model, so 'linear model' can disable.
upvoted 1 times

endeesa 2 years, 8 months ago
Answer is A

upvoted 1 times

MarinaMijailovic 2 years, 9 months ago

The correct answer is A, as deep learning models are inherently non-linear.

upvoted 5 times

bvkr 2 years, 10 months ago

ChatGPT answer is B: option B is the most appropriate choice: B. You should enable automatic featurization. Enabling automatic featurization will allow the Automated Machine Learning interface to automatically preprocess the CSV data and extract relevant features that are compatible with linear models. Automatic featurization can also handle missing values, categorical variables, and feature scaling, which can be time-consuming and error-prone if done manually. Disabling deep learning (option A) may be necessary if the dataset is small or if the use of deep learning is not feasible or desired, but it is not relevant to the given scenario. Setting the task type to Forecasting (option D) is also not relevant since the task type has already been specified as Classification. Disabling automatic featurization (option C) may be appropriate if the CSV data has already been preprocessed and feature engineering has been performed manually, but it is not necessary if the CSV data is in its raw form.

upvoted 2 times

phdykd 2 years, 11 months ago

The correct answer is B. You should enable automatic featurization. Automated Machine Learning (AutoML) is a process that automates various aspects of machine learning, such as data preprocessing, feature engineering, and model selection. It aims to make machine learning more accessible to individuals with limited expertise in data science. In this scenario, the user has been informed that only linear models need to be assessed. Enabling automatic featurization is important because it will allow the AutoML interface to transform the data and generate additional features that may improve the performance of the linear models. Disabling automatic featurization (option C) would prevent the AutoML interface from generating additional features, potentially limiting the performance of the linear models. Disabling deep learning (option A) is not necessary since the problem does not involve deep learning models. Setting the task type to Forecasting (option D) is also incorrect since the user has been informed that a classification model needs to be trained.

upvoted 1 times

mamau 3 years ago

C. You should disable automatic featurization. Automated Machine Learning (AutoML) provides the ability to automatically engineer features in order to simplify the process of building a machine learning model. However, in this case, the requirements specify that only linear models need to be assessed, so it would be best to disable automatic featurization to ensure that only linear models are considered. This would help ensure that the results are in line with the requirements.

upvoted 2 times

manuu97 3 years, 2 months ago

By exclusion the right answer is A

upvoted 1 times

lookaaaa 3 years, 2 months ago

featurization has no effect of linear models, so BC are wrong; As for D, you should choose classification not forecasting

upvoted 3 times

JTWang 3 years, 4 months ago

The Answer is A.

upvoted 3 times

synapse 3 years, 11 months ago

Without blocking algorithm options, the answer is A.

upvoted 3 times

RyanTsai 4 years, 5 months ago

ans: B. You should enable automatic featurization.

upvoted 2 times

RyanTsai 4 years, 5 months ago

ans: B. You should enable automatic featurization.you should be blocked for misleading people

upvoted 2 times

Mirjalol 3 years ago

you should be blocked for misleading people

upvoted 10 times

BenAji 4 years, 6 months ago

No right answer, as AML only support an option to "Blocked Algorithm" https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-automated-ml-for-ml-models

upvoted 6 times

You are preparing to train a regression model via automated machine learning. The data available to you has features with missing values, as well as categorical features with little discrete values.
You want to make sure that automated machine learning is configured as follows:
✑ missing values must be automatically imputed.
✑ categorical features must be encoded as part of the training task.
Which of the following actions should you take?

A. You should make use of the featurization parameter with the 'auto' value pair.

B. You should make use of the featurization parameter with the 'off' value pair.

C. You should make use of the featurization parameter with the 'on' value pair.

D. You should make use of the featurization parameter with the 'FeaturizationConfig' value pair.

Suggested Answer: *A*





Featurization str or FeaturizationConfig Values: 'auto' / 'off' / FeaturizationConfig Indicator for whether featurization step should be done automatically or not, or whether customized featurization should be used. Column type is automatically detected. Based on the detected column type preprocessing/featurization is done as follows: Categorical: Target encoding, one hot encoding, drop high cardinality categories, impute missing values. Numeric: Impute missing values, cluster distance, weight of evidence. DateTime: Several features such as day, seconds, minutes, hours etc. Text: Bag of words, pre-trained Word embedding, text target encoding. Reference: https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train.automl.automlconfig.automlconfig

Community vote distribution

A (89%) C (11%)

linearUpwind Most Recent 6 months, 2 weeks ago

https://learn.microsoft.com/en-us/azure/machine-learning/how-to-configure-auto-features?view=azureml-api-1#configure-featurization

upvoted 1 times

james2033 2 years, 4 months ago

The anwser is clear. Question keyword 'automatically imputed' . Answer keyword '_auto_ value pair' . https://learn.microsoft.com/en-us/azure/machine-learning/how-to-configure-auto-features?view=azureml-api-1#configure-featurization

upvoted 3 times

endeesa 2 years, 8 months ago

Answer is A, if you set featurization to "auto", azure ml will do this for you https://learn.microsoft.com/en-us/azure/machine-learning/how-to-configure-auto-features?view=azureml-api-1#configure-featurization

upvoted 4 times

MarinaMijailovic 2 years, 9 months ago

A) AUTO ML will handle the necessary preprocessing steps, such as imputing missing values and encoding categorical features, before training the regression model.

upvoted 2 times

lookaaaa 3 years, 2 months ago

"on" is not an option, "auto" is

upvoted 2 times

JTWang 3 years, 4 months ago

A is correct. https://learn.microsoft.com/zh-tw/azure/machine-learning/how-to-configure-auto-features#automatic-featurization

upvoted 2 times

exnaniantwort 3 years, 5 months ago

The following table shows the accepted settings for featurization in the AutoMLConfig class: Featurization configuration Description "featurization": 'auto' Specifies that, as part of preprocessing, data guardrails and featurization steps are to be done automatically. This setting is the default. "featurization": 'off' Specifies that featurization steps are not to be done automatically. "featurization": 'FeaturizationConfig' Specifies that customized featurization steps are to be used. Learn how to customize featurization.

upvoted 1 times

ning 3 years, 8 months ago

On is not an option

upvoted 2 times

pancman 3 years, 10 months ago

I change my answer, ON is not an option under featurization. Auto is correct.

upvoted 3 times

pancman 3 years, 10 months ago

They key word here is that you want to "ensure". Therefore, you need to turn the featurization option ON. Auto is not the correct answer. If you set it to auto, it is at the discretion of Auto ML to do the required actions stated in the question.I do not think "on" is an option 1. auto --> correct answer here 2. off --> nothing happens 3. an object of FeaturizationConfig class type --> use the special setting

upvoted 2 times

pancman 3 years, 10 months ago

Selected Answer: C

They key word here is that you want to "ensure". Therefore, you need to turn the featurization option ON. Auto is not the correct answer. If you set it to auto, it is at the discretion of Auto ML to do the required actions stated in the question.

upvoted 2 times

ning 3 years, 8 months ago

I do not think "on" is an option 1. auto --> correct answer here 2. off --> nothing happens 3. an object of FeaturizationConfig class type --> use the special setting

upvoted 2 times

Question #31 *Topic 1*

You make use of Azure Machine Learning Studio to develop a linear regression model. You perform an experiment to assess various algorithms. Which of the following is an algorithm that reduces the variances between actual and predicted values?

A. Fast Forest Quantile Regression

B. Poisson Regression

C. Boosted Decision Tree Regression

D. Linear Regression

Suggested Answer: *C*

Mean absolute error (MAE) measures how close the predictions are to the actual outcomes; thus, a lower score is better. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/boosted-decision-tree-regression https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/linear-regression

Community vote distribution



RyanTsai Highly Voted 4 years, 11 months ago

should be linear regressionLinear regression minimizes the sum of squares, i.e. it minimizes S=Sum[(y_i-f_i)^2,{i,1,n}], where y_i is the actual value and f_i the predicted value (we can see it as the average). Since the variance of {y_i} is S/n, minimizing S is equivalent to minimizing the variance.

upvoted 22 times

RyanTsai Highly Voted 4 years, 11 months ago

should be linear regression

upvoted 22 times

David_Tadeu 4 years, 4 months ago

Linear regression minimizes the sum of squares, i.e. it minimizes S=Sum[(y_i-f_i)^2,{i,1,n}], where y_i is the actual value and f_i the predicted value (we can see it as the average). Since the variance of {y_i} is S/n, minimizing S is equivalent to minimizing the variance.

upvoted 3 times

jed_elhak 4 years, 11 months ago

Correct answer is C : It builds each regression tree in a step-wise fashion, using a predefined loss function to measure the error in each step and correct for it in the next. Thus the prediction model is actually an ensemble of weaker prediction models. reference:https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/boosted-decision-tree-regression

upvoted 17 times

jed_elhak 4 years, 11 months ago

Correct answer is C : It builds each regression tree in a step-wise fashion, using a predefined loss function to measure the error in each step and correct for it in the next. Thus the prediction model is actually an ensemble of weaker prediction models. reference:https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/boosted-decision-tree-regressionAgree with you

upvoted 17 times

dija123 4 years, 7 months ago

Agree with you

upvoted 1 times

rysuf Most Recent 6 months, 2 weeks ago

Boosted Decision Tree Regression uses an ensemble of decision trees built sequentially. Each new tree focuses on correcting the errors (residuals) of the previous ones. This reduces variance between the actual and predicted values by learning from mistakes and stabilizing predictions. Linear Regression fits a single linear equation by minimizing squared error, but it does not explicitly reduce variance through ensembling. It's simple and interpretable, but limited when relationships are complex or non-linear.

upvoted 1 times

Guru_1 8 months, 1 week ago

C. Boosted Decision Tree Regression - Ensemble methods like Boosted Decision Tree Regression use multiple models to improve overall performance. Specifically, boosting combines weak learners sequentially, with each new learner attempting to correct the errors made by previous

ones. The aggregation of these models helps to reduce the variance of the predictions between the actual and predicted values. Linear Regression, on the other hand, can be highly sensitive to random errors in observed data, which can result in large variance.

upvoted 1 times

sar77 1 year, 1 month ago

Boosted Decision Tree Regression, including Azure ML's Boosted Decision Tree (based on LightGBM/MART), is designed to minimize errors between actual and predicted values by sequentially fitting trees that correct previous residuals. This iterative boosting process both reduces bias and lowers variance, resulting in more accurate predictions .

upvoted 1 times

PatAms538 1 year, 2 months ago

The Linear Regression is built upon the F-test. The logic of the F-test: the amount of systematic variance divided by the amount of unsystematic variance

upvoted 1 times

brzhanyu 1 year, 10 months ago

The correct answer is Boosted Decision Tree Regression. Boosted Decision Tree Regression is an algorithm that reduces the variance between actual and predicted values by iteratively combining multiple weak learners (decision trees) to create a stronger, more accurate model. This helps minimize the differences between the predicted values and the actual values by focusing on the errors made by previous models. Linear regression, while useful for linear relationships, does not specifically focus on reducing variance in the way ensemble methods like boosting do.

upvoted 2 times

6c3c83d 2 years, 3 months ago

D is linear and reduces the mean squared error which is the variance between actual and predicted. Regression trees do that too but are not linear

upvoted 1 times

GHill1982 2 years, 6 months ago

Linear regression computes the linear relationship between a dependent variable and one or more independent features. It uses the ordinary least squares method to minimize the sum of the squared errors between the actual and predicted values.

upvoted 2 times

jdada 2 years, 9 months ago

D. Linear Regression

upvoted 2 times

james2033 2 years, 10 months ago

Question keyword 'develop a linear regression model'. Answer keyword 'Linear regression' make (D) is correct answer. https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/linear-regression?view=azureml-api-2 Quote: 'create a linear regression model' , 'Linear regression is a common statistical method, ...'

upvoted 5 times

Nadine_nm 3 years ago

I think the question is poorly asked, since the algorithme used depends mainy on the data. The answer is linear regression, since Poisson regression is actually a form of generalized linear modeling and is often considered a non-linear regression model due to its underlying mathematical structure. Also Boosted decision tree, in a non linear regression as well.

upvoted 2 times

phydev 3 years, 1 month ago

Boosted decision tree regression a non-linear regression model.

upvoted 1 times

endeesa 3 years, 2 months ago

Selected Answer: C

Boosting will definitely reduce the variances

upvoted 1 times

endeesa 3 years, 2 months ago

Boosting will definitely reduce the variancesBut the the question says we're trying to develop a linear regression model. Shouldn't the influence the option and it's supposed to be D

upvoted 1 times

daviduzo 3 years, 1 month ago

But the the question says we're trying to develop a linear regression model. Shouldn't the influence the option and it's supposed to be D

upvoted 1 times

NickRafe 3 years, 3 months ago

Answer should be C. Reference: https://www.kdd.org/kdd2016/papers/files/adf0653-poyarkovA.pdf

upvoted 1 times

EXAMTOPICS2134 3 years, 4 months ago

Decision trees are non linear. Unlike Linear regression there is no equation to express relationship between independent and dependent variables

upvoted 1 times

bibabrian 3 years, 4 months ago

guys I know some of these questions are crap but read carefully it says "assess various algorithms.", which means a boosting algorithm like Boosted Decision Tree Regression makes sense.

upvoted 2 times

Question #32 *Topic 1*

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You have been tasked with constructing a machine learning model that translates language text into a different language text.
The machine learning model must be constructed and trained to learn the sequence of the.
Recommendation: You make use of Convolutional Neural Networks (CNNs).
Will the requirements be satisfied?

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Community vote distribution
B (100%)

Juttis Highly Voted 4 years ago
Use Reccurent Neural Network (RNN) for translations.
upvoted 13 times

james2033 1 year, 10 months ago
Convolution Neural Networks (CNN) for training images machine learning model. Recurrent Neural Networks (RNN) for training text (words sequence) in term of seq2seq, but now it is out of date, use Transformer as alternative solution. Anyway, CNN is incorrect for NLP.Transfer learning (Transformer architecture) for Machine translation https://learn.microsoft.com/en-us/azure/machine-learning/concept-deep-learning-vs-machine-learning?view=azureml-api-2#machine-translation .
upvoted 2 times

james2033 1 year, 10 months ago
Selected Answer: B
Convolution Neural Networks (CNN) for training images machine learning model.

Recurrent Neural Networks (RNN) for training text (words sequence) in term of seq2seq, but now it is out of date, use Transformer as alternative solution.

Anyway, CNN is incorrect for NLP.
upvoted 2 times

james2033 Most Recent 1 year, 10 months ago
Transfer learning (Transformer architecture) for Machine translation https://learn.microsoft.com/en-us/azure/machine-learning/concept-deep-learning-vs-machine-learning?view=azureml-api-2#machine-translation .
upvoted 1 times

EmmettBrown 2 years, 4 months ago
B is the answer
upvoted 1 times

eloyinaay 2 years, 5 months ago
The answer is correct RNN are for NLP and
upvoted 1 times

Arend78 2 years, 8 months ago
https://www.techtarget.com/searchenterpriseai/feature/CNN-vs-RNN-How-they-differ-and-where-they-overlap "CNNs are preferred in interpreting visual data, sparse data or data that does not come in sequence," [...] "Recurrent neural networks, on the other hand, are designed to recognize sequential or temporal data. They do better predictions considering the order or sequence of the data as they relate to previous or the next data nodes."
upvoted 4 times

ranjsi01 3 years, 6 months ago
cnn for image classification

upvoted 4 times

Question #33 *Topic 1*

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You have been tasked with constructing a machine learning model that translates language text into a different language text.
The machine learning model must be constructed and trained to learn the sequence of the.
Recommendation: You make use of Generative Adversarial Networks (GANs).
Will the requirements be satisfied?

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Community vote distribution
B (100%)

sar77 Most Recent 1 year, 1 month ago
Generative Adversarial Networks (GAN) for training generative AI. Recurrent Neural Networks (RNN) for training text (words sequence) in term of seq2seq, but now it is out of date, use Transformer as alternative solution.
upvoted 1 times

SwamiShubham0 1 year, 11 months ago
sequence of the....? is the question incomplete?
upvoted 2 times

james2033 2 years, 10 months ago
Generative Adversarial Networks (GAN) for training generative AI. Recurrent Neural Networks (RNN) for training text (words sequence) in term of seq2seq, but now it is out of date, use Transformer as alternative solution. Anyway, GAN is incorrect for NLP (translation). https://learn.microsoft.com/en-us/azure/machine-learning/concept-deep-learning-vs-machine-learning?view=azureml-api-2#generative-adversarial-network-gan https://learn.microsoft.com/en-us/azure/machine-learning/concept-deep-learning-vs-machine-learning?view=azureml-api-2#machine-translation
upvoted 2 times

dispoc 2 years, 11 months ago
GAN's are for new data generation.
upvoted 1 times

meysa 3 years, 7 months ago
No, Generative adversarial networks (GANs) are algorithmic architectures that use two neural networks, pitting one against the other (thus the “adversarial”) in order to generate new, synthetic instances of data that can pass for real data. They are used widely in image generation, video generation and voice generation.
upvoted 4 times

ranjsi01 4 years, 6 months ago
GANs used for image translation
upvoted 3 times

Question #34 *Topic 1*

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You have been tasked with constructing a machine learning model that translates language text into a different language text.
The machine learning model must be constructed and trained to learn the sequence of the.
Recommendation: You make use of Recurrent Neural Networks (RNNs).
Will the requirements be satisfied?

A. Yes

B. No

Suggested Answer: *A*
Note: RNNs are designed to take sequences of text as inputs or return sequences of text as outputs, or both. They're called recurrent because the network's hidden layers have a loop in which the output and cell state from each time step become inputs at the next time step. This recurrence serves as a form of memory. It allows contextual information to flow through the network so that relevant outputs from previous time steps can be applied to network operations at the current time step. Reference: https://towardsdatascience.com/language-translation-with-rnns-d84d43b40571

Community vote distribution
A (100%)

james2033 Most Recent 1 year, 10 months ago
'Machine translation' in sequence way, use Recurrent Neural Networks (RNN). If not the request did not contain 'sequence' way, use Transformer is contemporary best practise.
upvoted 1 times

dispoc 1 year, 11 months ago
RNN are good for learning of sequence
upvoted 1 times

noname_123 2 years, 6 months ago
https://learn.microsoft.com/en-us/azure/machine-learning/concept-deep-learning-vs-machine-learning Recurrent neural network (RNN) Recurrent neural networks are a widely used artificial neural network. These networks save the output of a layer and feed it back to the input layer to help predict the layer's outcome. Recurrent neural networks have great learning abilities. They're widely used for complex tasks such as time series forecasting, learning handwriting, and recognizing language.
upvoted 2 times

Gabonia 2 years, 12 months ago
Correct
upvoted 4 times

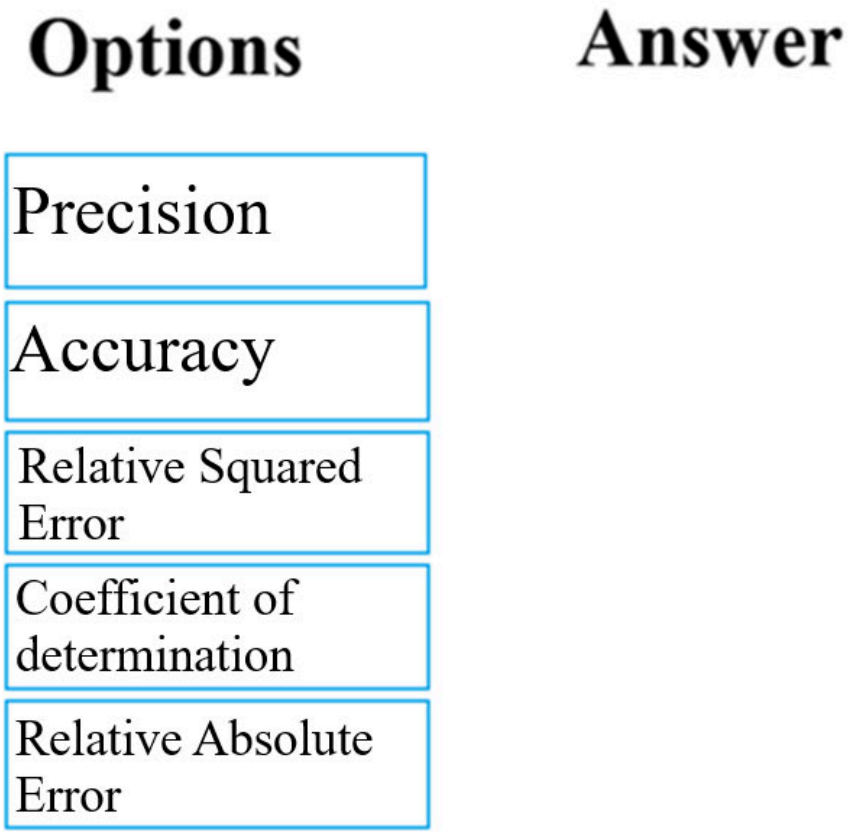
DRAG DROP -

You have been tasked with evaluating the performance of a binary classification model that you created.

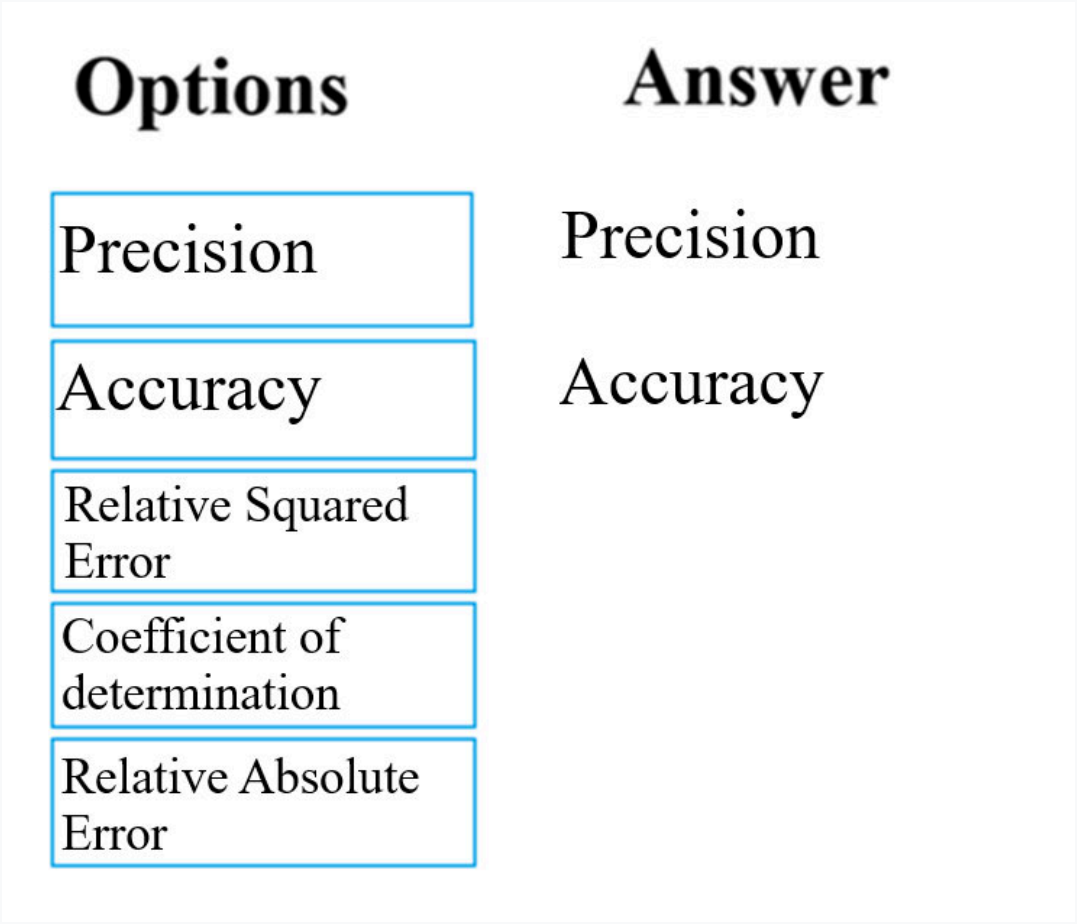
You need to choose evaluation metrics to achieve your goal.

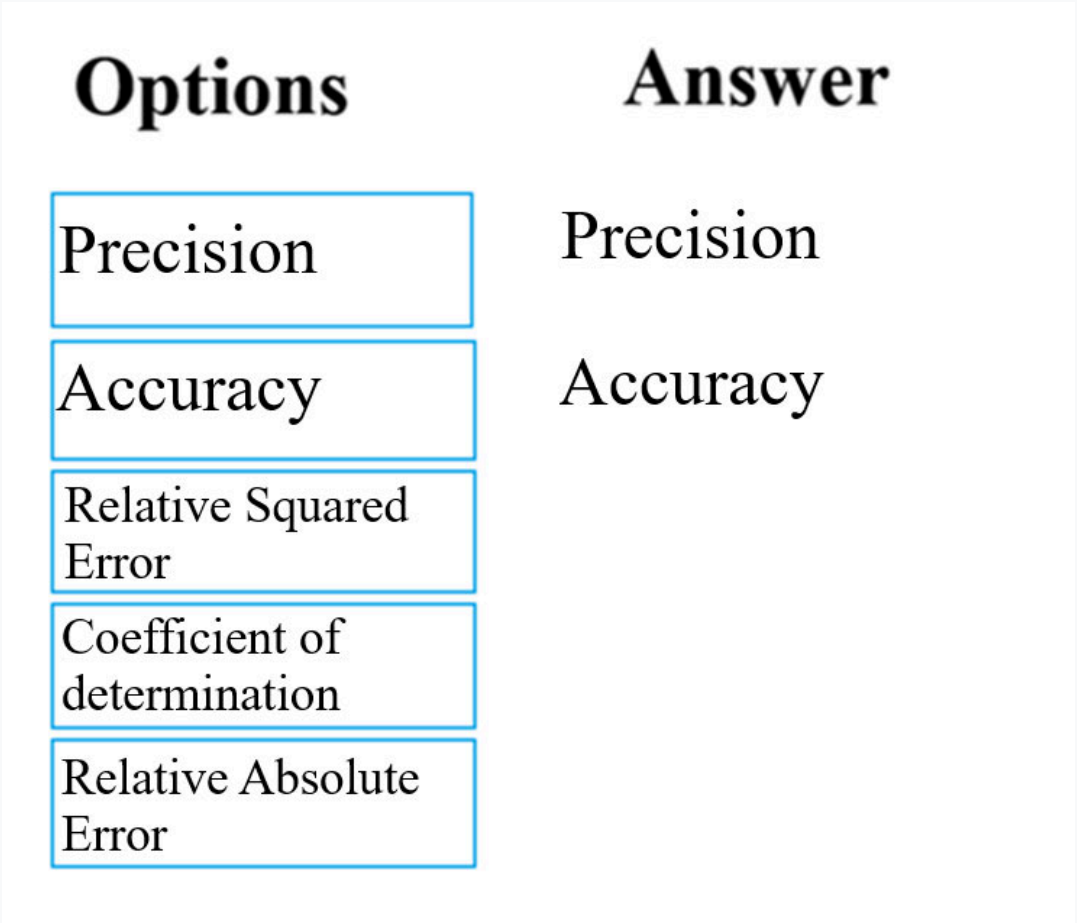
Which of the following are the metrics you would choose? Answer by dragging the correct options from the list to the answer area.

Select and Place:



Suggested Answer:





The evaluation metrics available for binary classification models are: Accuracy, Precision, Recall, F1 Score, and AUC. Note: A very natural question is: 'Out of the individuals whom the model, how many were classified correctly (TP)?' This question can be answered by looking at the

Precision of the model, which is the proportion of positives that are classified correctly. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio/evaluate-model-performance

pancman Highly Voted 2 years, 10 months ago

Answer is correct

upvoted 9 times

endeesa Most Recent 1 year, 8 months ago

Given answer is correct

upvoted 1 times

Mirjalol 2 years ago

Precision and accuracy, so yes, answer is correct

upvoted 2 times

Question #36 *Topic 1*

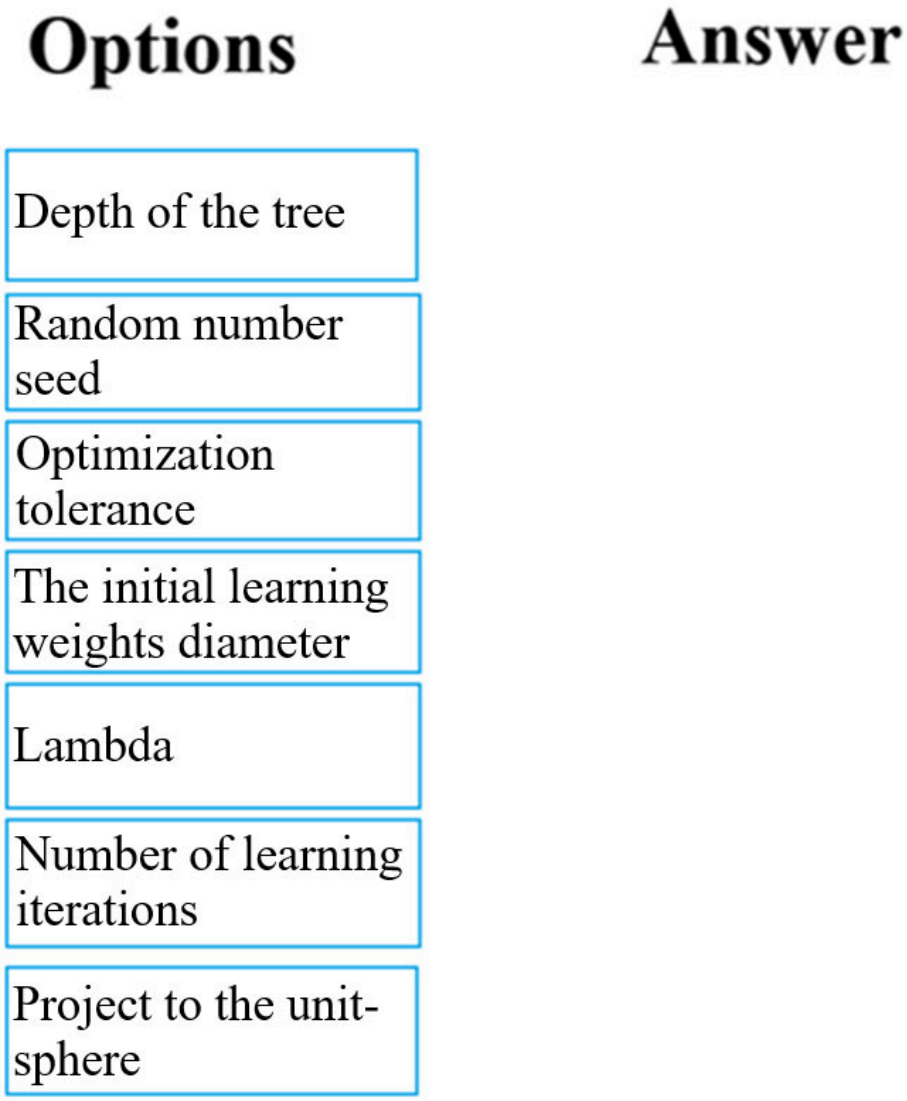
DRAG DROP -

You build a binary classification model using the Azure Machine Learning Studio Two-Class Neural Network module.

You are preparing to configure the Tune Model Hyperparameters module for the purpose of tuning accuracy for the model.

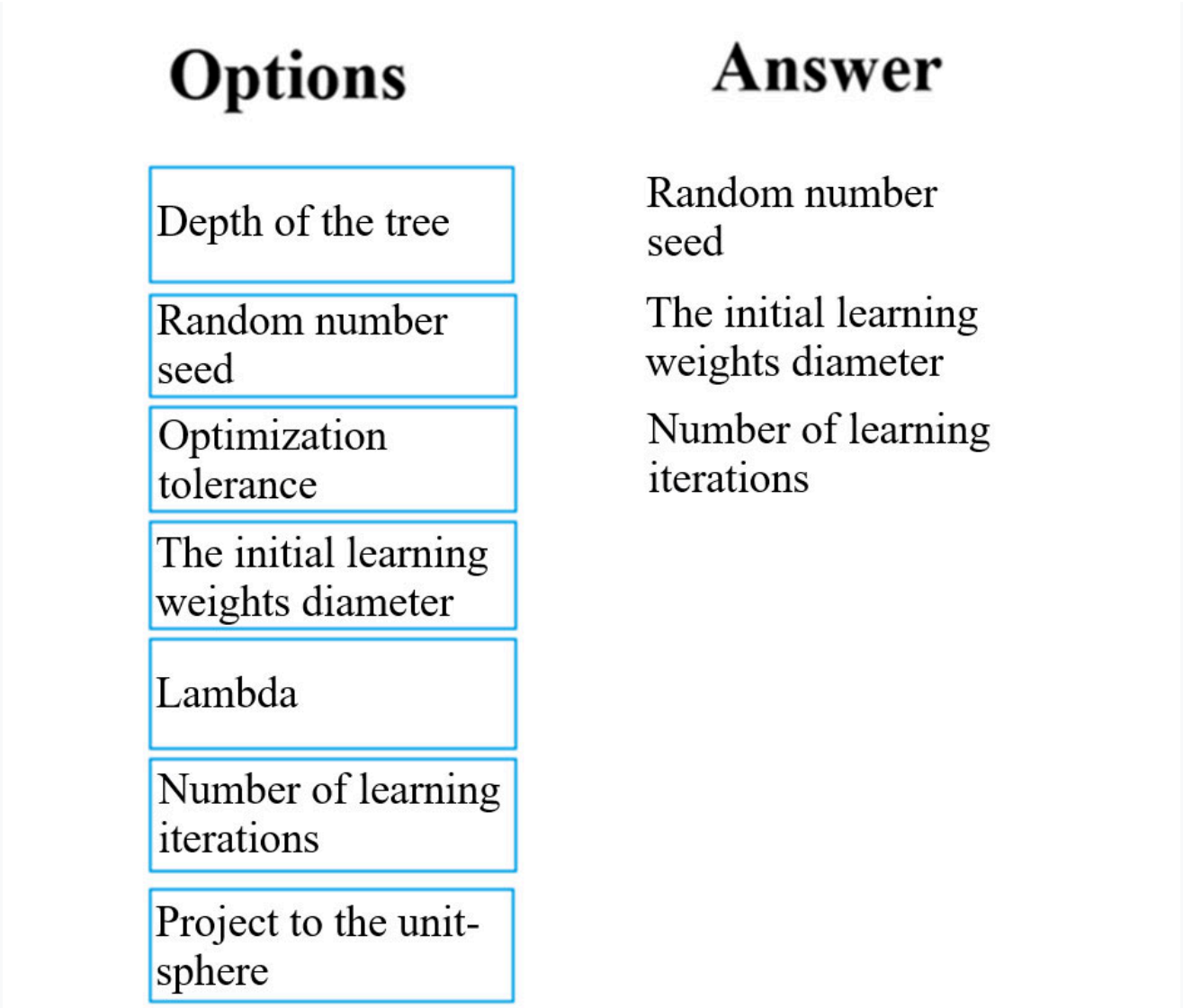
Which of the following are valid parameters for the Two-Class Neural Network module? Answer by dragging the correct options from the list to the answer area.

Select and Place:



Suggested Answer:

Options	Answer
Depth of the tree	Random number seed
Random number seed	The initial learning weights diameter
Optimization tolerance	Number of learning iterations
The initial learning weights diameter	
Lambda	
Number of learning iterations	
Project to the unit-sphere	



Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-neural-network

KingTN Highly Voted 3 years, 6 months ago

Random seed does not help in tuning accuracy, it just make experiments reproducible. I think it should not be used.

https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/two-class-neural-network

"For Random number seed, type a value to use as the seed. Specifying a seed value is useful when you want to ensure repeatability across runs of the same pipeline. Otherwise, a system clock value is used as the seed, which can cause slightly different results each time you run the pipeline."

upvoted 6 times

KingTN Highly Voted 3 years, 6 months ago

Random seed does not help in tuning accuracy, it just make experiments reproducible. I think it should not be used. https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/two-class-neural-network "For Random number seed, type a value to use as the seed. Specifying a seed value is useful when you want to ensure repeatability across runs of the same pipeline. Otherwise, a system clock value is used as the seed, which can cause slightly different results each time you run the pipeline."The question read: "Which of the following are valid parameters for the Two-Class Neural Network module?" And according the Microsoft's Azure documentation regarding 2-class NN's, Random Seed is a valid parameter.

upvoted 6 times

michaelmorar 3 years, 5 months ago

The question read: "Which of the following are valid parameters for the Two-Class Neural Network module?" And according the Microsoft's Azure documentation regarding 2-class NN's, Random Seed is a valid parameter.

upvoted 3 times

ning 4 years, 2 months ago

Random seed is a parameter for binary classification, but I do not understand "Hyperparameters" has anything to do with it???Random seed is not a parameter in the context of machine learning. Random seed is a Hyperparameter.

upvoted 5 times

ning 4 years, 2 months ago

Random seed is a parameter for binary classification, but I do not understand "Hyperparameters" has anything to do with it???

upvoted 5 times

SweetChilliPhilly 3 years, 9 months ago

Random seed is not a parameter in the context of machine learning. Random seed is a Hyperparameter.

upvoted 1 times

Xsesi Most Recent 2 years ago

Could anyone please provide insight into why one might choose not to use Optimization Tolerance? No question with selecting The Initial learning weights diameter & Number of learning iterations. Random seed shall not be selected since it does not optimise during the training process but merely ensure consistent results.

upvoted 1 times

EmmettBrown 3 years, 4 months ago

Answer is correct https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/two-class-neural-network?view=azureml-api-2#how-to-configure

upvoted 3 times

EmmettBrown 3 years, 4 months ago

Answer is correct https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/two-class-neural-network?view=azureml-api-2#how-to-configure100% agree with you

upvoted 3 times

prabhjot 2 years, 6 months ago

100% agree with you

upvoted 1 times

bvkr 3 years, 4 months ago

Chat GPT answer:The Two-Class Neural Network module has the following valid parameters: The Initial learning weights diameter: This parameter controls the magnitude of the weights in the neural network at the start of the training process. Lambda: This parameter controls the amount of regularization applied to the neural network during training. Number of learning iterations: This parameter controls the number of iterations performed during the training process.

upvoted 1 times

victorafb 3 years, 10 months ago

https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/two-class-neural-network Correct

upvoted 2 times

Question #37 *Topic 1*

You make use of Azure Machine Learning Studio to create a binary classification model.
You are preparing to carry out a parameter sweep of the model to tune hyperparameters. You have to make sure that the sweep allows for every possible combination of hyperparameters to be iterated. Also, the computing resources needed to carry out the sweep must be reduced.
Which of the following actions should you take?

A. You should consider making use of the Selective grid sweep mode.

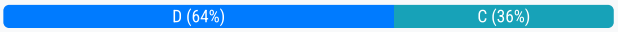
B. You should consider making use of the Measured grid sweep mode.

C. You should consider making use of the Entire grid sweep mode.

D. You should consider making use of the Random grid sweep mode.

Suggested Answer: *D*
Maximum number of runs on random grid: This option also controls the number of iterations over a random sampling of parameter values, but the values are not generated randomly from the specified range; instead, a matrix is created of all possible combinations of parameter values and a random sampling is taken over the matrix. This method is more efficient and less prone to regional oversampling or undersampling. If you are training a model that supports an integrated parameter sweep, you can also set a range of seed values to use and iterate over the random seeds as well. This is optional, but can be useful for avoiding bias introduced by seed selection. C: Entire grid: When you select this option, the module loops over a grid predefined by the system, to try different combinations and identify the best learner. This option is useful for cases where you don't know what the best parameter settings might be and want to try all possible combination of values. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/tune-model-hyperparameters

Community vote distribution



[Removed] Highly Voted 3 years, 7 months ago
I feel like like the two requirements are conflicting: every possible combination implies Entire grid while lower computational resources implies Random grid. "Entire grid: When you select this option, the component loops over a grid predefined by the system, to try different combinations and identify the best learner. This option is useful when you don't know what the best parameter settings might be and want to try all possible combinations of values." "Random sweep: When you select this option, the component will randomly select parameter values over a system-defined range. You must specify the maximum number of runs that you want the component to execute. This option is useful when you want to increase model performance by using the metrics of your choice but still conserve computing resources."For ensuring every possible combination of hyperparameters is explored, the "Entire grid sweep mode" is the appropriate choice.
upvoted 18 times

[Removed] Highly Voted 3 years, 7 months ago
I feel like like the two requirements are conflicting: every possible combination implies Entire grid while lower computational resources implies Random grid. "Entire grid: When you select this option, the component loops over a grid predefined by the system, to try different combinations and identify the best learner. This option is useful when you don't know what the best parameter settings might be and want to try all possible combinations of values."

"Random sweep: When you select this option, the component will randomly select parameter values over a system-defined range. You must specify the maximum number of runs that you want the component to execute. This option is useful when you want to increase model performance by using the metrics of your choice but still conserve computing resources."
upvoted 18 times

prabhjot 2 years, 6 months ago
For ensuring every possible combination of hyperparameters is explored, the "Entire grid sweep mode" is the appropriate choice.
upvoted 1 times

lookaaaa 3 years, 8 months ago
All combination + Reduce computing resource , because "Research has shown that this method (Random Grid Sweep) yields the same results, but is more efficient computationally." I think D would be the best choice
upvoted 9 times

Guru_1 Most Recent 8 months, 1 week ago

"to make sure that the sweep allows for every possible combination of hyperparameters to be iterated...", Random will not be an appropriate and It's not guaranteed every possible combination, but random. So, The answer would be "(c) You should consider making use of the Entire grid sweep mode"

upvoted 2 times

sim39 1 year, 8 months ago

ALLOWS for every possible combination doesn't mean it has to iterate through all. The requirement to reduce compute resources obviously points us away from a full grid search.

upvoted 3 times

Xsesi 2 years ago

Vote for Entire Grid Sweep Mode since one requirement is every possible combination of hyperparameters to be iterated. Other options reduce computational resources yet do not satisfy this requirement.

upvoted 1 times

Braxus 2 years, 5 months ago

"Maximum number of runs on random grid: This option also controls the number of iterations over a random sampling of parameter values, but the values are not generated randomly from the specified range; instead, a matrix is created of all possible combinations of parameter values and a random sampling is taken over the matrix. This method is more efficient and less prone to regional oversampling or undersampling."

upvoted 1 times

MarinaMijailovic 3 years, 1 month ago

The answer is D. Random seed ALLOWS every possible combination. It won't go through every possible combination but any random combination is possible.

upvoted 2 times

endeesa 3 years, 2 months ago

Question says "You have to make sure that the sweep allows for every possible combination of hyperparameters to be iterated". There is no way to guarantee Random sweep will get all possible combinations, answer is C

upvoted 2 times

phdykd 3 years, 6 months ago

D. You should consider making use of the Random grid sweep mode. The Random grid sweep mode randomly selects combinations of hyperparameters to test, reducing the number of total combinations and the computing resources needed to carry out the sweep. This method can still provide a good understanding of the relationship between hyperparameters and model performance, but may require multiple runs to converge on the optimal hyperparameters.

upvoted 1 times

meysa 3 years, 7 months ago

If we want every possible combi9nation we need entire grid, so C is correct.

upvoted 3 times

Sibajene 3 years, 7 months ago

C is correct..

upvoted 4 times

Edriv 3 years, 8 months ago

Keyword -> "sweep allows for every possible combination" so, option B https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/tune-model-hyperparametersI mean, option C

upvoted 1 times

Edriv 3 years, 8 months ago

Keyword -> "sweep allows for every possible combination" so, option B https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/tune-model-hyperparameters

upvoted 1 times

Edriv 3 years, 8 months ago

I mean, option C

upvoted 3 times

ning 4 years, 2 months ago

Entire grid is the only one can try all combinations, but random sweep is low computational cost

upvoted 3 times

David_Tadeu 4 years, 4 months ago

Related with this question https://www.examtopics.com/discussions/microsoft/view/43145-exam-dp-100-topic-2-question-80-discussion/

upvoted 2 times

synapse 4 years, 5 months ago

Answer D. there are two reqs... entire grid and low computational cost

upvoted 1 times

dija123 4 years, 8 months ago

Random grid

upvoted 1 times

RyanTsai 4 years, 11 months ago

ans: C

upvoted 4 times

Question #38 *Topic 1*

You are in the process of constructing a deep convolutional neural network (CNN). The CNN will be used for image classification.
You notice that the CNN model you constructed displays hints of overfitting.
You want to make sure that overfitting is minimized, and that the model is converged to an optimal fit.
Which of the following is TRUE with regards to achieving your goal?

A. You have to add an additional dense layer with 512 input units, and reduce the amount of training data.

B. You have to add L1/L2 regularization, and reduce the amount of training data.

C. You have to reduce the amount of training data and make use of training data augmentation.

D. You have to add L1/L2 regularization, and make use of training data augmentation.

E. You have to add an additional dense layer with 512 input units, and add L1/L2 regularization.

Suggested Answer: *B*
B: Weight regularization provides an approach to reduce the overfitting of a deep learning neural network model on the training data and improve the performance of the model on new data, such as the holdout test set. Keras provides a weight regularization API that allows you to add a penalty for weight size to the loss function. Three different regularizer instances are provided; they are: ✑ L1: Sum of the absolute weights. ✑ L2: Sum of the squared weights. ✑ L1L2: Sum of the absolute and the squared weights. Because a fully connected layer occupies most of the parameters, it is prone to overfitting. One method to reduce overfitting is dropout. At each training stage, individual nodes are either "dropped out" of the net with probability 1-p or kept with probability p, so that a reduced network is left; incoming and outgoing edges to a dropped-out node are also removed. By avoiding training all nodes on all training data, dropout decreases overfitting. Reference:
https://machinelearningmastery.com/how-to-reduce-overfitting-in-deep-learning-with-weight-regularization/
https://en.wikipedia.org/wiki/Convolutional_neural_network

Community vote distribution
D (100%)

exam_monkey1234 Highly Voted 4 years, 1 month ago
I would say answer Dagree, option B reduce amount of training data will lead to overfitting."data augmentation simply means increasing size of the data that is increasing the number of images present in the dataset.. using data augmentation a lot of similar images can be generated. This helps in increasing the dataset size and thus reduce overfitting." https://www.kdnuggets.com/2019/12/5-techniques-prevent-overfitting-neural-networks.html
upvoted 50 times

exam_monkey1234 Highly Voted 4 years, 1 month ago
I would say answer D
upvoted 50 times

Nghia1 2 years, 2 months ago
agree, option B reduce amount of training data will lead to overfitting.
upvoted 1 times

Moshekwa 4 years ago
"data augmentation simply means increasing size of the data that is increasing the number of images present in the dataset.. using data augmentation a lot of similar images can be generated. This helps in increasing the dataset size and thus reduce overfitting."
https://www.kdnuggets.com/2019/12/5-techniques-prevent-overfitting-neural-networks.html
upvoted 14 times

phdykd 2 years, 6 months ago
Moderator, you should correct the answers, not me! D. You have to add L1/L2 regularization and make use of training data augmentation. Overfitting occurs when a model is too complex for data it is being trained on and memorizes the training data instead of generalizing to new data. To reduce overfitting, you can use regularization techniques such as L1 or L2 regularization, which add a penalty term to the loss function to discourage the model from learning overly complex representations. Additionally, increasing the amount of training data can also help reduce overfitting by giving the model more information to learn from. One common way to increase the amount of training data is to use data augmentation, which involves transforming the existing data in ways that preserve the labels to generate additional training examples.
upvoted 9 times

dporwal04 Most Recent 1 year, 8 months ago
D is the correct ans

upvoted 2 times

rakeshmk 1 year, 10 months ago

The answer is D. Data augmentation really helps to reduce overfitting and L1L2 are most used regularization in Neural networks.. Dear moderator , please provide the correct answers. Many of them given here is misleading which can make chaos while attending exam

upvoted 1 times

endeesa 2 years, 2 months ago

The only option that makes sense to me is D. Adding regularisation will reduce overfitting, similarly adding more data adds more diversity to the training set allowing it to generalise better. So answer is D

upvoted 1 times

bvkr 2 years, 4 months ago

ChatGPT answer: Option D: You have to add L1/L2 regularization, and make use of training data augmentation. When a deep CNN model displays hints of overfitting, it means that the model is too complex and has learned to fit the training data too closely. One way to minimize overfitting is to add regularization to the model, which adds a penalty term to the loss function, encouraging the model to choose simpler solutions. L1/L2 regularization adds a penalty term to the loss function that discourages the model from using large weights in the network. This has the effect of reducing the complexity of the model and can help prevent overfitting. Data augmentation is another effective technique to minimize overfitting. It involves applying random transformations to the training data, such as random rotations or translations, to create new training examples that are similar to the original ones. This helps the model to generalize better to unseen data.

upvoted 2 times

Yoshizn 2 years, 6 months ago

Answer is D

upvoted 1 times

MansoorDataScientist 2 years, 6 months ago

Steps for reducing overfitting: Add more data. Use data augmentation. Use architectures that generalize well. Add regularization (mostly dropout, L1/L2 regularization are also possible) Reduce architecture complexity.

upvoted 2 times

Peeking 2 years, 7 months ago

D is definitely the answer.

upvoted 1 times

Edriv 2 years, 8 months ago

Why don't C?

upvoted 1 times

lookaaaa 2 years, 8 months ago

increse amount of data, simplify the model (decrese layers or NN unit, etc)

upvoted 2 times

zweic 2 years, 10 months ago

I would say D

upvoted 2 times

jlopezfelizzola 2 years, 11 months ago

My vote is for D. A & E discarded because that is increasing the complexity of the architecture. B, C are suggesting reducing the amount of data. D will generate more data for the CNN to be able to generalize more.Source https://towardsdatascience.com/deep-learning-3-more-on-cnns-handling-overfitting-2bd5d99abe5d

upvoted 2 times

jlopezfelizzola 2 years, 11 months ago

Selected Answer: D

My vote is for D. A & E discarded because that is increasing the complexity of the architecture. B, C are suggesting reducing the amount of data. D will generate more data for the CNN to be able to generalize more.

upvoted 2 times

jlopezfelizzola 2 years, 11 months ago

Source https://towardsdatascience.com/deep-learning-3-more-on-cnns-handling-overfitting-2bd5d99abe5d

upvoted 1 times

synapse 3 years, 5 months ago

Definitely D

upvoted 1 times

dija123 3 years, 8 months ago

I vote for D

upvoted 1 times

jed_elhak 3 years, 11 months ago

Answer is D :)

upvoted 4 times

Maryam89 4 years ago

Answer is D

upvoted 4 times

Question #39 *Topic 1*

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You are planning to make use of Azure Machine Learning designer to train models.
You need choose a suitable compute type.
Recommendation: You choose Attached compute.
Will the requirements be satisfied?

A. Yes

B. No

Suggested Answer: *B*
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-studio

Community vote distribution
B (67%) A (33%)

dushmantha Highly Voted 4 years, 11 months ago
Should be True. Because we can use databricks or vm as attached compute for training purposes
upvoted 13 times

dushmantha Highly Voted 4 years, 11 months ago
Should be True. Because we can use databricks or vm as attached compute for training purposesAgree, Yes is the correct answerDesigner! Only computer instance or computer cluster
upvoted 13 times

beny 4 years, 11 months ago
Agree, Yes is the correct answer
upvoted 5 times

ning 4 years, 2 months ago
Designer! Only computer instance or computer cluster
upvoted 5 times

azayra 4 years, 9 months ago
Azure Machine Learning designer to train models. its the key so answer is compute cluster.
upvoted 8 times

smaha_18 Most Recent 8 months, 1 week ago
I checked the official documentation of Microsoft. https://learn.microsoft.com/en-us/azure/machine-learning/concept-compute-target?view=azureml-api-2 . It clearly says that for the development approach of Azure ML Designer, Compute targets outside of Azure ML workspaces like Attaching Computes, Remote VM's are not recommended. For the developemt approach of Azure AutoML, Azure NOtebooks, Yes Attached Compute is an option
upvoted 1 times

testgm 1 year, 9 months ago
Attached compute is already supported by Azure Machine Learning Designer
upvoted 1 times

sl_mslconsulting 2 years, 2 months ago
There are now 5 tabs in compute, and "Attached computes" is one of them. When you configure & submit a pipeline job, you can choose attached compute.
upvoted 2 times

phdykd 3 years ago
A-Yes is answer
upvoted 1 times

SunilB 3 years, 5 months ago
We can use attached compute. Just tried it
upvoted 5 times

SunilB 3 years, 5 months ago

Correct Answer: B See the blue box in the link below which says 'Attached compute is not supported, use compute instances or clusters instead.' https://learn.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-train-score

upvoted 5 times

phdykd 3 years, 6 months ago

It is not possible to determine whether the requirements will be satisfied based on the information provided. The recommendation to choose Attached compute is a specific solution, but the requirements are not stated. It is important to have a clear understanding of the requirements and constraints before making a decision on which compute type to use in Azure Machine Learning Designer.

upvoted 1 times

Starlite 3 years, 6 months ago

ChatGPT says A

upvoted 1 times

Edriv 3 years, 8 months ago

https://learn.microsoft.com/en-us/azure/machine-learning/concept-designer#compute

upvoted 1 times

lookaaaa 3 years, 8 months ago

As is shown in the links below, Designer only supports Azure Machine Learning Compute (AML Compute cluster or AML compute instance) for training. 1. https://learn.microsoft.com/en-us/azure/machine-learning/concept-designer#compute 2. https://learn.microsoft.com/en-us/azure/machine-learning/concept-compute-target#training-compute-targets

upvoted 4 times

fvil 3 years, 9 months ago

Cannot use Attached Compute Cluster in designer: https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target Same context but slightly different question on exam 07/11/2022

upvoted 2 times

amokrane_mancer 3 years, 10 months ago

Answer is A

upvoted 1 times

JTWang 3 years, 10 months ago

No is the correct answer. https://learn.microsoft.com/en-us/azure/machine-learning/concept-compute-target ML Designer support 3 target: 1.Azure Machine Learning compute cluste 2.Azure Machine Learning compute instance 3.Azure Machine Learning Kubernetesplus 4. Azure Machine Learning serverless compute

upvoted 2 times

JTWang 3 years, 10 months ago

Selected Answer: B

No is the correct answer. https://learn.microsoft.com/en-us/azure/machine-learning/concept-compute-target ML Designer support 3 target: 1.Azure Machine Learning compute cluste 2.Azure Machine Learning compute instance 3.Azure Machine Learning Kubernetes

upvoted 2 times

hiyoww 2 years, 4 months ago

plus 4. Azure Machine Learning serverless compute

upvoted 1 times

Steven2022 3 years, 11 months ago

i just tried in AML designer, can choose attached compute now

upvoted 6 times

WeiD 4 years, 3 months ago

can attach Kubernetes service for training and inferencing (preview) Azure Machine Learning provides you with the following options to attach your own Kubernetes clusters for training and inferencing: Azure Kubernetes Service. Azure Kubernetes Service provides a managed cluster in Azure. Azure Arc Kubernetes. Use Azure Arc-enabled Kubernetes clusters if your cluster is hosted outside of Azure. Those 2 also listed in the https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target

upvoted 1 times

Question #40 *Topic 1*

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You are planning to make use of Azure Machine Learning designer to train models.
You need choose a suitable compute type.
Recommendation: You choose Inference cluster.
Will the requirements be satisfied?

A. Yes

B. No

Suggested Answer: *B*
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-studio

Community vote distribution
B (100%)

Moshekwa Highly Voted 3 years, 6 months ago
inference cluster is to deploy not train hence B
upvoted 11 times

krishna1818 Most Recent 1 year, 8 months ago
Its compute cluster
upvoted 1 times

synapse 2 years, 11 months ago
Training non on inference cluster
upvoted 2 times

Question #41 *Topic 1*

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You are planning to make use of Azure Machine Learning designer to train models.
You need choose a suitable compute type.
Recommendation: You choose Compute cluster.
Will the requirements be satisfied?

A. Yes

B. No

Suggested Answer: *A*
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-studio

Community vote distribution
A (100%)

phdykd Most Recent 2 years ago
Create a cloud-based compute instance to use for your development environment. Create a cloud-based compute cluster to use for training your model.
upvoted 4 times

MonaHallaq 2 years, 1 month ago
According to the link shared below https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target#train even compute instance can be used. Keep in mind that Azure is changing and evolving all the time
upvoted 2 times

PremPatrick 2 years, 3 months ago
https://learn.microsoft.com/en-us/azure/machine-learning/concept-compute-target#train Says even Compute Instance is possible Answer?
upvoted 3 times

pancman 2 years, 10 months ago
Selected Answer: A
Yes is the correct answer
upvoted 4 times

pancman 2 years, 10 months ago
Yes is the correct answercorrect
upvoted 4 times

Gabonia 2 years, 6 months ago
correct
upvoted 1 times

YipingRuan 3 years, 7 months ago
Can be "Compute instance“Nope see the table at https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target#train
upvoted 2 times

YipingRuan 3 years, 7 months ago
Can be "Compute instance“
upvoted 2 times

a_sisbarra Highly Voted 3 years, 7 months ago
Nope see the table at https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target#train
upvoted 6 times

Question #42 *Topic 1*

You are making use of the Azure Machine Learning to designer construct an experiment.
After dividing a dataset into training and testing sets, you configure the algorithm to be Two-Class Boosted Decision Tree.
You are preparing to ascertain the Area Under the Curve (AUC).
Which of the following is a sequential combination of the models required to achieve your goal?

A. Train, Score, Evaluate.

B. Score, Evaluate, Train.

C. Evaluate, Export Data, Train.

D. Train, Score, Export Data.

Suggested Answer: *A*
Official Answer: A

Community vote distribution
A (100%)

synapse Highly Voted 4 years, 5 months ago
Answer is correct
upvoted 7 times

geethavkr Most Recent 2 years ago
Option A
upvoted 1 times

Roger_Sampaio1 2 years, 8 months ago
No Azure Machine Learning Designer seriam os módulos: 1 - Treinar. Ocorre literalmente o treinamento do modelo. 2 - Pontuar. Efetua-se as previsões com os dados de teste. 3 - Avaliar. Efetua-se a avaliação por meio das métricas. Resposta: A
upvoted 2 times

krishna1818 3 years, 2 months ago
Option A
upvoted 1 times

Question #43 *Topic 1*

DRAG DROP

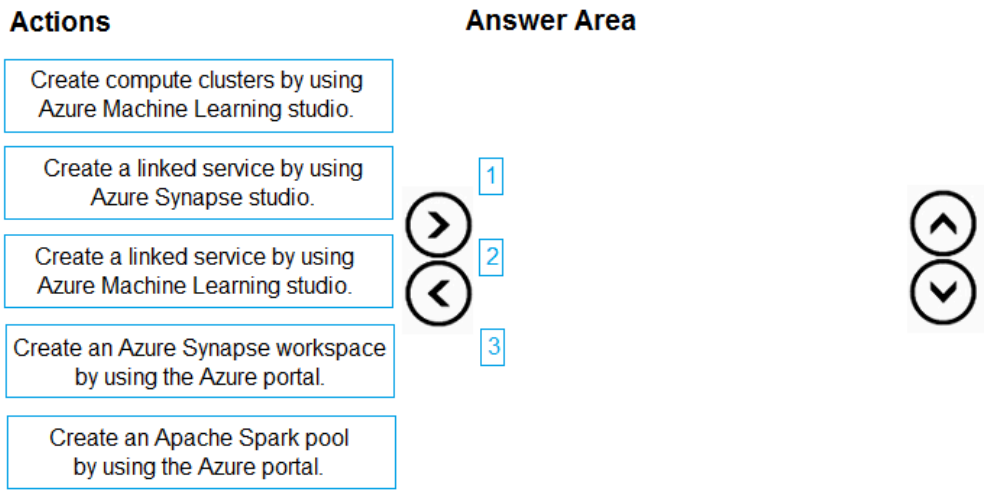
-

You create an Azure Machine Learning workspace.

You must implement dedicated compute for model training in the workspace by using Azure Synapse compute resources. The solution must attach the dedicated compute and start an Azure Synapse session.

You need to implement the computer resources.

Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Suggested Answer:

Answer Area

1. Create an Azure Synapse workspace by using the Azure portal.
2. Create an Apache Spark pool by using the Azure portal.
3. Create a linked service by using Azure Machine Learning studio.

Answer Area

1. Create an Azure Synapse workspace by using the Azure portal.
2. Create an Apache Spark pool by using the Azure portal.
3. Create a linked service by using Azure Machine Learning studio.

PatAms538 Most Recent 1 year ago

Was in the exam 31/07/25

upvoted 1 times

PCRamirez 2 years, 2 months ago

According to Copilot Pro: D) Create an Azure Synapse Workspace by using the Azure Portal: This is the first step because you need to have an Azure Synapse Workspace set up before you can link it to Azure Machine Learning or create compute resources within it1. E) Create an Apache Spark pool by using the Azure Portal: Once the Azure Synapse Workspace is created, you should create an Apache Spark pool within it. This pool will provide the compute resources needed for model training2. B) Create a linked service by using Azure Synapse Studio: After setting up the Azure Synapse Workspace and the Apache Spark pool, you need to create a linked service in Azure Synapse Studio to connect the Azure Synapse Workspace with the Azure Machine Learning workspace.

upvoted 2 times

Ratz 3 years ago

However, a new Synapse workspace can no longer be registered with Azure Machine Latest Update: Learning as a linked service. We recommend using Managed (Automatic) Synapse compute and attached Synapse Spark pools available in CLI v2 and Python SDK v2. Please see https://aka.ms/aml-spark for more details.

upvoted 2 times

naruto1012 3 years, 3 months ago

The answer is correct from this link https://learn.microsoft.com/en-us/azure/machine-learning/v1/how-to-link-synapse-ml-workspaces?view=azureml-api-1

upvoted 2 times

naruto1012 3 years, 3 months ago

The answer is correct from this link https://learn.microsoft.com/en-us/azure/machine-learning/v1/how-to-link-synapse-ml-workspaces?view=azureml-api-1That's outdated

upvoted 2 times

daviduzo 3 years, 1 month ago

That's outdated

upvoted 2 times

Tommo565 3 years, 4 months ago

Answer is correct (I think)

upvoted 1 times

Jin_22 3 years, 5 months ago

I think the correct answer should be: Create an Azure Synapse workspace Create an Azure Synapse Compute Attach the Compute to Azure ML Workspace

upvoted 4 times

Question #1 *Topic 2*

You are developing a hands-on workshop to introduce Docker for Windows to attendees.
You need to ensure that workshop attendees can install Docker on their devices.
Which two prerequisite components should attendees install on the devices? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Microsoft Hardware-Assisted Virtualization Detection Tool

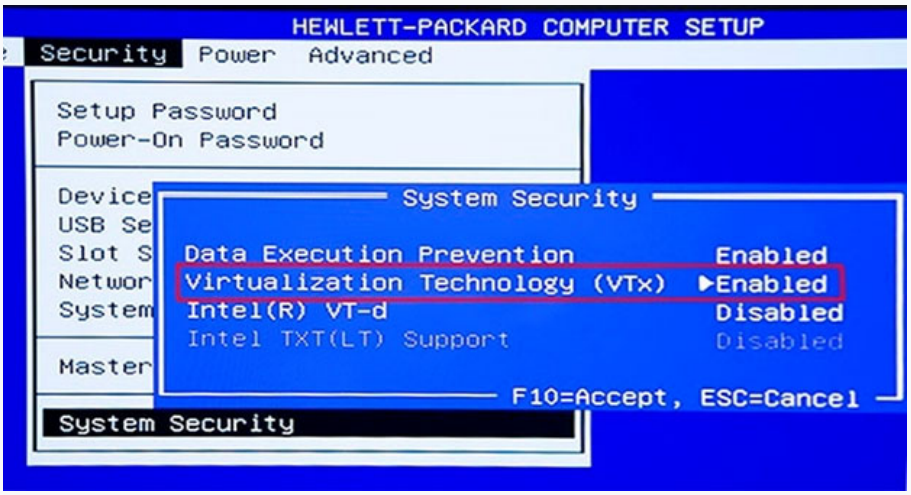
B. Kitematic

C. BIOS-enabled virtualization

D. VirtualBox

E. Windows 10 64-bit Professional

Suggested Answer: *C, E*





C: Make sure your Windows system supports Hardware Virtualization Technology and that virtualization is enabled. Ensure that hardware virtualization support is turned on in the BIOS settings. For example:

E: To run Docker, your machine must have a 64-bit operating system running Windows 7 or higher. Reference: https://docs.docker.com/toolbox/toolbox_install_windows/ https://blogs.technet.microsoft.com/canitpro/2015/09/08/step-by-step-enabling-hyper-v-for-use-on-windows-10/

Community vote distribution

CE (100%)

fredgu Highly Voted 3 years, 2 months ago

Just took the DP-100 test today, Dec 5, 2020, and found that none of the questions or knowledge points from page 1 to page 26 have been tested or mentioned. Nearly 95% of the questions tested in the exam were about Azure Data Science Service. And you are suggested to type codes to get familiar with the new coding style instead of wasting time on researching ML Studio (Classic) here.

upvoted 13 times

fredgu Highly Voted 3 years, 2 months ago

Just took the DP-100 test today, Dec 5, 2020, and found that none of the questions or knowledge points from page 1 to page 26 have been tested or mentioned. Nearly 95% of the questions tested in the exam were about Azure Data Science Service. And you are suggested to type codes to get familiar with the new coding style instead of wasting time on researching ML Studio (Classic) here.Can you clarify? what do you mean by the new coding style or Azure Data Science Service.Same with @fredgu, mostly it use new azure learning studio (ml.azure.com)can you be more clear of what consists P1 to P26? Which questions are these?I totally disagree with @fredgu. I took the test on the last week of May 2021 and most but not all questions were in the exam. I Used this as revision and I did well in the exam.I did pass with excellent over 900 scores with help here. But expect questions outside here. I used this practice as style of the test. I should say questions some times similar but not exact.

upvoted 13 times

SamiMelke 3 years, 1 month ago

Can you clarify? what do you mean by the new coding style or Azure Data Science Service.

upvoted 4 times

SamiMelke 3 years, 1 month ago

Can you clarify? what do you mean by the new coding style or Azure Data Science Service.Same with @fredgu, mostly it use new azure learning studio (ml.azure.com)

upvoted 4 times

Edhotp 3 years, 1 month ago

Same with @fredgu, mostly it use new azure learning studio (ml.azure.com)

upvoted 4 times

woyaodp100 2 years, 10 months ago

can you be more clear of what consists P1 to P26? Which questions are these?

upvoted 2 times

Winny 2 years, 8 months ago

I totally disagree with @fredgu. I took the test on the last week of May 2021 and most but not all questions were in the exam. I Used this as revision and I did well in the exam.

upvoted 6 times

Wayne888 2 years, 2 months ago

I did pass with excellent over 900 scores with help here. But expect questions outside here. I used this practice as style of the test. I should say questions some times similar but not exact.

upvoted 1 times

ning Most Recent 1 year, 9 months ago

Documentation mentioned two things: virtualization and 64 bits operation system; mapping to answers will be C / E

upvoted 2 times

austin06112000 1 year, 11 months ago

Answer is right. by Mar 4,2022.

upvoted 2 times

leraS28 2 years ago

I think we need virtualization, because the Students use Windows Machine, and we need Linux for machine learning. In Docker you can only use the same operating system, that means that first we need to install a Linux VM and Docker on top. Otherwise we should not need virtualizationI found even better explanation here https://sloopstash.com/blog/can-docker-containers-run-on-any-operating-system.html

upvoted 1 times

leraS28 2 years ago

I think we need virtualization, because the Students use Windows Machine, and we need Linux for machine learning. In Docker you can only use the same operating system, that means that first we need to install a Linux VM and Docker on top. Otherwise we should not need virtualization

upvoted 1 times

leraS28 2 years ago

I found even better explanation here https://sloopstash.com/blog/can-docker-containers-run-on-any-operating-system.html

upvoted 1 times

Akki0120 2 years, 7 months ago

If anyone wants all questions ping me 9403778084Need your help with questions.can you help me too?HI Akki, Please help me with the questions

upvoted 1 times

bikydrum 1 year, 12 months ago

Need your help with questions.

upvoted 1 times

Sandhya_Soman 2 years, 1 month ago

HI Akki, Please help me with the questions

upvoted 1 times

Akki0120 2 years, 7 months ago

If anyone wants all questions ping me 9403778084can you help me too?can you help me too?hi Akki , Can you help me with the Question Sets

upvoted 4 times

Akki0120 2 years, 7 months ago

If anyone wants all questions ping me 9403778084

upvoted 4 times

morojenie 2 years, 5 months ago

can you help me too?

upvoted 1 times

unonw 2 years, 5 months ago

can you help me too?

upvoted 1 times

shaimag 2 years, 6 months ago

hi Akki , Can you help me with the Question Sets

upvoted 2 times

abhijmk 2 years, 7 months ago

C&E are absolutely correct , for more details can refer here https://docs.docker.com/docker-for-windows/install-windows-home

upvoted 2 times

narendralv 2 years, 8 months ago

C & E are correct. Please follow the link for more information https://docs.docker.com/docker-for-windows/install/

upvoted 4 times

prashantjoge 2 years, 9 months ago

azure databricks support clustering while azure data factory supports orchestration (https://docs.microsoft.com/en-us/azure/databricks/clusters/configure). The orchestration here should be in the context of data processing (think SSIS, ETL, informatica etc.) Answer should be B. Azure containers instances provide some basic orchestration capabilities, but then again the context is different. https://docs.microsoft.com/en-us/azure/container-instances/container-instances-orchestrator-relationship

upvoted 1 times

dev2dev 2 years, 11 months ago

I think its c and d. because the question is asking about the components. how OS is a component?VirtualBox is not required for Docker. C is correct, I would assume the OS is already installed, but anyway I think the answer is correct, just tricky

upvoted 2 times

dev2dev 2 years, 11 months ago

I think its c and d. because the question is asking about the components. how OS is a component?

upvoted 2 times

levm39 2 years, 10 months ago

VirtualBox is not required for Docker. C is correct, I would assume the OS is already installed, but anyway I think the answer is correct, just tricky

upvoted 1 times

foorx 3 years, 8 months ago

Docker can now run on Win 10 Home with WSL2 now?Yes, https://docs.docker.com/docker-for-windows/install-windows-home/

upvoted 4 times

foorx 3 years, 8 months ago

Docker can now run on Win 10 Home with WSL2 now?

upvoted 4 times

Parzival 3 years, 6 months ago

Yes, https://docs.docker.com/docker-for-windows/install-windows-home/

upvoted 2 times

Question #2 *Topic 2*

Your team is building a data engineering and data science development environment.
The environment must support the following requirements:
✑ support Python and Scala
✑ compose data storage, movement, and processing services into automated data pipelines
✑ the same tool should be used for the orchestration of both data engineering and data science
✑ support workload isolation and interactive workloads
✑ enable scaling across a cluster of machines
You need to create the environment.
What should you do?

A. Build the environment in Apache Hive for HDInsight and use Azure Data Factory for orchestration.

B. Build the environment in Azure Databricks and use Azure Data Factory for orchestration.

C. Build the environment in Apache Spark for HDInsight and use Azure Container Instances for orchestration.

D. Build the environment in Azure Databricks and use Azure Container Instances for orchestration.

Suggested Answer: *B*
In Azure Databricks, we can create two different types of clusters. ✑ Standard, these are the default clusters and can be used with Python, R, Scala and SQL ✑ High-concurrency Azure Databricks is fully integrated with Azure Data Factory. Incorrect Answers: D: Azure Container Instances is good for development or testing. Not suitable for production workloads. Reference: https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/data-science-and-machine-learning

Community vote distribution
B (100%)

Adi06 Highly Voted 4 years, 9 months ago
Is the answer not D?? They are trying to build a development environment (line 1). Nowhere it says its for production environment.
upvoted 9 times

Adi06 Highly Voted 4 years, 9 months ago
Is the answer not D?? They are trying to build a development environment (line 1). Nowhere it says its for production environment.Not D. There is no autoscaling with ACIdefinitely dAgreed. If it was production environment, it should be Kubernetes services. Since it's development it should be container services. https://docs.microsoft.com/en-us/learn/modules/register-and-deploy-model-with-amls/2-deploy-modelyou cant do orchestration with ACI, only with data factory, answer is correct.
upvoted 9 times

strikchao 4 years ago
Not D. There is no autoscaling with ACI
upvoted 1 times

prashantjoge 4 years, 2 months ago
definitely d
upvoted 2 times

allanm 4 years, 3 months ago
Agreed. If it was production environment, it should be Kubernetes services. Since it's development it should be container services. https://docs.microsoft.com/en-us/learn/modules/register-and-deploy-model-with-amls/2-deploy-modelyou cant do orchestration with ACI, only with data factory, answer is correct.
upvoted 1 times

allanm 4 years, 3 months ago
Agreed. If it was production environment, it should be Kubernetes services. Since it's development it should be container services.

https://docs.microsoft.com/en-us/learn/modules/register-and-deploy-model-with-amls/2-deploy-model
upvoted 1 times

levm39 4 years, 2 months ago
you cant do orchestration with ACI, only with data factory, answer is correct.

upvoted 8 times

phdykd 2 years, 6 months ago

B. Azure Databricks is a fast, easy, and collaborative Apache Spark-based analytics platform that provides a complete environment for data engineering, machine learning, and data science. It supports Python and Scala, and allows you to compose data storage, movement, and processing services into automated data pipelines. Azure Data Factory, on the other hand, is a cloud-based data integration service that allows you to create, schedule, and orchestrate your data pipelines. By using both Databricks and Data Factory together, you can have a unified platform for both data engineering and data science that also supports workload isolation and interactive workloads, as well as enables scaling across a cluster of machines.

upvoted 8 times

phydev Most Recent 2 years, 1 month ago

Azure Databricks is an obvious choice for the environment. To decide between Data Factory vs Container Instances for orchestration makes the difference here. ADF would be a more suitable choice compared to ACI for orchestration in this scenario due to 1. Native Orchestration Capabilities 2. Visual Workflow Designer 3. Integration with Diverse Data Sources and Services 4. Built-in Monitoring and Management Therefore, for the purpose of orchestrating data pipelines in a data engineering and data science environment, ADF would be the recommended choice due to its dedicated orchestration features, data integration capabilities, visual workflow designer, and integration with diverse data sources and services.

upvoted 4 times

dija123 3 years, 8 months ago

B without doubts

upvoted 3 times

kolakone 4 years, 1 month ago

B is the right answer. C and D are out as there is need for data engineering Since there is need for "both data engineering and data science", there is need for Data Factory, hence C and D are out. Due to need for Scala and Python support, Databricks (B) is the correct answer.

upvoted 6 times

Navishmamta1111111111111 4 years, 1 month ago

B is correct

upvoted 3 times

okeyken1 4 years, 1 month ago

the correct answer is B

upvoted 1 times

MAGGCol 4 years, 2 months ago

Previewed last year, Microsoft's Azure Container Instances (ACI) is now ready for production usage, according to the company. ... Microsoft promises an uptime service level agreement of 99.9 percent for any container group. Each container is secured and isolated through a VM hypervisor.

upvoted 1 times

prashantjoge 4 years, 3 months ago

azure databricks support clustering while azure data factory supports orchestration (https://docs.microsoft.com/en-us/azure/databricks/clusters/configure). The orchestration here should be in the context of data processing (think SSIS, ETL, informatica etc.) Answer should be B. Azure containers instances provide some basic orchestration capabilities, but then again the context is different. https://docs.microsoft.com/en-us/azure/container-instances/container-instances-orchestrator-relationship

upvoted 5 times

chaudha4 4 years, 3 months ago

I think the best answer is B. ACI is used to deploy a model. ACI is just like docker - for orchestration you would need something like kubernetes not docker.

upvoted 5 times

LakeSky 4 years, 4 months ago

Wow, so what's the correct answer really? Why is Azure Container not an option?I think Azure Container instances cannot do orchestration

upvoted 2 times

LakeSky 4 years, 4 months ago

Wow, so what's the correct answer really? Why is Azure Container not an option?

upvoted 2 times

cab123 4 years, 4 months ago

I think Azure Container instances cannot do orchestration

upvoted 2 times

Question #3 *Topic 2*

DRAG DROP -

You are building an intelligent solution using machine learning models.

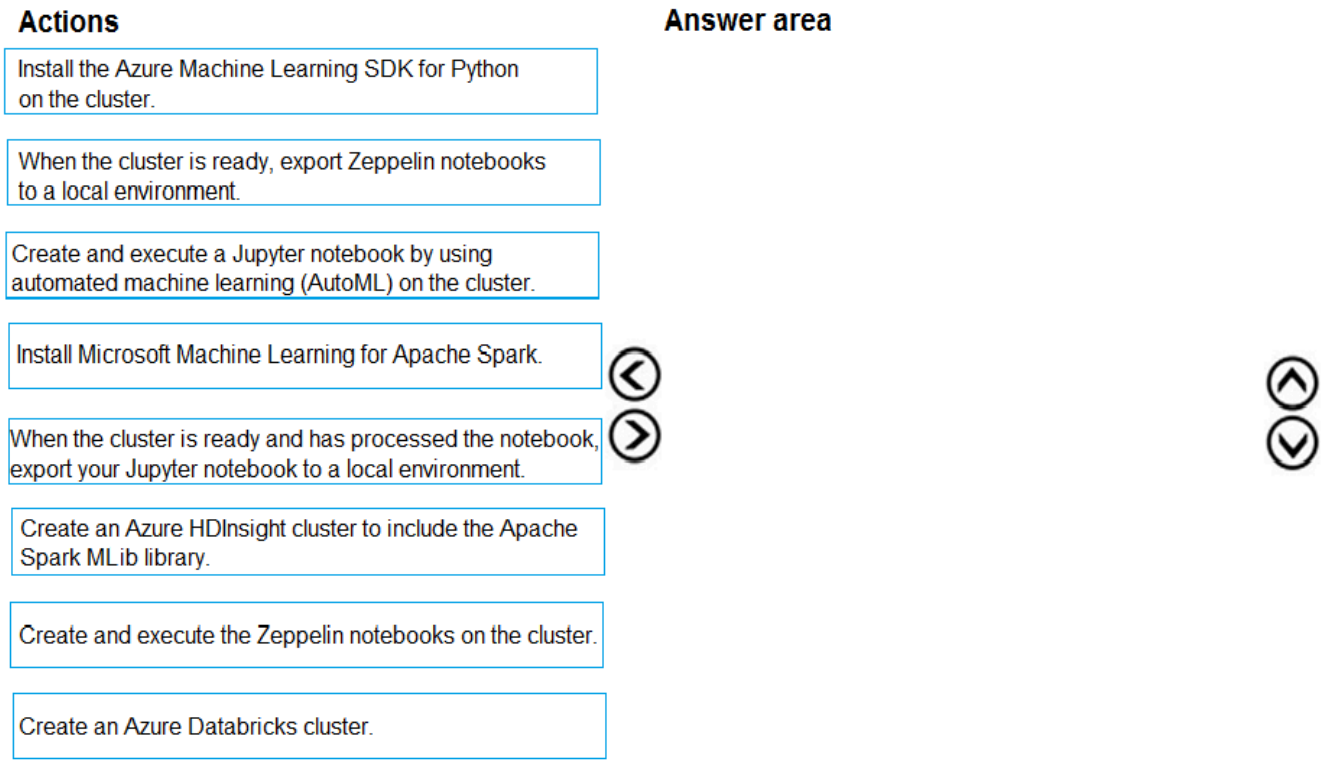
The environment must support the following requirements:

- ✑ Data scientists must build notebooks in a cloud environment
- ✑ Data scientists must use automatic feature engineering and model building in machine learning pipelines.
- ✑ Notebooks must be deployed to retrain using Spark instances with dynamic worker allocation.
- ✑ Notebooks must be exportable to be version controlled locally.

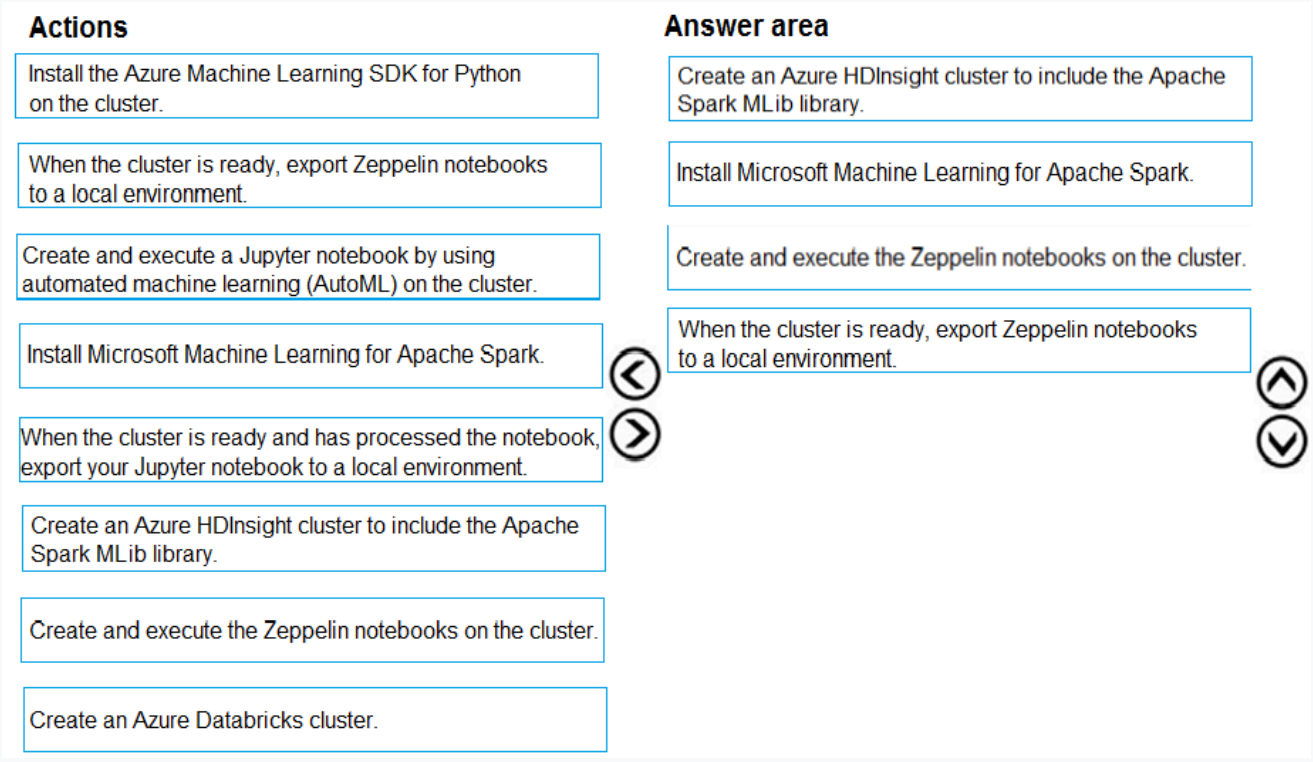
You need to create the environment.

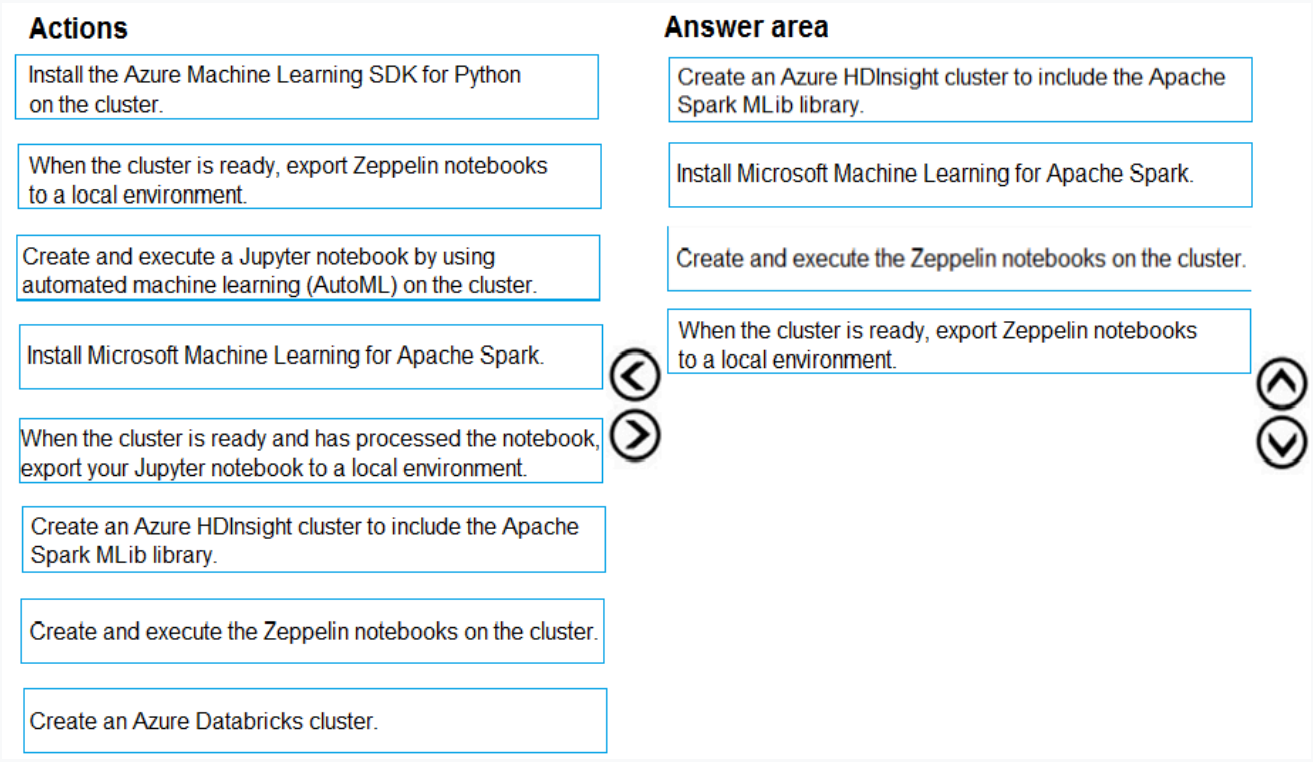
Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Select and Place:



Suggested Answer:





Step 1: Create an Azure HDInsight cluster to include the Apache Spark Mlib library Step 2: Install Microsot Machine Learning for Apache Spark You install AzureML on your Azure HDInsight cluster. Microsoft Machine Learning for Apache Spark (MMLSpark) provides a number of deep learning and data science tools for Apache Spark, including seamless integration of Spark Machine Learning pipelines with Microsoft Cognitive Toolkit (CNTK) and OpenCV, enabling you to quickly create powerful, highly-scalable predictive and analytical models for large image and text datasets. Step 3: Create and execute the Zeppelin notebooks on the cluster Step 4: When the cluster is ready, export Zeppelin notebooks to a local environment. Notebooks must be exportable to be version controlled locally. Reference: https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-zeppelin-notebook https://azuremlbuild.blob.core.windows.net/pysparkapi/intro.html

Dasist Highly Voted 5 years, 4 months ago

Should be: Create Azure Databricks cluster -> Install Azure ML SDK for Python -> Create and exec Jupyter notebook using AutoML -> Export Jupyter to local env. That because you need auto feature engineering provided by autoMLincase anyone still has doubts about this, refer:

https://docs.microsoft.com/en-us/azure/machine-learning/how-to-configure-databricks-automl-environmentdont see the option for install ML SDK for python for databricksyou can add ML lib using a script action when you create the HDinsight service. databricks-rg-azdatabrickspt-fkd2ogyzogbagNotebooks must be deployed to retrain using Spark instances with dynamic worker allocation- This condition won't be satisfied with Jupyter

upvoted 51 times

Dasist Highly Voted 5 years, 4 months ago

Should be: Create Azure Databricks cluster -> Install Azure ML SDK for Python -> Create and exec Jupyter notebook using AutoML -> Export Jupyter to local env. That because you need auto feature engineering provided by autoML

upvoted 51 times

spaceykacey 4 years, 9 months ago

incase anyone still has doubts about this, refer: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-configure-databricks-automl-environment

upvoted 3 times

prashantjoge 5 years, 3 months ago

dont see the option for install ML SDK for python for databricksyou can add ML lib using a script action when you create the HDinsight service. databricks-rg-azdatabrickspt-fkd2ogyzogbag

upvoted 3 times

prashantjoge 5 years, 3 months ago

dont see the option for install ML SDK for python for databricks

upvoted 3 times

prashantjoge 5 years, 3 months ago

you can add ML lib using a script action when you create the HDinsight service. databricks-rg-azdatabrickspt-fkd2ogyzogbag

upvoted 3 times

bruce 5 years, 4 months ago

Notebooks must be deployed to retrain using Spark instances with dynamic worker allocation- This condition won't be satisfied with Jupyter

upvoted 5 times

ajay_1233456 3 years, 12 months ago

1. Create Azure Databricks cluster 2. Install Azure ML SDK for Python 3. Create and exec Jupyter notebook using AutoML 4. Export Jupyter to local env

upvoted 13 times

OdaNabunaga 2 years, 1 month ago

1. Create an Azure Databricks cluster 2. Install the Azure Machine Learning SDK for Python on the cluster 3. Create and execute a Jupyter notebook by using automated machine learning (AutoML) on the cluster 4. When the cluster is ready and has processed the notebook, export your Jupyter notebook to a local environmentI think it's most accurate

upvoted 3 times

OdaNabunaga 2 years, 1 month ago

1. Create an Azure Databricks cluster 2. Install the Azure Machine Learning SDK for Python on the cluster 3. Create and execute a Jupyter notebook by using automated machine learning (AutoML) on the cluster 4. When the cluster is ready and has processed the notebook, export your Jupyter notebook to a local environment

upvoted 3 times

larimalarima Most Recent 2 years, 1 month ago

I think it's most accurate

upvoted 2 times

PI_Team 3 years, 1 month ago

Create an Azure Databricks cluster to provide a cloud environment for data scientists to build their notebooks. Install the Azure ML SDK for Python on the cluster to enable data scientists to use automatic feature engineering and model building in machine learning pipelines. Create and execute the Zeppelin notebooks on the cluster to build and train machine learning models using Spark instances with dynamic worker allocation. When the cluster is ready, export Zeppelin notebooks to a local environment to enable version control of the notebooks locally. SaM

upvoted 2 times

phdykd 3 years, 6 months ago

Here is the most accurate sequence of actions for creating the desired environment: Create an Azure Databricks cluster Install Microsoft Machine Learning for Apache Spark on the cluster Create and execute Jupyter notebooks using AutoML on the cluster When the cluster is ready and has processed the notebook, export your Jupyter notebook to a local environment for version control. This sequence of actions will allow you to take advantage of the Azure Databricks platform for cloud-based data processing, and the Microsoft Machine Learning for Apache Spark library for

automating feature engineering and model building in your Jupyter notebooks. Additionally, exporting the notebooks to a local environment will allow you to version control them and collaborate with other team members.

upvoted 7 times

shubhangi2612 3 years, 7 months ago

https://industry40.co.in/azure-hdinsight-and-azure-databricks/

upvoted 3 times

ning 4 years, 3 months ago

Totally agree 1. Create Azure Databricks cluster 2. Install Azure ML SDK for Python 3. Create and exec Jupyter notebook using AutoML 4. Export Jupyter to local env

upvoted 4 times

DingDongSingSong 4 years, 4 months ago

Reference this link: https://docs.microsoft.com/en-us/azure/machine-learning/concept-automated-ml Answer is as per DASIST noted: 1. Create Azure Databricks cluster 2. Install Azure ML SDK for Python 3. Create and exec Jupyter notebook using AutoML 4. Export Jupyter to local env Another link that supports this rationale is :https://industry40.co.in/azure-hdinsight-and-azure-databricks/. It clearly outlines why for Spark based environments, Databricks is a better option than HDInsight

upvoted 3 times

ajayjha123 4 years, 8 months ago

Should be: Create Azure Databricks cluster -> Install Azure ML SDK for Python -> Create and exec Jupyter notebook using AutoML -> Export Jupyter to local env. That because you need auto feature engineering provided by autoML

upvoted 3 times

RyanTsai 4 years, 11 months ago

agree: Create Azure Databricks cluster -> Create and exec Jupyter notebook using AutoML -> Install Azure ML SDK for Python -> Export Jupyter to local env

upvoted 4 times

dija123 5 years ago

Create Azure Databricks cluster -> Create and exec Jupyter notebook using AutoML -> Install Azure ML SDK for Python -> Export Jupyter to local env

upvoted 3 times

Akki0120 5 years, 1 month ago

If anyone wants all questions ping me 9403778084Can you please share the email

upvoted 3 times

Akki0120 5 years, 1 month ago

If anyone wants all questions ping me 9403778084

upvoted 3 times

Lutendo 4 years, 11 months ago

Can you please share the email

upvoted 1 times

tamoor 5 years, 6 months ago

you can use only azure hdinsights because of condition you can use only apache-spark for data bricks, you must use Hadoop.

upvoted 1 times

dzzz 5 years, 8 months ago

I believe Data Bricks is capable, but if you choose that as first step, there is no further actions can be chosen, which all around Zeppelin, but Data Bricks doesn't support Zeppelin.https://docs.microsoft.com/en-us/azure/databricks/dev-tools/databricks-connect#:~:text=Databricks%20Connect%20allows%20you%20to,applications%20to%20Azure%20Databricks%20clusters. From the above link, it is seen that Databricks can support Zepplin.data bricks is a an analytics platform. it does not support feature engineering

upvoted 3 times

dzzz 5 years, 8 months ago

I believe Data Bricks is capable, but if you choose that as first step, there is no further actions can be chosen, which all around Zeppelin, but Data Bricks doesn't support Zeppelin.

upvoted 3 times

Srivathsan 5 years, 6 months ago

https://docs.microsoft.com/en-us/azure/databricks/dev-tools/databricks-connect#:~:text=Databricks%20Connect%20allows%20you%20to,applications%20to%20Azure%20Databricks%20clusters. From the above link, it is seen that Databricks can support Zepplin.

upvoted 1 times

prashantjoge 5 years, 3 months ago

data bricks is a an analytics platform. it does not support feature engineering

upvoted 1 times

valkyrieShadow 5 years, 8 months ago

Azure data bricks meets all the requirements. HDInsight does not. Example: automatic feature engineering is included with autoML. HDinsight does not include this feature.

HDinsight: https://docs.microsoft.com/en-us/azure/hdinsight/

Azure Databricks: https://docs.microsoft.com/en-us/azure/databricks/applications/machine-learning/automl-hyperparam-tuning/

upvoted 3 times

valkyrieShadow 5 years, 8 months ago

Azure data bricks meets all the requirements. HDInsight does not. Example: automatic feature engineering is included with autoML. HDinsight does not include this feature. HDinsight: https://docs.microsoft.com/en-us/azure/hdinsight/ Azure Databricks: https://docs.microsoft.com/en-us/azure/databricks/applications/machine-learning/automl-hyperparam-tuning/No, HDinsight also provides all the autoML and auto feature engineering features https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-run-machine-learning-automlthe given answer is correct

upvoted 3 times

HkIsCrazY 5 years, 6 months ago

No, HDinsight also provides all the autoML and auto feature engineering features https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-run-machine-learning-automlthe given answer is correct

upvoted 3 times

HkIsCrazY 5 years, 6 months ago

No, HDinsight also provides all the autoML and auto feature engineering features

https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-run-machine-learning-automl

upvoted 3 times

prashantjoge 5 years, 3 months ago

the given answer is correct

upvoted 3 times

Karen_12321 5 years, 9 months ago

Why not jupyter note book?

upvoted 1 times

Karen_12321 5 years, 9 months ago

Why not jupyter note book?Maybe because jupyter notebook don't provide intepreter for Spark like Zeppelin? https://medium.com/ankitakumar140494/a-comprehensive-comparison-between-jupyter-notebook-and-apache-zeppelin-911501981bfbNo, it supports too.

upvoted 1 times

LakeSky 5 years, 4 months ago

Maybe because jupyter notebook don't provide intepreter for Spark like Zeppelin? https://medium.com/ankitakumar140494/a-comprehensive-comparison-between-jupyter-notebook-and-apache-zeppelin-911501981bfbNo, it supports too.

upvoted 1 times

LakeSky 5 years, 4 months ago

Maybe because jupyter notebook don't provide intepreter for Spark like Zeppelin? https://medium.com/ankitakumar140494/a-comprehensive-comparison-between-jupyter-notebook-and-apache-zeppelin-911501981bfb

upvoted 1 times

zehraoneexam 4 years, 5 months ago

No, it supports too.

upvoted 1 times

sayak17 5 years, 11 months ago

first option can be Azure Databricks also right? It can also do all of the steps after that.

upvoted 2 times

sayak17 5 years, 11 months ago

first option can be Azure Databricks also right? It can also do all of the steps after that.Why would both spark mllib and mmlspark be required? I feel first option will be databricks

upvoted 2 times

sayak17 5 years, 11 months ago

Why would both spark mllib and mmlspark be required? I feel first option will be databricks

upvoted 1 times

Question #4 *Topic 2*

You plan to build a team data science environment. Data for training models in machine learning pipelines will be over 20 GB in size.
You have the following requirements:
➪ Models must be built using Caffe2 or Chainer frameworks.
➪ Data scientists must be able to use a data science environment to build the machine learning pipelines and train models on their personal devices in both connected and disconnected network environments.
Personal devices must support updating machine learning pipelines when connected to a network.
You need to select a data science environment.
Which environment should you use?

A. Azure Machine Learning Service

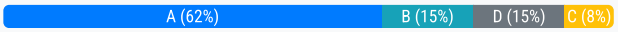
B. Azure Machine Learning Studio

C. Azure Databricks

D. Azure Kubernetes Service (AKS)

Suggested Answer: *A*
The Data Science Virtual Machine (DSVM) is a customized VM image on Microsoft's Azure cloud built specifically for doing data science. Caffe2 and Chainer are supported by DSVM. DSVM integrates with Azure Machine Learning. Incorrect Answers: B: Use Machine Learning Studio when you want to experiment with machine learning models quickly and easily, and the built-in machine learning algorithms are sufficient for your solutions. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview

Community vote distribution



chaudha4 Highly Voted 4 years, 3 months ago
It seems like the answer explanation is mixing Azure Machine Learning Studio with Azure Machine Learning Designer since the description is for designer not studio. Studio includes designer, notebook and autoML. I think the correct answer is actually B since Azure Machine Learning Service can only be used from within Azure Machine Learning Studio.
upvoted 14 times

Wayland Highly Voted 2 years, 11 months ago
I just did some digging and here is what I find: https://docs.microsoft.com/en-us/azure/machine-learning/overview-what-is-machine-learning-studio. The "Machine Learning Studio" in this question is actually referring to "Machine Learning Studio Classic", which is an outdated platform that only offering web service (no offline), and the "Machine Learning Service" is actually the "Azure Machine Learning" you can find in Azure Portal. What's the more the new Studio is now part of the "Azure Machine Learning" as we speak. So for this question, A is the correct answer at the time when it was firstly created, but it no long applies right now. Now, A and B are pretty much the same thing.
upvoted 10 times

sar77 Most Recent 1 year, 1 month ago
A. Azure Machine Learning Service Supports large datasets and custom frameworks (like Caffe2, Chainer via Docker). Allows local development and offline work using Azure ML SDK and local compute. Pipelines can be updated when reconnected. □ Best fit for all requirements.
upvoted 1 times

SaraGG28 1 year, 7 months ago
Azure Machine Learning Service is the most appropriate choice based on the requirements: Supports Caffe2 and Chainer frameworks: Azure Machine Learning Service allows flexibility in choosing machine learning frameworks, including Caffe2 and Chainer, through custom environments. Supports both connected and disconnected environments: With Azure Machine Learning Service, data scientists can work locally on their personal devices using the Azure ML SDK or CLI, even offline. Pipelines and models can be updated and synced when reconnected to the network. Large dataset support: The service can handle datasets larger than 20 GB by enabling integration with cloud storage (e.g., Azure Blob Storage) and local caching mechanisms for disconnected environments. Personal device flexibility: Azure Machine Learning Service supports local development and allows deploying models on personal devices with proper synchronization capabilities. Chat GPT
upvoted 1 times

uncleeeesam 1 year, 8 months ago
ChatGPT says D.
upvoted 1 times

Ran2025 1 year, 10 months ago

A is correct! Azure Machine Learning Service supports local machine compute!

upvoted 1 times

Yoshizn 2 years, 6 months ago

https://www.codit.eu/blog/azure-machine-learning-studio-vs-services/

upvoted 3 times

phdykd 2 years, 6 months ago

A. Azure Machine Learning Service would be the best option for building a team data science environment with the given requirements. It allows building machine learning pipelines using Caffe2 or Chainer frameworks, supports training models on personal devices in both connected and disconnected network environments, and provides a mechanism for updating machine learning pipelines when connected to a network.

upvoted 3 times

shubhangi2612 2 years, 7 months ago

on this link, I found the difference between azure ML service and studio and the conclusion is azure Ml service is hybrid(on cloud or premise)
https://www.codit.eu/blog/azure-machine-learning-studio-vs-services/?country_sel=uk

upvoted 1 times

Sibajene 2 years, 7 months ago

C is correct

upvoted 1 times

Edriv 2 years, 8 months ago

Option A

upvoted 1 times

Edriv 2 years, 8 months ago

option Ahttps://www.codit.eu/blog/azure-machine-learning-studio-vs-services/?country_sel=be

upvoted 1 times

Edriv 2 years, 8 months ago

https://www.codit.eu/blog/azure-machine-learning-studio-vs-services/?country_sel=be

upvoted 1 times

JTWang 2 years, 10 months ago

Answer is A. Because Azure Machine Learning Studio need network! Azure Machine Learning Service can support local compute.
https://www.codit.eu/blog/azure-machine-learning-studio-vs-services/?country_sel=be

upvoted 4 times

Nav727 3 years, 2 months ago

Selected Answer: B

Why is B incorrect??

upvoted 2 times

Nav727 3 years, 2 months ago

Why is B incorrect??The ML studio seems to refer to the Web UI ml.azure.com, which is unavailable in an offline setting.

upvoted 2 times

chevyli 2 years, 11 months ago

The ML studio seems to refer to the Web UI ml.azure.com, which is unavailable in an offline setting.

upvoted 2 times

turtle666 3 years, 3 months ago

answer should be A, but already out-dated ? https://www.codit.eu/blog/azure-machine-learning-studio-vs-services/?country_sel=be

upvoted 1 times

DingDongSingSong 3 years, 4 months ago

What's the consensus on the answer? Is it A or B? I cannot find any learning document that provides clarity on compatibility with caffe2 or chainer, and any information on off network local machine usage.

upvoted 2 times

spaceykacey 3 years, 9 months ago

I think the answer is A, Azure ML service is hybrid (can be run both on cloud and on premise) https://www.codit.eu/blog/azure-machine-learning-studio-vs-services/

upvoted 2 times

Question #5 *Topic 2*

You are implementing a machine learning model to predict stock prices.
The model uses a PostgreSQL database and requires GPU processing.
You need to create a virtual machine that is pre-configured with the required tools.
What should you do?

A. Create a Data Science Virtual Machine (DSVM) Windows edition.

B. Create a Geo AI Data Science Virtual Machine (Geo-DSVM) Windows edition.

C. Create a Deep Learning Virtual Machine (DLVM) Linux edition.

D. Create a Deep Learning Virtual Machine (DLVM) Windows edition.

Suggested Answer: *A*
In the DSVM, your training models can use deep learning algorithms on hardware that's based on graphics processing units (GPUs). PostgreSQL is available for the following operating systems: Linux (all recent distributions), 64-bit installers available for macOS (OS X) version 10.6 and newer ع€" Windows (with installers available for 64-bit version; tested on latest versions and back to Windows 2012 R2. Incorrect Answers: B: The Azure Geo AI Data Science VM (Geo-DSVM) delivers geospatial analytics capabilities from Microsoft's Data Science VM. Specifically, this VM extends the AI and data science toolkits in the Data Science VM by adding ESRI's market-leading ArcGIS Pro Geographic Information System. C, D: DLVM is a template on top of DSVM image. In terms of the packages, GPU drivers etc are all there in the DSVM image. Mostly it is for convenience during creation where we only allow DLVM to be created on GPU VM instances on Azure. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview

Community vote distribution
C (100%)

Avii123 Highly Voted 5 years, 2 months ago
Correct answer is DSVM on Linux as postgreSQL is only supported on Linux VMS. https://docs.microsoft.com/en-us/azure/virtual-machines/linux/postgresql-install
upvoted 27 times

chaudha4 Highly Voted 4 years, 9 months ago
I think that correct answer is A. I think this is a trick question. Option C and D are not even a valid option. Most people are reading it as "DSVM linux edition" while it is "DLVM" (not DSVM). There is almost no mention on DLVM anywhere in Azure docs. If there was an option called DSVM Linux edition, then that would have been the correct answer. The correct answer to this question is DVSM linux or windows edition.pgsql is supported on windows and linux. https://www.postgresqltutorial.com/install-postgresql/But isn't DSVM Windows edition pre-configured.DLVM is a specialized version of DSVM optimized for Deep Learning. DSVM for Windows doesn't come pre-configured with Postgres and I am unsure if it is supported.
upvoted 13 times

chaudha4 4 years, 9 months ago
I think that correct answer is A. I think this is a trick question. Option C and D are not even a valid option. Most people are reading it as "DSVM linux edition" while it is "DLVM" (not DSVM). There is almost no mention on DLVM anywhere in Azure docs. If there was an option called DSVM Linux edition, then that would have been the correct answer. The correct answer to this question is DVSM linux or windows edition.
upvoted 13 times

prashantjoge 4 years, 9 months ago
pgsql is supported on windows and linux. https://www.postgresqltutorial.com/install-postgresql/
upvoted 2 times

prashantjoge 4 years, 9 months ago
pgsql is supported on windows and linux. https://www.postgresqltutorial.com/install-postgresql/But isn't DSVM Windows edition pre-configured.
upvoted 2 times

trickerk 4 years, 6 months ago
But isn't DSVM Windows edition pre-configured.
upvoted 4 times

Rdninja 3 years ago
DLVM is a specialized version of DSVM optimized for Deep Learning. DSVM for Windows doesn't come pre-configured with Postgres and I am unsure if it is supported.

upvoted 1 times

sl_mslconsulting Most Recent 1 year, 8 months ago

https://learn.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/tools-included?view=azureml-api-2#build-deep-learning-and-machine-learning-solutions

upvoted 2 times

Vikyyy 2 years, 2 months ago

Some places it says DSVM windows doesn't support Postgres but Chat GPT says "PostgreSQL can run on any system, including both Windows and Linux, so it doesn't affect the choice here." I believe DLVM LInux or WInsdows DSVM are correct .

upvoted 1 times

ABosco 2 years, 6 months ago

DSVM supports PostgreSQL. So the answer is A. Link below: https://learn.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/linux-dsvm-walkthrough?view=azureml-api-2PostgreSQL in your link is only mentioned for Linux

upvoted 1 times

ABosco 2 years, 6 months ago

DSVM supports PostgreSQL. So the answer is A. Link below: https://learn.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/linux-dsvm-walkthrough?view=azureml-api-2

upvoted 1 times

BR_CS 2 years, 6 months ago

PostgreSQL in your link is only mentioned for Linux

upvoted 2 times

daviduzo 2 years, 8 months ago

The answer should be DSVM Linux edition. The Windows DSVM doesn't come pre-configured with Postgres. Man these options should be updated, I doubt the exam would exclude the DSVM Linux edition. C and D are not even correct either and B isn't useful here.

upvoted 1 times

phdykd 3 years ago

Moderator should provide correct answers! Answers are mess in this portal! C. Create a Deep Learning Virtual Machine (DLVM) Linux edition. DLVM is pre-configured with tools such as CUDA and cuDNN for GPU processing, and it is optimized for deep learning workloads. Additionally, DLVM supports PostgreSQL, which is required for the stock price prediction model. The Linux edition of DLVM would provide a cost-effective and reliable environment for running the GPU-intensive machine learning workloads.

upvoted 5 times

ning 3 years, 9 months ago

Only Windows and Linux Ubuntu two options when talking about machine / deep learning VMs; unless very specific MS / Windows products, e.g., SQL Server, otherwise always select Linux Ubuntu as your answerAgree - C

upvoted 4 times

ning 3 years, 9 months ago

Selected Answer: C

Only Windows and Linux Ubuntu two options when talking about machine / deep learning VMs; unless very specific MS / Windows products, e.g., SQL Server, otherwise always select Linux Ubuntu as your answer

upvoted 4 times

Edriv 3 years, 2 months ago

Agree - C

upvoted 1 times

DingDongSingSong 3 years, 10 months ago

According to this link: https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/linux-dsvm-walkthrough DSVM on Linux has PostgreSQL preinstalled. Windows does not. So the answer is C even though it says DLVM (which is nothing but a variation of DSVM with GPU, which is also the requirement per the question)

upvoted 2 times

trickerk 4 years, 6 months ago

Question #10 in next page is about DLVM. So the correct answer is C, because DSVM for Windows doesn't have pre-configured PostgreSQL. DLVM (DSVM framework) for Linux have pre-configured PostgreSQL.

upvoted 2 times

thhvancouver 4 years, 6 months ago

Windows DSVM appears to be correct: https://medium.com/@Hackyroot/azure-deep-learning-a7c1953e2542

upvoted 2 times

trickerk 4 years, 7 months ago

I believe the correct answer is C. Because the Microsoft documents don't mention PostgreSQL for DSVM for Windows. Just compare: https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/vm-do-ten-things https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/linux-dsvm-walkthrough

upvoted 1 times

Akki0120 4 years, 7 months ago

If anyone wants all questions ping me 9403778084

upvoted 3 times

Akki0120 4 years, 7 months ago

If anyone wants all questions ping me 9403778084Thanks a lot!Please

upvoted 3 times

cccwahtever 4 years, 1 month ago

Thanks a lot!

upvoted 1 times

Frank90T 4 years, 3 months ago

Please

upvoted 1 times

luisdanielse 5 years ago

Postgress is not available on DSVM windows. I'd go for C. Actually, another question on this set is related to postgress-Windows. The explanation on this other question is that Postgress cannot be used with windows vm

upvoted 4 times

luisdanielse 5 years ago

Postgress is not available on DSVM windows. I'd go for C. Actually, another question on this set is related to postgress-Windows. The explanation on this other question is that Postgress cannot be used with windows vmtotally agree!I think a Deep Learning Virtual Machine is not available anymore. Only option is DSVM.postgress cant be used with windows is a joke.

upvoted 4 times

vippi66 5 years ago

totally agree!

upvoted 2 times

treadst0ne 4 years, 12 months ago

I think a Deep Learning Virtual Machine is not available anymore. Only option is DSVM.

upvoted 4 times

dev2dev 4 years, 11 months ago

postgress cant be used with windows is a joke.

upvoted 1 times

Srivathsan 5 years ago

To all who feel confused, refer the below docs: https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview

It clearly states the DSVM windows edition means Windows Server 2019. Hence we shall say Option A is correct here. DSVM has GPU enabled [Refer "Deep learning with GPUs" section in the docs link]

As per the Docs: "The Windows editions of the DSVM come pre-installed with GPU drivers, frameworks, and GPU versions of deep learning frameworks. On the Linux editions, deep learning on GPUs is enabled on the Ubuntu DSVMs."

upvoted 7 times

Srivathsan 5 years ago

To all who feel confused, refer the below docs: https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview It clearly states the DSVM windows edition means Windows Server 2019. Hence we shall say Option A is correct here. DSVM has GPU enabled [Refer "Deep learning with GPUs" section in the docs link] As per the Docs: "The Windows editions of the DSVM come pre-installed with GPU drivers, frameworks, and GPU versions of deep learning frameworks. On the Linux editions, deep learning on GPUs is enabled on the Ubuntu DSVMs."Sorry, not correct. The requirement is for GPU + PostgreSQL. Both of those are available on DSVM on Linux, not on Windows. Answer is C

upvoted 7 times

DingDongSingSong 3 years, 10 months ago

Sorry, not correct. The requirement is for GPU + PostgreSQL. Both of those are available on DSVM on Linux, not on Windows. Answer is C

upvoted 1 times

nepketo 5 years, 2 months ago

The answer is C since we need Deep Learning VM for GPU capability and Linux to support PostgreSQL.

upvoted 6 times

ipindado2020 5 years, 3 months ago

Agree with A: B-> Customized for Geospatial Analytics C & D-> Customized for Deep Learning (DLVM)

upvoted 1 times

Question #6 *Topic 2*

You are developing deep learning models to analyze semi-structured, unstructured, and structured data types.
You have the following data available for model building:
✑ Video recordings of sporting events
✑ Transcripts of radio commentary about events
✑ Logs from related social media feeds captured during sporting events
You need to select an environment for creating the model.
Which environment should you use?

A. Azure Cognitive Services

B. Azure Data Lake Analytics

C. Azure HDInsight with Spark MLib

D. Azure Machine Learning Studio

Suggested Answer: *A*
Azure Cognitive Services expand on Microsoft's evolving portfolio of machine learning APIs and enable developers to easily add cognitive features ג€" such as emotion and video detection; facial, speech, and vision recognition; and speech and language understanding ג€" into their applications. The goal of Azure Cognitive Services is to help developers create applications that can see, hear, speak, understand, and even begin to reason. The catalog of services within Azure Cognitive Services can be categorized into five main pillars - Vision, Speech, Language, Search, and Knowledge. Reference: https://docs.microsoft.com/en-us/azure/cognitive-services/welcome

Community vote distribution



Herr Highly Voted 6 years ago
The answer should be "C", the question talks about an environment for model building. Cognitive services is primarily for those with little or no DS skills (Azure Cognitive Services are APIs, SDKs, and services available to help developers build intelligent applications without having direct AI or data science skills or knowledge)
upvoted 35 times

Herr Highly Voted 6 years ago
The answer should be "C", the question talks about an environment for model building. Cognitive services is primarily for those with little or no DS skills (Azure Cognitive Services are APIs, SDKs, and services available to help developers build intelligent applications without having direct AI or data science skills or knowledge)Agree - CI agree.
upvoted 35 times

Edriv 3 years, 8 months ago
Agree - C
upvoted 1 times

treadstOne 5 years, 5 months ago
I agree.
upvoted 4 times

levm39 5 years, 2 months ago
Answer is A: explanation is here: https://docs.microsoft.com/en-us/azure/cognitive-services/big-data/cognitive-services-for-big-data
upvoted 11 times

Guru_1 Most Recent 8 months ago
D. Azure Machine Learning Studio Azure Machine Learning Studio is the correct environment for this task because it is a comprehensive platform designed for building, deploying, and managing machine learning models, including deep learning, using all data types (structured, semi-structured, and unstructured). It provides the necessary tools and flexibility to handle complex data like video, transcripts, and social media logs for advanced model training.
upvoted 1 times

sar77 1 year, 1 month ago
Azure Machine Learning Studio Supports custom deep learning models. Can handle structured, semi-structured, and unstructured data. Integrates with PyTorch, TensorFlow, Keras, and supports GPU compute. □ Best choice for building and training deep learning models on diverse data types.

upvoted 1 times

f82411e 1 year, 3 months ago

Azure Cognitive Services Usa modelos ya entrenados. Útil para tareas simples, pero no sirve para entrenar modelos personalizados complejos. B. Azure Data Lake Analytics Es una solución para consultas y análisis de datos a gran escala, no para entrenar modelos. Azure HDInsight con Spark MLlib Bueno para ML tradicional en grandes volúmenes de datos, pero limitado para deep learning y datos no estructurados como video. La respuesta es la D

upvoted 1 times

jl420 1 year, 9 months ago

While Azure Cognitive Services is a powerful suite of APIs and services for adding AI capabilities to applications, it is more focused on providing pre-built models for specific tasks such as vision, speech, language, and decision-making. These services are excellent for integrating AI features into applications quickly and easily, but they may not offer the same level of flexibility and control needed for developing custom deep learning models from scratch. On the other hand, Azure Machine Learning Studio provides a comprehensive environment for developing, training, and deploying custom machine learning models, including deep learning models. It supports various data types and offers tools for data preparation, model training, and deployment, making it a more suitable choice for your needs.

upvoted 3 times

jefimija 1 year, 9 months ago

Azure Cognitive Services is for using prebuilt models, not for training

upvoted 1 times

sl_mslconsulting 2 years, 2 months ago

I would pick D and ChapGPT 4 agreed with me. I wrote AI-102 (the updated one) a while back. AI cognitive service is not what you think it is as people discussed here.

upvoted 2 times

Shariq 2 years, 2 months ago

Should not it be Azure AI Service?

upvoted 1 times

prabhjot 2 years, 6 months ago

ANS D - Azure Machine Learning Studio: Azure Machine Learning Studio provides a comprehensive environment for building, training, and deploying machine learning models, including deep learning models. It supports various data types, and you can use popular deep learning frameworks like TensorFlow and PyTorch. This option allows for a more integrated and customizable approach to deep learning model development.

upvoted 1 times

Ran2025 2 years, 10 months ago

A is correct!

upvoted 1 times

phydev 3 years, 1 month ago

ChatGPT says A.

upvoted 2 times

PopeyeDS 3 years, 1 month ago

question talks about the training hence we should go with A

upvoted 1 times

MarinaMijailovic 3 years, 1 month ago

It seems this isnt DP 100 question, but AI 102.

upvoted 2 times

MarinaMijailovic 3 years, 1 month ago

Can we build models with Cognitive services?

upvoted 1 times

SoftAI 3 years, 4 months ago

i vote for cognitive services, simply because spark is not a better alternative in my opinion. real time processing is not mentioned anywhere

upvoted 2 times

sahithi2004 3 years, 5 months ago

Azure HDInsight with Spark MLib, you can easily build and deploy machine learning models for analysing semi-structured, unstructured, and structured data types. It provides a scalable computing environment that can handle large amounts of data, and you can use various programming languages such as Python, R, and Scala to develop your models. Azure Cognitive Services is a cloud-based platform that provides various APIs and pre-built models for developing applications with natural language processing, computer vision, and speech recognition capabilities. While it can be used to analyse unstructured data such as text and images, it may not be the most suitable choice for analysing semi-structured and structured data such as video recordings and social media logs.

upvoted 3 times

Question #7 *Topic 2*

You must store data in Azure Blob Storage to support Azure Machine Learning.
You need to transfer the data into Azure Blob Storage.
What are three possible ways to achieve the goal? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. Bulk Insert SQL Query

B. AzCopy

C. Python script

D. Azure Storage Explorer

E. Bulk Copy Program (BCP)

Suggested Answer: *B, C, D*
You can move data to and from Azure Blob storage using different technologies: ➪ Azure Storage-Explorer ➪ AzCopy ➪ Python ➪ SSIS
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/move-azure-blob

Community vote distribution
BCD (100%)

dija123 Highly Voted 4 years, 2 months ago
BCD are correct
upvoted 8 times

ipindado2020 Highly Voted 5 years, 3 months ago
Agree with BCD
upvoted 5 times

evangelist Most Recent 1 year, 12 months ago
python+AzCopy+Storage Explorer
upvoted 1 times

huyastromodeus 2 years, 5 months ago
dejavu
upvoted 1 times

krishna1818 2 years, 8 months ago
and SIS
upvoted 1 times

Depayser 2 years, 9 months ago
BCD Correctos
upvoted 1 times

phdykd 3 years ago
B. AzCopy C. Python script D. Azure Storage Explorer B. AzCopy is a command-line utility that is used to transfer data to and from Azure Blob Storage. C. A Python script can be written to transfer data to Azure Blob Storage using the Azure Blob Storage API. D. Azure Storage Explorer is a graphical user interface tool that can be used to transfer data to and from Azure Blob Storage.
upvoted 1 times

phylogenetics 4 years, 4 months ago
BCD are correct!
upvoted 4 times

duytran216 4 years, 11 months ago
Correct! A & E is for SQL
upvoted 4 times

Question #8 *Topic 2*

You are moving a large dataset from Azure Machine Learning Studio to a Weka environment.
You need to format the data for the Weka environment.
Which module should you use?

A. Convert to CSV

B. Convert to Dataset

C. Convert to ARFF

D. Convert to SVMLight

Suggested Answer: *C*
Use the Convert to ARFF module in Azure Machine Learning Studio, to convert datasets and results in Azure Machine Learning to the attribute-relation file format used by the Weka toolset. This format is known as ARFF. The ARFF data specification for Weka supports multiple machine learning tasks, including data preprocessing, classification, and feature selection. In this format, data is organized by entites and their attributes, and is contained in a single text file. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/convert-to-arff

Community vote distribution
C (100%)

chaudha4 Highly Voted 4 years, 9 months ago
This applies only to ML Studio (classic) - it won't get asked. Skip this one !!
upvoted 13 times

chaudha4 Highly Voted 4 years, 9 months ago
This applies only to ML Studio (classic) - it won't get asked. Skip this one !!thank you for your comment!
upvoted 13 times

Saurabhjain507 3 years, 8 months ago
thank you for your comment!
upvoted 1 times

evangelist Most Recent 1 year, 12 months ago
WEKA can use any format but ARFF is the native data format of course it is the answer.
upvoted 1 times

krishna1818 2 years, 8 months ago
Convert to ARFF
upvoted 1 times

duytran216 4 years, 11 months ago
Correct. ARFF is Weka extension.
upvoted 3 times

ipindado2020 5 years, 3 months ago
Agree with C
upvoted 1 times

Question #9 *Topic 2*

You plan to create a speech recognition deep learning model.
The model must support the latest version of Python.
You need to recommend a deep learning framework for speech recognition to include in the Data Science Virtual Machine (DSVM).
What should you recommend?

A. Rattle

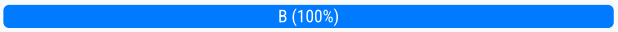
B. TensorFlow

C. Weka

D. Scikit-learn

Suggested Answer: *B*
TensorFlow is an open-source library for numerical computation and large-scale machine learning. It uses Python to provide a convenient front-end API for building applications with the framework TensorFlow can train and run deep neural networks for handwritten digit classification, image recognition, word embeddings, recurrent neural networks, sequence- to-sequence models for machine translation, natural language processing, and PDE (partial differential equation) based simulations. Incorrect Answers: A: Rattle is the R analytical tool that gets you started with data analytics and machine learning. C: Weka is used for visual data mining and machine learning software in Java. D: Scikit-learn is one of the most useful libraries for machine learning in Python. It is on NumPy, SciPy and matplotlib, this library contains a lot of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction. Reference: https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html

Community vote distribution



evangelist Most Recent 1 year, 12 months ago
without using Tensorflow what else you can use?
upvoted 1 times

farruxnb 2 years, 5 months ago
I want to buy Microsoft 62-193 Exam questions with answers. Where can i buy should be the latest updated base
upvoted 1 times

Mal42 2 years, 6 months ago
Correct
upvoted 2 times

ipindado2020 5 years, 3 months ago
Agree with B
upvoted 3 times

Question #10 *Topic 2*

You plan to use a Deep Learning Virtual Machine (DLVM) to train deep learning models using Compute Unified Device Architecture (CUDA) computations.
You need to configure the DLVM to support CUDA.
What should you implement?

A. Solid State Drives (SSD)

B. Computer Processing Unit (CPU) speed increase by using overclocking

C. Graphic Processing Unit (GPU)

D. High Random Access Memory (RAM) configuration

E. Intel Software Guard Extensions (Intel SGX) technology

Suggested Answer: *C*
A Deep Learning Virtual Machine is a pre-configured environment for deep learning using GPU instances. Reference:
https://azuremarketplace.microsoft.com/en-au/marketplace/apps/microsoft-ads.dsvm-deep-learning

Community vote distribution
C (88%) A (12%)

chaudha4 Highly Voted 4 years, 9 months ago
This question seems to be out of scope for the exam. https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE3VUjA
upvoted 7 times

Rosh4yuh Highly Voted 4 years, 7 months ago
on 17/7/2021
upvoted 6 times

evangelist Most Recent 1 year, 12 months ago
Deep Learning modeling for CUDA computation, without GPU there is nothing you can do
upvoted 1 times

Ahmed_Gehad 2 years, 6 months ago
CUDA is a parallel computing platform and application programming interface (API) developed by NVIDIA for general-purpose computing on graphics processing units (GPUs). CUDA allows developers to use the power of GPUs to accelerate the processing of data-intensive applications. In order to use CUDA, a DLVM must have a GPU that is compatible with CUDA. The DLVM must also have the CUDA software installed. The other options are not relevant to configuring a DLVM to support CUDA. SSDs can improve the performance of a DLVM, but they are not required for CUDA. Overclocking a CPU can improve its performance, but it is not necessary for CUDA. RAM is required for CUDA, but a high RAM configuration is not necessary. Intel SGX is a security technology that is not related to CUDA. Therefore, the best option is to implement a GPU in the DLVM.
upvoted 2 times

azurelearner666 3 years, 10 months ago
Selected Answer: C GPU/CUDA all the way! Also this is out of scope it seems..
upvoted 2 times

Nand4 3 years, 11 months ago
GPU is correct
upvoted 4 times

ashii007 4 years, 1 month ago
SSDs has nothing to do with CUDA, so A cannot be the right answer.
upvoted 1 times

Smith_1396 4 years, 1 month ago
https://docs.microsoft.com/en-us/azure/virtual-machines/sizes-gpuWrong! A (SSD) has nothing to do with CUDA/GPU
upvoted 1 times

Smith_1396 4 years, 1 month ago
Selected Answer: A
https://docs.microsoft.com/en-us/azure/virtual-machines/sizes-gpu
upvoted 1 times

azurelearner666 3 years, 10 months ago

Wrong! A (SSD) has nothing to do with CUDA/GPU

upvoted 4 times

Smith_1396 4 years, 1 month ago

If GPU processing thn C should correct answer. CUDA is there so A should be correct answer

upvoted 1 times

Smith_1396 4 years, 1 month ago

Answer should be A

upvoted 1 times

ipindado2020 5 years, 3 months ago

Agree with C

upvoted 2 times

Question #11 *Topic 2*

You plan to use a Data Science Virtual Machine (DSVM) with the open source deep learning frameworks Caffe2 and PyTorch.
You need to select a pre-configured DSVM to support the frameworks.
What should you create?

A. Data Science Virtual Machine for Windows 2012

B. Data Science Virtual Machine for Linux (CentOS)

C. Geo AI Data Science Virtual Machine with ArcGIS

D. Data Science Virtual Machine for Windows 2016

E. Data Science Virtual Machine for Linux (Ubuntu)

Suggested Answer: *E*
Caffe2 and PyTorch is supported by Data Science Virtual Machine for Linux. Microsoft offers Linux editions of the DSVM on Ubuntu 16.04 LTS and CentOS 7.4. Only the DSVM on Ubuntu is preconfigured for Caffe2 and PyTorch. Incorrect Answers: D: Caffe2 and PytOCH are only supported in the Data Science Virtual Machine for Linux. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview

Community vote distribution
E (100%)

James_James Highly Voted 5 years, 9 months ago
They are on the list https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/dsvm-tools-deep-learning-frameworks
upvoted 13 times

James_James 5 years, 9 months ago
They are on the list https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/dsvm-tools-deep-learning-frameworksThus, answer "E" is correct
upvoted 13 times

HkIsCrazY 5 years ago
Thus, answer "E" is correct
upvoted 7 times

evangelist Most Recent 1 year, 12 months ago
Deep Learning must be performed on Linux system and Ubuntu is natural choice.
upvoted 2 times

michaelmorar 2 years, 12 months ago
Ubuntu is legendary for everything
upvoted 4 times

phdykd 3 years ago
E. Data Science Virtual Machine for Linux (Ubuntu) is the most accurate choice. Justification: Both Caffe2 and PyTorch are widely used deep learning frameworks and are well supported on Linux platforms, particularly Ubuntu. The Data Science Virtual Machine for Linux (Ubuntu) is pre-configured with the necessary libraries, tools, and deep learning frameworks to help you get started quickly with your deep learning projects. CentOS is a stable and reliable Linux distribution, but it may not have the latest version of Caffe2 or PyTorch, or the necessary dependencies to run them. Windows Server operating systems such as Windows 2012 and Windows 2016 are not as commonly used for deep learning and may not have the necessary support for these frameworks. The Geo AI Data Science Virtual Machine with ArcGIS is a specialized DSVM specifically designed for geospatial data and may not include support for Caffe2 and PyTorch.
upvoted 1 times

MansoorDataScientist 3 years ago
The DSVM is available on: Windows Server 2019 Ubuntu 20.04 LTS
upvoted 1 times

racnaoamo 3 years, 9 months ago
similar question on 18-5-22
upvoted 1 times

TheYazan 3 years, 11 months ago

On march-9-2022

upvoted 2 times

ashii007 4 years, 1 month ago

A-D are wrong. None of the OS mentioned in options A-D are supported.

upvoted 1 times

chaudha4 4 years, 9 months ago

Has anyone seen this question asked recently. It seems like out of scope based on https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE3VUjAcaffe is outdated and is replaced by pytorch which is available on windows and linux. if the question was about horovod, then ubuntu would make sense

upvoted 3 times

chaudha4 4 years, 9 months ago

Has anyone seen this question asked recently. It seems like out of scope based on https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE3VUjA

upvoted 3 times

prashantjoge 4 years, 8 months ago

caffe is outdated and is replaced by pytorch which is available on windows and linux. if the question was about horovod, then ubuntu would make sense

upvoted 4 times

sambis 5 years, 1 month ago

Correct ans: D Ref: https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/dsvm-tools-deep-learning-frameworks

upvoted 2 times

sambis 5 years, 1 month ago

Correct ans: D Ref: https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/dsvm-tools-deep-learning-frameworksFrom the same link, I am providing you the necessary proofs: Caffe2 - Supported DSVM editions Ubuntu 16.04 PyTorch - Supported DSVM editions Windows Server 2019 Ubuntu 18.04 Ubuntu 16.04 And, DSVMs are available only in Windows Server 2019 and Linux(Ubuntu). https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview So, the answer is clearly E. Because DSVM don't work in Windows Server 2012/2016 and CentOS is not supported, as per the documentation.

upvoted 2 times

Srivathsan Highly Voted 5 years ago

From the same link, I am providing you the necessary proofs: Caffe2 - Supported DSVM editions Ubuntu 16.04 PyTorch - Supported DSVM editions Windows Server 2019 Ubuntu 18.04 Ubuntu 16.04 And, DSVMs are available only in Windows Server 2019 and Linux(Ubuntu). https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview So, the answer is clearly E. Because DSVM don't work in Windows Server 2012/2016 and CentOS is not supported, as per the documentation.

upvoted 14 times

ipindado2020 5 years, 3 months ago

Agree with D https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/tools-included

upvoted 2 times

ipindado2020 5 years, 3 months ago

Agree with D https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/tools-includedYour link specifically say that Caffe2 is not supported by Window...!!!! The answer is E

upvoted 2 times

111ssy 5 years, 2 months ago

Your link specifically say that Caffe2 is not supported by Window...!!!! The answer is E

upvoted 1 times

Kamalraj 5 years, 5 months ago

Caffe2 in Ubuntu 16.04 and PyTorch in Ubuntu 16.04 & Ubuntu 18.04.

upvoted 1 times

Rajuuu 5 years, 6 months ago

Correct answer should be Linux which is not in the option. Link below points Cafffe2 to Linux and not UBUNTU. https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/tools-included

upvoted 1 times

jkuz 5 years, 6 months ago

List of supported configuration: https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/tools-included

upvoted 3 times

Zhuo 5 years, 9 months ago

The inccorecct answer "D: Caffe2 and PytOCH are only supported in the Data Science Virtual Machine for Linux." is not in the list. Besides I thought if it is in the list, it should be the correct answer. Not E.

upvoted 1 times

Question #12 *Topic 2*

HOTSPOT -

You are performing sentiment analysis using a CSV file that includes 12,000 customer reviews written in a short sentence format. You add the CSV file to Azure

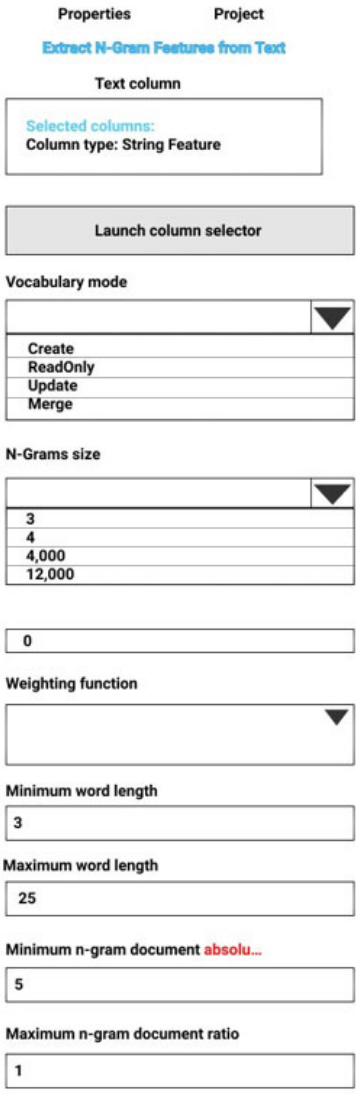
Machine Learning Studio and configure it as the starting point dataset of an experiment. You add the Extract N-Gram Features from Text module to the experiment to extract key phrases from the customer review column in the dataset.

You must create a new n-gram dictionary from the customer review text and set the maximum n-gram size to trigrams.

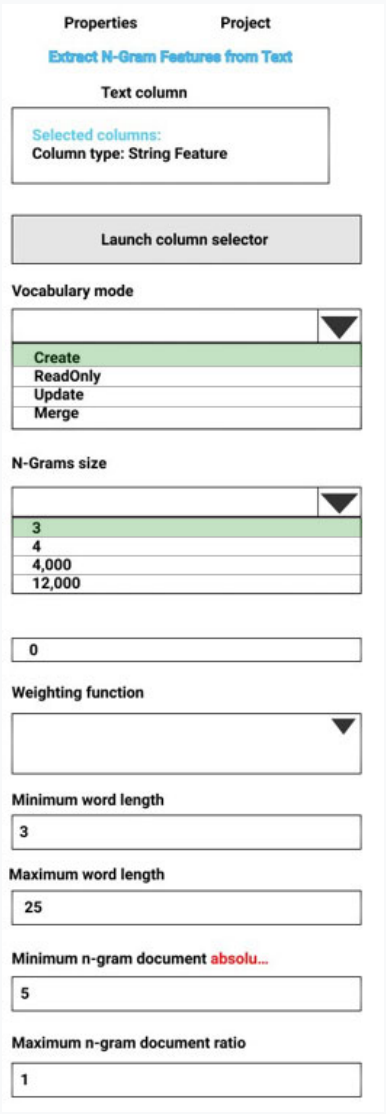
What should you select? To answer, select the appropriate options in the answer area.

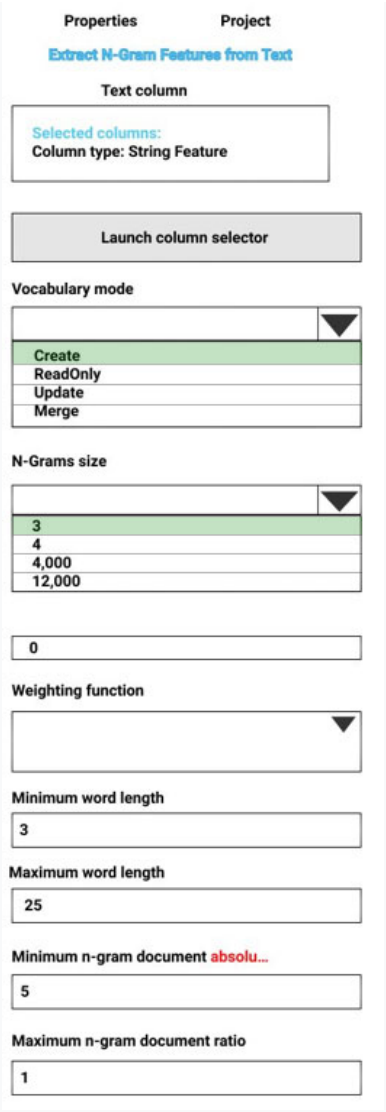
NOTE: Each correct selection is worth one point.

Hot Area:



Suggested Answer:





Vocabulary mode: Create - For Vocabulary mode, select Create to indicate that you are creating a new list of n-gram features.

N-Grams size: 3 - For N-Grams size, type a number that indicates the maximum size of the n-grams to extract and store. For example, if you type 3, unigrams, bigrams, and trigrams will be created.

Weighting function: Leave blank - The option, Weighting function, is required only if you merge or update vocabularies. It specifies how terms in the two vocabularies and their scores should be weighted against each other. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/extract-n-gram-features-from-text

ashii007 Highly Voted 4 years, 8 months ago

vocab = create ngram = 3

upvoted 5 times

Roger_Sampaio1 2 years, 8 months ago

Resposta:

1. Vocabulary mode: Criar. 2. N-Grams size: 3. Já que a questão diz trigrama. Isso é o tamanho máximo das palavras-chaves.

upvoted 1 times

Roger_Sampaio1 2 years, 8 months ago

Resposta: 1. Vocabulary mode: Criar. 2. N-Grams size: 3. Já que a questão diz trigrama. Isso é o tamanho máximo das palavras-chaves.English: 2. N-Grams size: 3. Since the question says trigram. This is the maximum keyword length.

upvoted 1 times

Sadhak Most Recent 1 year, 9 months ago

English: 2. N-Grams size: 3. Since the question says trigram. This is the maximum keyword length.

upvoted 1 times

ning 4 years, 2 months ago

Weighted function, you have to select one when creating n-gram, just tested in azure ml studio

upvoted 1 times

ning 4 years, 2 months ago

Weighted function, you have to select one when creating n-gram, just tested in azure ml studioPer documentation, The option, Weighting function, is required only if you merge or update vocabularies. So no need to select an option here

upvoted 1 times

manvel 3 years, 6 months ago

Per documentation, The option, Weighting function, is required only if you merge or update vocabularies. So no need to select an option here

upvoted 3 times

Question #13 *Topic 2*

You are developing a data science workspace that uses an Azure Machine Learning service.
You need to select a compute target to deploy the workspace.
What should you use?

A. Azure Data Lake Analytics

B. Azure Databricks

C. Azure Container Service

D. Apache Spark for HDInsight

Suggested Answer: *C*
Azure Container Instances can be used as compute target for testing or development. Use for low-scale CPU-based workloads that require less than 48 GB of RAM. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-deploy-and-where

Community vote distribution
B (56%) C (44%)

cenkersoy Highly Voted 6 years, 2 months ago
Should it not say "deploying the trained model" instead of "workspace" ?
upvoted 18 times

modschegiebsch Highly Voted 6 years, 1 month ago
Before reading this question, I never heard of "Azure Container Service". Here is why: https://azure.microsoft.com/de-de/updates/azure-container-service-will-retire-on-january-31-2020/

One can choose Azure Kubernetes Service (AKS) or Azure Container Instances as a "compute target" as well as Databricks clusters or Spark on HDInsight clusters.

https://docs.microsoft.com/de-de/azure/machine-learning/concept-compute-target

But the question is probably about deploying the model as a web service. Then there is actually no correct answer, because the Azure Container Service retired.
upvoted 14 times

modschegiebsch 6 years, 1 month ago
Before reading this question, I never heard of "Azure Container Service". Here is why: https://azure.microsoft.com/de-de/updates/azure-container-service-will-retire-on-january-31-2020/ One can choose Azure Kubernetes Service (AKS) or Azure Container Instances as a "compute target" as well as Databricks clusters or Spark on HDInsight clusters. https://docs.microsoft.com/de-de/azure/machine-learning/concept-compute-target But the question is probably about deploying the model as a web service. Then there is actually no correct answer, because the Azure Container Service retired.no, its not retired. its still used fir developmentYou are right. Answer is Azure Container Instance. C looks like it was corrected in the solution.
upvoted 14 times

prashantjoge 5 years, 2 months ago
no, its not retired. its still used fir development
upvoted 3 times

Laredo 5 years, 9 months ago
You are right. Answer is Azure Container Instance. C looks like it was corrected in the solution.
upvoted 4 times

sar77 Most Recent 1 year, 1 month ago
To deploy an Azure Machine Learning workspace, the most appropriate compute target would be: B. Azure Databricks
upvoted 1 times

f82411e 1 year, 3 months ago
porque todas las preguntas hablan de databricks nunca se menciona contenedores
upvoted 1 times

84b1989 1 year, 7 months ago

Azure Databricks is the most suitable compute target for deploying a data science workspace that uses Azure Machine Learning service. Here's why: Integration with Azure Machine Learning: Azure Databricks integrates seamlessly with Azure Machine Learning, allowing you to train and deploy machine learning models at scale. It supports collaborative data science workflows and provides a unified platform for data engineering and machine learning. Scalability: Azure Databricks is built on Apache Spark and provides a highly scalable environment for processing large datasets and running complex machine learning algorithms. Support for diverse workloads: Azure Databricks supports both batch and real-time data processing, making it ideal for data science tasks like data preparation, model training, and deployment. Collaboration: Azure Databricks provides a collaborative workspace for data scientists, data engineers, and machine learning engineers, enabling teams to work together efficiently.

upvoted 1 times

84b1989 1 year, 7 months ago

Azure Databricks is the most suitable compute target for deploying a data science workspace that uses Azure Machine Learning service. Here's why: Integration with Azure Machine Learning: Azure Databricks integrates seamlessly with Azure Machine Learning, allowing you to train and deploy machine learning models at scale. It supports collaborative data science workflows and provides a unified platform for data engineering and machine learning. Scalability: Azure Databricks is built on Apache Spark and provides a highly scalable environment for processing large datasets and running complex machine learning algorithms. Support for diverse workloads: Azure Databricks supports both batch and real-time data processing, making it ideal for data science tasks like data preparation, model training, and deployment. Collaboration: Azure Databricks provides a collaborative workspace for data scientists, data engineers, and machine learning engineers, enabling teams to work together efficientl

upvoted 1 times

maguido 1 year, 12 months ago

Azure Databricks is more suited for the development and training phases of machine learning, while Azure Kubernetes Service (AKS) is the preferred choice for deploying machine learning models as it offers the necessary infrastructure for production-level serving of these models

upvoted 1 times

Hisayuki 2 years, 9 months ago

Once we have the Databricks workspace, we can link it to the Azure ML workspace so that they can talk to each other.

upvoted 1 times

Ran2025 2 years, 10 months ago

C is correct! Azure Container Services include AKS and ACI. https://learn.microsoft.com/en-us/azure/containers/

upvoted 4 times

Ahmed_Gehad 3 years ago

I think it's C as per https://learn.microsoft.com/en-us/azure/machine-learning/concept-compute-target?view=azureml-api-2#compute-targets-for-inference

upvoted 3 times

damaldon 3 years, 1 month ago

Containers only for small deployments, otherwise Databricks100% agree with you

upvoted 2 times

damaldon 3 years, 1 month ago

Containers only for small deployments, otherwise Databricks

upvoted 2 times

prabhjot 2 years, 6 months ago

100% agree with you

upvoted 1 times

rishi_ram 3 years, 2 months ago

While Azure Container Service (ACS) can be used to deploy and manage containers for various applications, including machine learning workloads, it is not the recommended compute target for deploying a data science workspace using Azure Machine Learning service. Azure Databricks is a more suitable choice for deploying the workspace in this scenario. It provides a collaborative and scalable environment specifically designed for data engineering and data science tasks. With Azure Databricks, you can leverage the power of Apache Spark for distributed data processing and machine learning tasks, making it an ideal compute target for data science workloads. Azure Container Service, on the other hand, is a container orchestration service that supports various container runtimes, such as Docker and Kubernetes. While it can be used to deploy machine learning models encapsulated in containers, it is not optimized for the specific requirements of a data science workspace and may not provide the same level of integration and ease of use as Azure Databricks. Therefore, in this context, the recommended choice for deploying the data science workspace using Azure Machine Learning service is B. Azure Databricks.

upvoted 1 times

sahithi2004 3 years, 5 months ago

it is databricks

upvoted 3 times

phdykd 3 years, 6 months ago

B. Azure Databricks is the most accurate answer if you're looking to deploy a data science workspace using Azure Machine Learning service as a compute target. Azure Databricks is a fast, easy, and collaborative Apache Spark-based analytics platform that helps you build and deploy machine learning models. It provides a seamless integration with the Azure Machine Learning service and offers features such as auto-scaling, collaboration and management of the machine learning workflow, and a secure and scalable infrastructure for running your data science projects. While you can run machine learning models inside containers in Azure Container Service, it doesn't provide the seamless integration and specific features for managing and deploying machine learning models that Azure Databricks does. For example, you wouldn't have access to the auto-scaling, collaboration, and management features for the machine learning workflow that Azure Databricks provides.

upvoted 4 times

ning 4 years, 3 months ago

Deployment target, only two options: Container Instance, Azure Kubernetes Cluster

upvoted 4 times

Rosh4yuh 5 years, 1 month ago

on 17/7/2021

upvoted 5 times

Narendra05 5 years, 1 month ago

The service is discontinued from 31Jan 2020 the Question is not relavent anymore

upvoted 3 times

Question #14 *Topic 2*

You are solving a classification task.
The dataset is imbalanced.
You need to select an Azure Machine Learning Studio module to improve the classification accuracy.
Which module should you use?

A. Permutation Feature Importance

B. Filter Based Feature Selection

C. Fisher Linear Discriminant Analysis

D. Synthetic Minority Oversampling Technique (SMOTE)

Suggested Answer: *D*
Use the SMOTE module in Azure Machine Learning Studio (classic) to increase the number of underrepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases. You connect the SMOTE module to a dataset that is imbalanced. There are many reasons why a dataset might be imbalanced: the category you are targeting might be very rare in the population, or the data might simply be difficult to collect. Typically, you use SMOTE when the class you want to analyze is under- represented. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote

Community vote distribution
D (100%)

crushgear Highly Voted 5 years, 4 months ago
But these things are not covered in Microsoft Learn.it did mention oversamplinghttps://docs.microsoft.com/en-us/azure/machine-learning/component-reference/smote The link is from Azure ML documentation. So the Microsoft Learn does cover this concepthaha they will not metioned everything.
upvoted 11 times

crushgear Highly Voted 5 years, 4 months ago
But these things are not covered in Microsoft Learn.
upvoted 11 times

ckkobe24 5 years ago
it did mention oversampling
upvoted 2 times

DingDongSingSong 4 years, 4 months ago
https://docs.microsoft.com/en-us/azure/machine-learning/component-reference/smote The link is from Azure ML documentation. So the Microsoft Learn does cover this concept
upvoted 2 times

dadofworld 1 year, 10 months ago
haha they will not metioned everything.
upvoted 1 times

sar77 Most Recent 1 year, 1 month ago
The best choice for handling an imbalanced dataset in a classification task is: D. Synthetic Minority Oversampling Technique (SMOTE)
upvoted 1 times

David_Tadeu 4 years, 4 months ago
Related with https://www.examtopics.com/discussions/microsoft/view/55875-exam-dp-100-topic-2-question-23-discussion/
upvoted 2 times

nbboy 4 years, 7 months ago
this's right
upvoted 2 times

pg13 5 years, 2 months ago
D is correct
upvoted 1 times

Photons 5 years, 6 months ago

D is the right answer

upvoted 2 times

Vipuls 5 years, 7 months ago

Yes Answer id D

upvoted 1 times

ipindado2020 5 years, 9 months ago

Agree with D

upvoted 2 times

Question #15 *Topic 2*

DRAG DROP -

You configure a Deep Learning Virtual Machine for Windows.

You need to recommend tools and frameworks to perform the following:

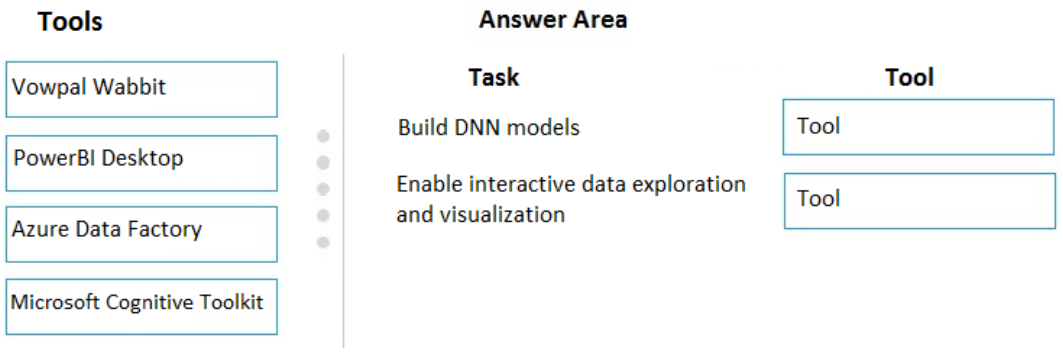
➪ Build deep neural network (DNN) models

➪ Perform interactive data exploration and visualization

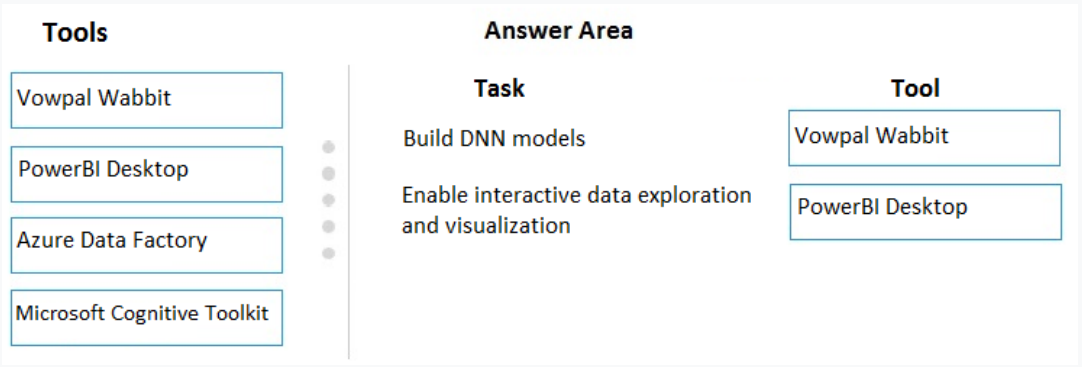
Which tools and frameworks should you recommend? To answer, drag the appropriate tools to the correct tasks. Each tool may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

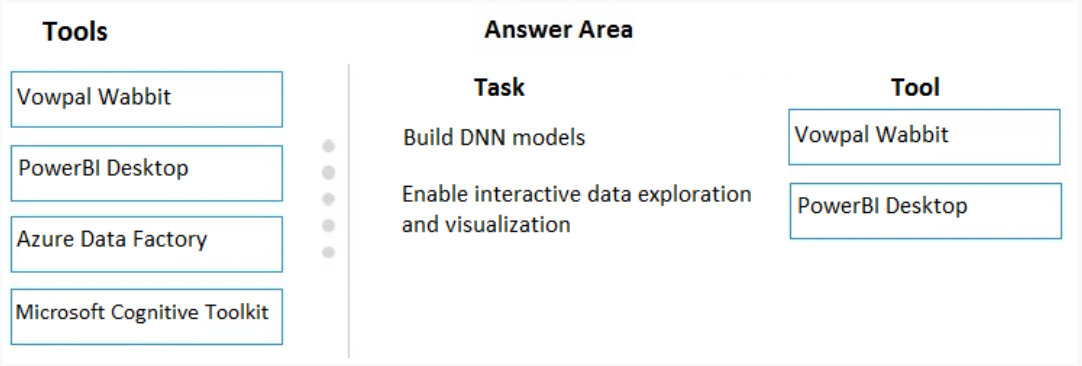
NOTE: Each correct selection is worth one point.

Select and Place:



Suggested Answer:





Box 1: Vowpal Wabbit - Use the Train Vowpal Wabbit Version 8 module in Azure Machine Learning Studio (classic), to create a machine learning model by using Vowpal Wabbit.

Box 2: PowerBI Desktop - Power BI Desktop is a powerful visual data exploration and interactive reporting tool BI is a name given to a modern approach to business decision making in which users are empowered to find, explore, and share insights from data across the enterprise.

Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/train-vowpal-wabbit-version-8-model
https://docs.microsoft.com/en-us/azure/architecture/data-guide/scenarios/interactive-data-exploration

jameswoo Highly Voted 5 years, 9 months ago

I think it should be Microsoft Cognitive Kit for DNN.

upvoted 35 times

sakshi239 Highly Voted 5 years, 9 months ago

It should be Microsoft Cognitive Kit. Vowpal Wabbit is for Machine Learning.

upvoted 10 times

NullVoider_0 Most Recent 2 years, 2 months ago

For building DNN models we should use Microsoft Cognitive tool kit, not Vowpal Wabbit. Vowpal Wabbit: While it's a fast machine learning system, it's more focused on online learning scenarios and may not be as comprehensive as CNTK for building complex DNN models.

upvoted 2 times

phdykd 3 years ago

For building deep neural network (DNN) models: Microsoft Cognitive Toolkit (CNTK) For interactive data exploration and visualization: PowerBI DesktopYes, ChatGPT agrees.

upvoted 5 times

phdykd 3 years ago

For building deep neural network (DNN) models:

Microsoft Cognitive Toolkit (CNTK) For interactive data exploration and visualization:

PowerBI Desktop

upvoted 5 times

phydev 2 years, 7 months ago

Yes, ChatGPT agrees.

upvoted 1 times

JTWang 3 years, 10 months ago

CNTK allows the user to easily realize and combine popular model types such as feed-forward DNNs, convolutional neural networks (CNNs) and recurrent neural networks (RNNs/LSTMs). CNTK implements stochastic gradient descent (SGD, error backpropagation) learning with automatic differentiation and parallelization across multiple GPUs and servers https://docs.microsoft.com/en-us/cognitive-toolkit/

upvoted 3 times

ashii007 4 years, 1 month ago

this question is based on outdated topic. Cognitive toolkit is being retired and DSVMs are recommended way to go.

upvoted 3 times

RyanTsai 4 years, 5 months ago

agree: Microsoft Conginitive Kitabe chup saale

upvoted 3 times

RyanTsai 4 years, 5 months ago

agree: Microsoft Conginitive Kit

upvoted 3 times

viplov 4 years, 1 month ago

abe chup saale

upvoted 3 times

Akki0120 4 years, 7 months ago

If anyone wants all questions ping me 9403778084

upvoted 3 times

azurecert2021 4 years, 7 months ago

yes Microsoft Cognitive TooKit for DNN.

upvoted 2 times

sayak17 5 years, 2 months ago

I will go with microsoft cognitive toolkit for DNN https://docs.microsoft.com/en-us/cognitive-toolkit/examples#c-examples

upvoted 6 times

ARC 5 years, 5 months ago

Microsoft Conginitive Kit fits better for DNN

upvoted 2 times

amigi 5 years, 5 months ago

Vowpal Wabbit can also be used for Neural Networks: https://stackoverflow.com/questions/33441183/vowpal-wabbit-input-of-neural-network But it seems that only for nn with 1 hidden layer

upvoted 1 times

jkuz 5 years, 6 months ago

https://docs.microsoft.com/en-us/cognitive-toolkit/ "CNTK allows the user to easily realize and combine popular model types such as feed-forward DNNs, "

upvoted 3 times

jkuz 5 years, 6 months ago

https://docs.microsoft.com/en-us/cognitive-toolkit/ "CNTK allows the user to easily realize and combine popular model types such as feed-forward DNNs, "I just also that.

upvoted 3 times

Parzival 5 years, 5 months ago

I just also that.

upvoted 1 times

kukushka 5 years, 7 months ago

The link is wrong here is correct one - https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/train-vowpal-wabbit-version-8-model

upvoted 3 times

Question #16 *Topic 2*

You use Azure Machine Learning Studio to build a machine learning experiment.
You need to divide data into two distinct datasets.
Which module should you use?

A. Assign Data to Clusters

B. Load Trained Model

C. Partition and Sample

D. Tune Model-Hyperparameters

Suggested Answer: *C*
Partition and Sample with the Stratified split option outputs multiple datasets, partitioned using the rules you specified. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample

Community vote distribution
C (100%)

Nugi Highly Voted 3 years, 6 months ago
Either Partition and Sample, or Split Data.
upvoted 18 times

chaudha4 Highly Voted 2 years, 9 months ago
I doubt we will see this question going forward since it applies only to Studio (classic). Similar drag and drop modules have been added to Azure Machine Learning designer which is where we should expect new questions.
upvoted 10 times

ougullamaija Most Recent 1 year, 10 months ago
Correctamunda.
upvoted 1 times

EricWu 4 years ago
so the answer is C
upvoted 4 times

Question #17 *Topic 2*

DRAG DROP -

You are creating an experiment by using Azure Machine Learning Studio.

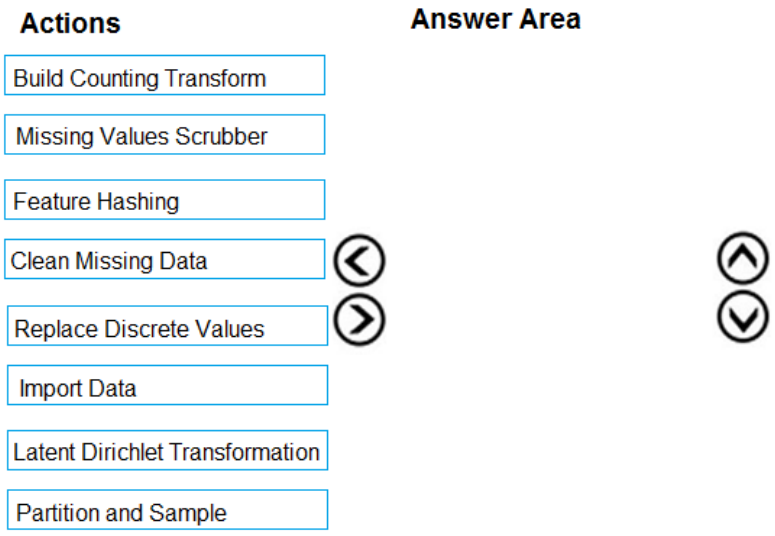
You must divide the data into four subsets for evaluation. There is a high degree of missing values in the data. You must prepare the data for analysis.

You need to select appropriate methods for producing the experiment.

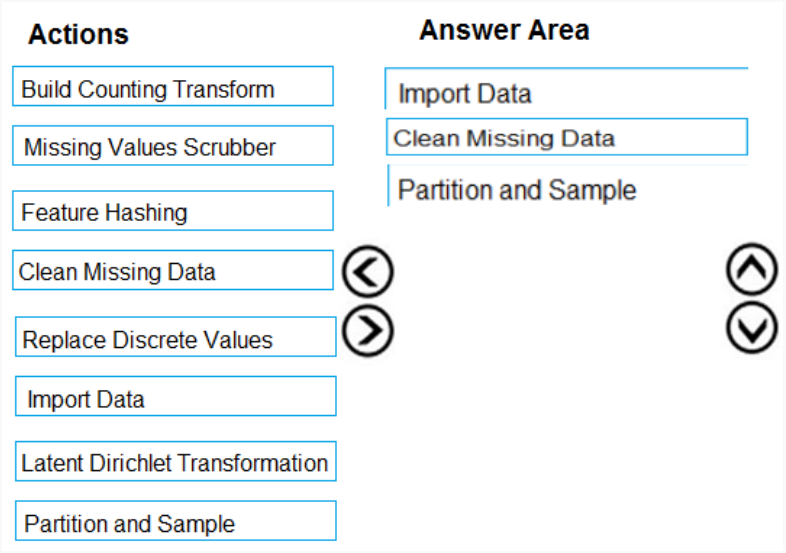
Which three modules should you run in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.

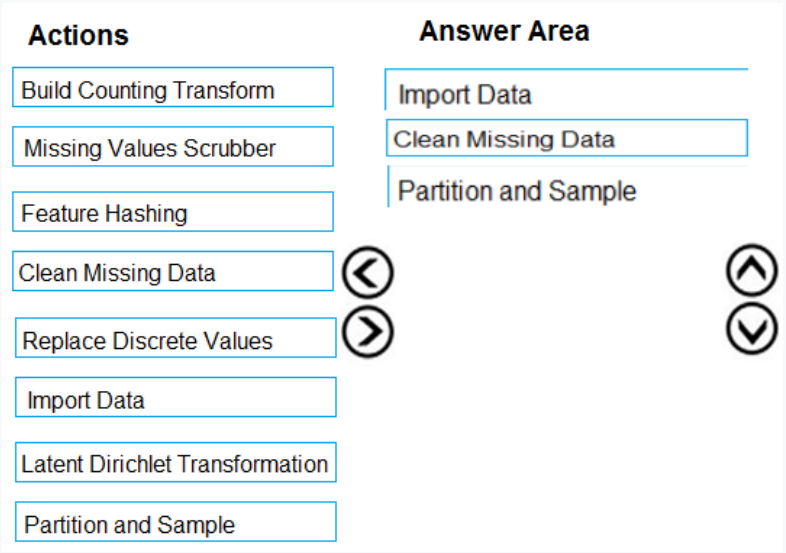
Select and Place:



Suggested Answer:



.



The Clean Missing Data module in Azure Machine Learning Studio, to remove, replace, or infer missing values. Incorrect Answers: ✑ Latent Direchlet Transformation: Latent Dirichlet Allocation module in Azure Machine Learning Studio, to group otherwise unclassified text into a number of categories. Latent Dirichlet Allocation (LDA) is often used in natural language processing (NLP) to find texts that are similar. Another common term is topic modeling. ✑ Build Counting Transform: Build Counting Transform module in Azure Machine Learning Studio, to analyze training data. From this data, the module builds a count table as well as a set of count-based features that can be used in a predictive model. Missing Value Scrubber: The Missing Values Scrubber module is deprecated.

✑ Feature hashing: Feature hashing is used for linguistics, and works by converting unique tokens into integers. ✑ Replace discrete values: the Replace Discrete Values module in Azure Machine Learning Studio is used to generate a probability score that can be used to represent a discrete value. This score can be useful for understanding the information value of the discrete values. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

ougullamaija Highly Voted 3 years, 10 months ago

Correct as f*uck.

upvoted 19 times

NullVoider_0 Most Recent 2 years ago

On exam 12-02-2024.

upvoted 1 times

ZIMARAKI 2 years, 11 months ago

correct

upvoted 2 times

azurelearner666 3 years, 10 months ago

correct!

upvoted 2 times

hargur 4 years, 4 months ago

on 19Oct2021

upvoted 4 times

Akki0120 4 years, 7 months ago

If anyone wants all questions ping me 9403778084

upvoted 1 times

HOTSPOT -

You are retrieving data from a large datastore by using Azure Machine Learning Studio.

You must create a subset of the data for testing purposes using a random sampling seed based on the system clock.

You add the Partition and Sample module to your experiment.

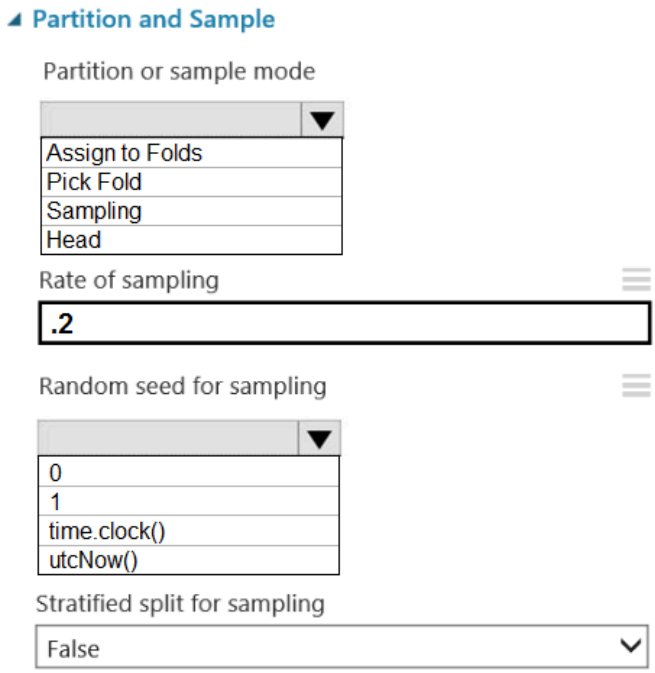
You need to select the properties for the module.

Which values should you select? To answer, select the appropriate options in the answer area.

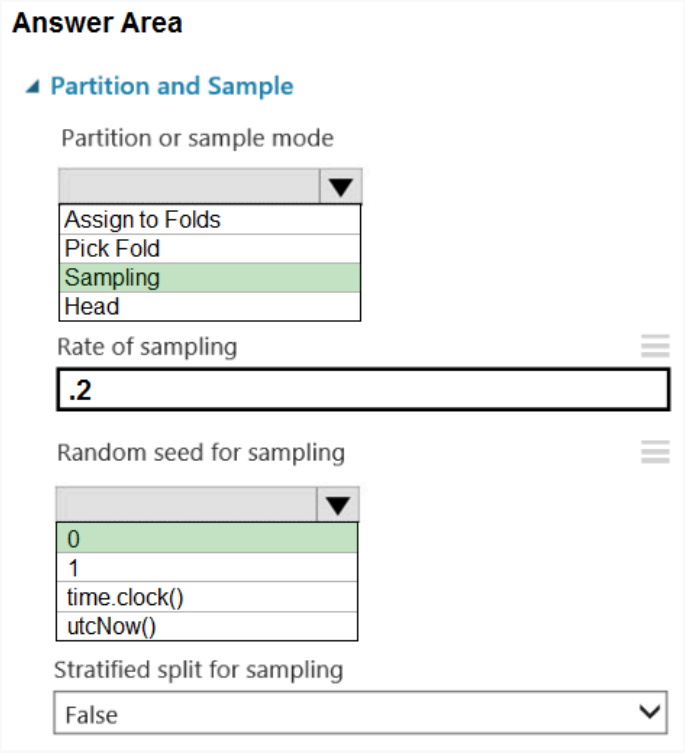
NOTE: Each correct selection is worth one point.

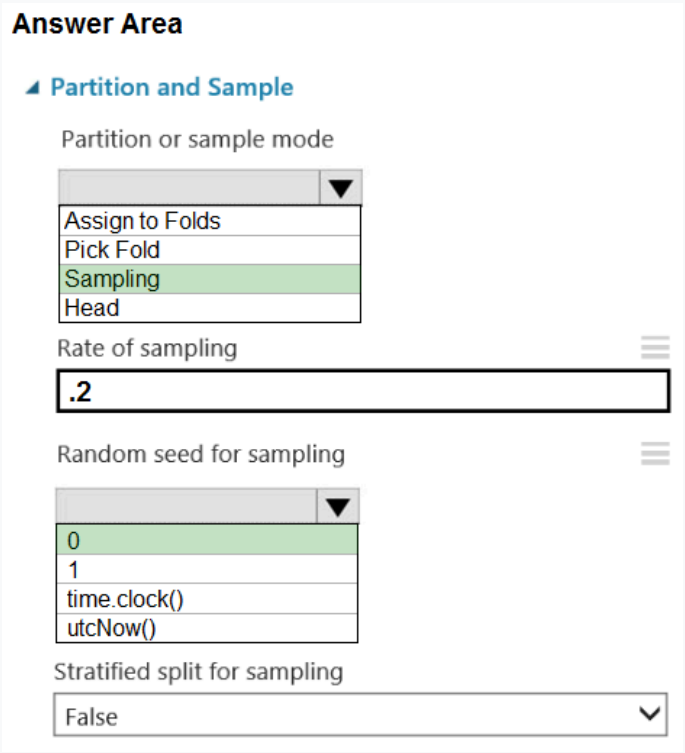
Hot Area:

Answer Area



Suggested Answer:





Box 1: Sampling -

Create a sample of data - This option supports simple random sampling or stratified random sampling. This is useful if you want to create a smaller representative sample dataset for testing. 1. Add the Partition and Sample module to your experiment in Studio, and connect the dataset. 2. Partition or sample mode: Set this to Sampling. 3. Rate of sampling. See box 2 below.

Box 2: 0 - 3. Rate of sampling. Random seed for sampling: Optionally, type an integer to use as a seed value. This option is important if you want the rows to be divided the same way every time. The default value is 0, meaning that a starting seed is generated based on the system clock. This can lead to slightly different results each time you run the experiment. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample

BilJon Highly Voted 4 years, 4 months ago

Random seed for sampling: Optionally, type an integer to use as a seed value. This option is important if you want the rows to be divided the same way every time. The default value is 0, meaning that a starting seed is generated based on the system clock. This can lead to slightly different results each time you run the experiment. https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample

upvoted 22 times

David_Tadeu 3 years, 4 months ago

So they just had the option time.clock() to trick us into failing?

upvoted 10 times

Hisayuki Most Recent 1 year, 9 months ago

time.clock() - Trick us. Microsoft learning says "This option is important if you want the rows to be divided the same way every time. The default value is 0, meaning that a starting seed is generated based on the system clock."

upvoted 1 times

ning 3 years, 3 months ago

correct! 0 --> system clock, random sampling not stratified sampling

upvoted 2 times

dzzz 4 years, 8 months ago

The question says "a random sampling seed based on the system clock". I believe that implies the 2nd drop down is time.clock() - machine time.there's no such option as time.clock(). 0 is the correct answer.I think, "0" is a correct answer. Because the value to enter should be an integer. And look here as well: Random seed for sampling: Optionally, enter an integer to use as a seed value. This option is important if you want the rows to be divided the same way every time. The default value is 0, meaning that a starting seed is generated based on the system clock. This value can lead to slightly different results each time you run the pipeline. https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/partition-and-sample

upvoted 5 times

dzzz 4 years, 8 months ago

The question says "a random sampling seed based on the system clock". I believe that implies the 2nd drop down is time.clock() - machine time.

upvoted 5 times

synapse 3 years, 5 months ago

there's no such option as time.clock(). 0 is the correct answer.

upvoted 6 times

damirbek369 Highly Voted 4 years, 8 months ago

I think, "0" is a correct answer. Because the value to enter should be an integer. And look here as well: Random seed for sampling: Optionally, enter an integer to use as a seed value. This option is important if you want the rows to be divided the same way every time. The default value is 0, meaning that a starting seed is generated based on the system clock. This value can lead to slightly different results each time you run the pipeline. https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/partition-and-sample

upvoted 17 times

Question #19 *Topic 2*

You are creating a machine learning model. You have a dataset that contains null rows.
You need to use the Clean Missing Data module in Azure Machine Learning Studio to identify and resolve the null and missing data in the dataset.
Which parameter should you use?

A. Replace with mean

B. Remove entire column

C. Remove entire row

D. Hot Deck

E. Custom substitution value

F. Replace with mode

Suggested Answer: *C*
Remove entire row: Completely removes any row in the dataset that has one or more missing values. This is useful if the missing value can be considered randomly missing. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

Community vote distribution
C (100%)

ljljljlj Highly Voted 4 years, 7 months ago
On exam 2021/7/10
upvoted 8 times

sar77 Most Recent 1 year, 1 month ago
For a dataset with null rows, the most fitting parameter in Azure Machine Learning Studio's Clean Missing Data module would be: C. Remove entire row
upvoted 1 times

Ejire 1 year, 8 months ago
The correct answer is C. If you need voucher for any DP or AI exam contact me on +2348139103938
upvoted 1 times

krishna1818 2 years, 8 months ago
Remove the entire row if there are less number of missing values
upvoted 1 times

MarinaMijailovic 2 years, 9 months ago
Null rows means the entire the row is null so C
upvoted 2 times

David_Tadeu 3 years, 11 months ago
Remove entire row: removes a row which contains a null value for a specified (it is a parameter of this procedure) set of columns
upvoted 1 times

synapse 3 years, 11 months ago
"Null Rows" So I am going with C
upvoted 2 times

guddusao 4 years, 6 months ago
I think mode will work both numerical and categorical data. May be mode can impute the missing value.
upvoted 1 times

Abdulraoufhakeem 4 years, 7 months ago
I agree to answer C
upvoted 2 times

Haet 4 years, 9 months ago
Answer C is correct
upvoted 3 times

dev2dev 4 years, 11 months ago

A, B, C and F are valid.They are not since we are only targeting NULL rows. C is correct

upvoted 3 times

dev2dev 4 years, 11 months ago

A, B, C and F are valid.

upvoted 3 times

Dasist Highly Voted 4 years, 10 months ago

They are not since we are only targeting NULL rows. C is correct

upvoted 6 times

Abhinav_nasaiitkgp 5 years, 1 month ago

How other options are not valid? We can impute missing values with mean median and mode also. Something is missing in the question

upvoted 3 times

Abhinav_nasaiitkgp 5 years, 1 month ago

How other options are not valid? We can impute missing values with mean median and mode also. Something is missing in the questionyou can replace with mean or mode only if the column is of integer type since the values are nulls here it implies the data type is String or text. and hence from the options give only remove rows looks like the correct answer.hello ck1729 Does null means that the data type is only String or text ? Does this implies that missing values for integers and float are Nan and not Null? Otherwise how can we know the datatype of the missing values pleas"You have a dataset that contains null rows" : If all the columns are Null for the row you don't need to keep itIf you have ENTIRE rows that are null, chances are you don't need them.

upvoted 3 times

ck1729 5 years, 1 month ago

you can replace with mean or mode only if the column is of integer type since the values are nulls here it implies the data type is String or text. and hence from the options give only remove rows looks like the correct answer.hello ck1729 Does null means that the data type is only String or text ? Does this implies that missing values for integers and float are Nan and not Null? Otherwise how can we know the datatype of the missing values pleas

upvoted 2 times

ck1729 5 years, 1 month ago

you can replace with mean or mode only if the column is of integer type since the values are nulls here it implies the data type is String or text. and hence from the options give only remove rows looks like the correct answer.

upvoted 2 times

ML_Novice 4 years, 3 months ago

hello ck1729 Does null means that the data type is only String or text ? Does this implies that missing values for integers and float are Nan and not Null? Otherwise how can we know the datatype of the missing values pleas

upvoted 1 times

Adrien_B 5 years ago

"You have a dataset that contains null rows" : If all the columns are Null for the row you don't need to keep it

upvoted 6 times

allanm 4 years, 9 months ago

If you have ENTIRE rows that are null, chances are you don't need them.

upvoted 6 times

HOTSPOT -

The finance team asks you to train a model using data in an Azure Storage blob container named finance-data.

You need to register the container as a datastore in an Azure Machine Learning workspace and ensure that an error will be raised if the container does not exist.

How should you complete the code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

```
datastore = Datastore.[ register_azure_blob_container | register_azure_file_share | register_azure_data_lake | register_azure_sql_database ] (workspace = ws,
    datastore_name = 'finance_datastore',
    container_name = 'finance-data',
    account_name = 'fintrainingdatastorage',
    account_key = 'FWUYORRv3XoyNe...',
    [ create_if_not_exists = True | create_if_not_exists = False | overwrite = True | overwrite = False ])
```

Suggested Answer:

Answer Area

```
datastore = Datastore. [register_azure_blob_container | register_azure_file_share | register_azure_data_lake | register_azure_sql_database] (workspace = ws,
    datastore_name = 'finance_datastore',
    container_name = 'finance-data',
    account_name = 'fintrainingdatastorage',
    account_key = 'FWUYORRv3XoyNe...',
    [create_if_not_exists = True | create_if_not_exists = False | overwrite = True | overwrite = False]
```

Answer Area

```
datastore = Datastore. [register_azure_blob_container | register_azure_file_share | register_azure_data_lake | register_azure_sql_database] (workspace = ws,
    datastore_name = 'finance_datastore',
    container_name = 'finance-data',
    account_name = 'fintrainingdatastorage',
    account_key = 'FWUYORRv3XoyNe...',
    [create_if_not_exists = True | create_if_not_exists = False | overwrite = True | overwrite = False]
```

Box 1: register_azure_blob_container Register an Azure Blob Container to the datastore. Box 2: create_if_not_exists = False Create the file share if it does not exist, defaults to False. Reference: https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.datastore.datastore

MattAnya Highly Voted 2 years, 1 month ago

was on exam 01/03/2023

upvoted 13 times

atyagi55 3 years, 9 months ago

give answer is correct.Please refer link https://azure.github.io/azureml-sdk-for-r/reference/register_azure_blob_container_datastore.html create_if_not_exists "If TRUE, creates the blob container if it does not exists."

upvoted 9 times

MarinaMijailovic Most Recent 1 year, 9 months ago

Answer is correct. We are not creating a container, only registering it and we need an error message if it does not exist. If we set "create_if_not_exists" to true, it will not display the error message but create the container and we dont want that.

upvoted 3 times

ougullamaija 2 years, 10 months ago

You can resolve this by thinking logically. overwrite = False would mean, that we don't overwrite the existing file, but that we create a copy! create_if_not_exists = False won't create anything if there is an existing folder.

upvoted 3 times

azurelearner666 2 years, 10 months ago

Correct, A: Register_azure_blob_container B: Create_if_not_exists = False First is easy, we are talking about storage accounts and a blob container. Second too, It should raise an error if the container does not exist, so it should not create it in that case, as there would be no error then.

upvoted 4 times

kkkk_jjjj 2 years, 11 months ago

on exam 18/03/2022

upvoted 2 times

hargur 3 years, 4 months ago

on 19Oct2021

upvoted 2 times

kisskeo 3 years, 4 months ago

On Exam 01 Oct 2021

upvoted 3 times

mthombenindhl84 3 years, 5 months ago

on exam 11/9/2021

upvoted 2 times

snsnsnsn 3 years, 5 months ago

on exam 2/9/21

upvoted 1 times

dushmantha 3 years, 5 months ago

On exam 2021/08/31

upvoted 1 times

datamijn 3 years, 6 months ago

on 2/8/2021

upvoted 3 times

Rosh4yuh 3 years, 7 months ago

on 17/7/2021

upvoted 2 times

ljljljlj 3 years, 7 months ago

On exam 2021/7/10

upvoted 3 times

prashantjoge 3 years, 8 months ago

the answer does not satisfy the question "ensure that an error will be raised if the container does not exist."

upvoted 2 times

ACSC 3 years, 10 months ago

Answer is correct. There is a mistake in the documentation. Where it says 'file share', it is 'container'.

upvoted 4 times

dev2dev 3 years, 11 months ago

2nd option should be overwrite = False create_if_not_exists is for the file share.

upvoted 3 times

dev2dev 3 years, 11 months ago

2nd option should be overwrite = False create_if_not_exists is for the file share.the answer is correct don't comment if you not sure !Agreed. Incredible how many wrong answers can be found here.Answer is correct, if you are not sure don't post anything, you are adding to the confusion. Have a look at the documentation instead: https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.datastore(class)?view=azure-ml-py#register-azure-blob-container-workspace--datastore-name--container-name--account-name--sas-token-none--account-key-none--protocol-none--endpoint-none--overwrite-false--create-if-not-exists-false--skip-validation-false--blob-cache-timeout-none--grant-workspace-access-false--subscription-id-none--resource-group-none-

upvoted 3 times

jamessnow 3 years, 10 months ago

the answer is correct don't comment if you not sure !

upvoted 12 times

Anty85 3 years, 11 months ago

Agreed. Incredible how many wrong answers can be found here.

upvoted 4 times

l2azure Highly Voted 3 years, 10 months ago

Answer is correct, if you are not sure don't post anything, you are adding to the confusion. Have a look at the documentation instead: https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.datastore(class)?view=azure-ml-py#register-azure-blob-container-workspace--datastore-name--container-name--account-name--sas-token-none--account-key-none--protocol-none--endpoint-none--overwrite-false--create-if-not-exists-false--skip-validation-false--blob-cache-timeout-none--grant-workspace-access-false--subscription-id-none--resource-group-none-

upvoted 13 times

Question #21 *Topic 2*

You plan to provision an Azure Machine Learning Basic edition workspace for a data science project.
You need to identify the tasks you will be able to perform in the workspace.
Which three tasks will you be able to perform? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. Create a Compute Instance and use it to run code in Jupyter notebooks.

B. Create an Azure Kubernetes Service (AKS) inference cluster.

C. Use the designer to train a model by dragging and dropping pre-defined modules.

D. Create a tabular dataset that supports versioning.

E. Use the Automated Machine Learning user interface to train a model.

Suggested Answer: *A, B, D*
Incorrect Answers: C, E: The UI is included the Enterprise edition only. Reference: https://azure.microsoft.com/en-us/pricing/details/machine-learning/

Community vote distribution



brendal89 Highly Voted 4 years, 10 months ago
I can confirm that I had this question on the exam in april 2021Thanks. That makes me a bit nervous. I was ignoring outdated questions like this. There are many more question on Studio(classic) that I have ignore hoping that similar question will get asked for designer instead.
upvoted 18 times

brendal89 Highly Voted 4 years, 10 months ago
I can confirm that I had this question on the exam in april 2021
upvoted 18 times

chaudha4 4 years, 9 months ago
Thanks. That makes me a bit nervous. I was ignoring outdated questions like this. There are many more question on Studio(classic) that I have ignore hoping that similar question will get asked for designer instead.
upvoted 4 times

NullVoider_0 2 years, 2 months ago
As of December 2023, you can use A) Create a Compute Instance and use it to run code in Jupyter notebooks. C) Use the designer to train a model by dragging and dropping pre-defined modules. E) Use the Automated Machine Learning user interface to train a model. Currently, Create AKS deployment clusters, and Create versioned tabular datasets are not supported in Azure ML studio. If you want to access these services then you will have to go to AKS (Azure Kubernetes Service) for AKS-related stuff.
upvoted 13 times

sar77 Most Recent 1 year ago
Correct Answers: A. Create a Compute Instance and use it to run code in Jupyter notebooks. B. Create an Azure Kubernetes Service (AKS) inference cluster. D. Create a tabular dataset that supports versioning.
upvoted 2 times

f82411e 1 year, 3 months ago
ADE es mas facil automl que el diseñador
upvoted 1 times

Sam2213 1 year, 5 months ago
Can Anyone please confirm correct ans based on latest update ?
upvoted 1 times

Edriv 3 years, 1 month ago
I am AgreeHi, Agree!
upvoted 2 times

Edriv 3 years, 1 month ago
I am Agree
upvoted 2 times

phydev 2 years, 7 months ago

Hi, Agree!

upvoted 9 times

ning 3 years, 9 months ago

No idea, just set up an Azure Machine Learning Studio, all are available, also, I have never seem anything version selection thereYes, this comment is correct. There is also data versioning available when you create datasets.

upvoted 5 times

ning 3 years, 9 months ago

No idea, just set up an Azure Machine Learning Studio, all are available, also, I have never seem anything version selection there

upvoted 5 times

Saurabhjain507 3 years, 8 months ago

Yes, this comment is correct. There is also data versioning available when you create datasets.

upvoted 1 times

azayra 4 years, 3 months ago

all are correct because now its just basic

upvoted 1 times

ljljljlj 4 years, 7 months ago

On exam 2021/7/10

upvoted 5 times

chaudha4 4 years, 10 months ago

This seems like an outdated question (as of April 2021). There is no mention of any basic or enterprise options. There is only one and based on what I see, you can do everything listed in the question.

upvoted 10 times

ACSC 4 years, 10 months ago

Beginning December 21st, all Enterprise Edition workspaces will be automatically set to Basic Edition, which has the same capabilities. No downtime will occur during this process. On January 1, 2021, Enterprise Edition will be formally retired. https://docs.microsoft.com/en-us/azure/machine-learning/concept-workspace#what-happened-to-enterprise-edition

upvoted 8 times

vippi66 5 years ago

I don't think they have a premium offering now.

upvoted 2 times

Question #22 *Topic 2*

HOTSPOT -

A coworker registers a datastore in a Machine Learning services workspace by using the following code:

You need to write code to access the datastore from a notebook.

How should you complete the code segment? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

```
Datastore.register_azure_blob_container(workspace=ws,
 datastore_name='demo_datastore',
 container_name='demo_datacontainer',
 account_name='demo_account',
 account_key='0A0A0A-0A0A00A-0A00A0A0A0A0A',
 create_if_not_exists=True)
```

Answer Area

```
import azureml.core
from azureml.core import Workspace, Datastore
ws = Workspace.from_config()
datastore = [ ▼ ].get( [ ▼ ] , ' [ ▼ ] ')
```

Workspace	ws	demo_datastore
Datastore	run	demo_datacontainer
Experiment	experiment	demo_account
Run	log	Datastore

Suggested Answer:

Answer Area

```
import azureml.core
from azureml.core import Workspace, Datastore
ws = Workspace.from_config()
datastore = [Workspace / Datastore / Experiment / Run].get([ws / run / experiment / log], ', ['demo_datastore' / 'demo_datacontainer' / 'demo_account' / 'Datastore'])
```

Answer Area

```
import azureml.core
from azureml.core import Workspace, Datastore
ws = Workspace.from_config()
datastore = [Workspace / Datastore / Experiment / Run].get([ws / run / experiment / log], ', ['demo_datastore' / 'demo_datacontainer' / 'demo_account' / 'Datastore'])
```

Box 1: DataStore - To get a specific datastore registered in the current workspace, use the get() static method on the Datastore class: # Get a named datastore from the current workspace datastore = Datastore.get(ws, datastore_name='your datastore name')

Box 2: ws -

Box 3: demo_datastore - Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-access-data

Lucario95 Highly Voted 4 years, 9 months ago
Seems correct
upvoted 14 times

MattAnya Highly Voted 3 years, 1 month ago
On 03 Jan2023
upvoted 10 times

thisiston Most Recent 1 year, 9 months ago
Correct answer is: 'Datastore', 'ws', 'blob_datastore_name' Ex:blob_datastore = Datastore.get(ws, blob_datastore_name) Link: https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.datastore.datastore?view=azure-ml-py
upvoted 1 times

k1ngs1zed 3 years, 10 months ago
answer is correct
upvoted 1 times

kisskeo 4 years, 4 months ago
On Exam 01 Oct 2021
upvoted 2 times

snsnsnsn 4 years, 5 months ago
on exam 2/9/21
upvoted 1 times

datamijn 4 years, 6 months ago
on 2/8/2021
upvoted 2 times

Rosh4yuh 4 years, 7 months ago
on 17/7/2021 Answer is correct
upvoted 4 times

ljljljlj 4 years, 7 months ago

On exam 2021/7/10

upvoted 2 times

okeyken1 4 years, 8 months ago

yes correct

upvoted 4 times

Question #23 *Topic 2*

A set of CSV files contains sales records. All the CSV files have the same data schema.
Each CSV file contains the sales record for a particular month and has the filename sales.csv. Each file is stored in a folder that indicates the month and year when the data was recorded. The folders are in an Azure blob container for which a datastore has been defined in an Azure Machine Learning workspace. The folders are organized in a parent folder named sales to create the following hierarchical structure:
At the end of each month, a new folder with that month's sales file is added to the sales folder.
You plan to use the sales data to train a machine learning model based on the following requirements:
✑ You must define a dataset that loads all of the sales data to date into a structure that can be easily converted to a dataframe.
✑ You must be able to create experiments that use only data that was created before a specific previous month, ignoring any data that was added after that month.
✑ You must register the minimum number of datasets possible.
You need to register the sales data as a dataset in Azure Machine Learning service workspace.
What should you do?



A. Create a tabular dataset that references the datastore and explicitly specifies each 'sales/mm-yyyy/sales.csv' file every month. Register the dataset with the name sales_dataset each month, replacing the existing dataset and specifying a tag named month indicating the month and year it was registered. Use this dataset for all experiments.

B. Create a tabular dataset that references the datastore and specifies the path 'sales/*/sales.csv', register the dataset with the name sales_dataset and a tag named month indicating the month and year it was registered, and use this dataset for all experiments.

C. Create a new tabular dataset that references the datastore and explicitly specifies each 'sales/mm-yyyy/sales.csv' file every month. Register the dataset with the name sales_dataset_MM-YYYY each month with appropriate MM and YYYY values for the month and year. Use the appropriate month-specific dataset for experiments.

D. Create a tabular dataset that references the datastore and explicitly specifies each 'sales/mm-yyyy/sales.csv' file. Register the dataset with the name sales_dataset each month as a new version and with a tag named month indicating the month and year it was registered. Use this dataset for all experiments, identifying the version to be used based on the month tag as necessary.

Suggested Answer: *B*
Specify the path. Example: The following code gets the workspace existing workspace and the desired datastore by name. And then passes the datastore and file locations to the path parameter to create a new TabularDataset, weather_ds. from azureml.core import Workspace, Datastore, Dataset datastore_name = 'your datastore name' # get existing workspace workspace = Workspace.from_config() # retrieve an existing datastore in the workspace by name datastore = Datastore.get(workspace, datastore_name) # create a TabularDataset from 3 file paths in datastore datastore_paths = [(datastore, 'weather/2018/11.csv'), (datastore, 'weather/2018/12.csv'), (datastore, 'weather/2019/*.csv')] weather_ds = Dataset.Tabular.from_delimited_files(path=datastore_paths)

Community vote distribution
D (89%) B (11%)

gamezone25 Highly Voted 4 years, 10 months ago
D seems to be the correct answer. B does not allow you to get the data from before a specific month. With D you create only one dataset with multiple versions (1 version per month). Similar example in 'Versioning best practice': https://docs.microsoft.com/en-us/azure/machine-learning/how-to-version-track-datasetsI agree. The example shown in the link below does exactly what is being asked in the question. https://docs.microsoft.com/en-us/azure/machine-learning/how-to-version-track-datasets#versioning-best-practiceYou must register the minimum number of datasets possible. D is not correct, because you will have to do this manually each month,?But B you can't select by (each) Month.But D don't satisfy the last requirement that register minimal data set possible since each specific sales file need to register in option D. Given answer B seems correct as it fulfils all conditions.I guess you can by using module like Split or Filter data? You can specify the condition to get data before a particular month

upvoted 28 times

gamezone25 Highly Voted 4 years, 10 months ago

D seems to be the correct answer. B does not allow you to get the data from before a specific month. With D you create only one dataset with multiple versions (1 version per month). Similar example in 'Versioning best practice': https://docs.microsoft.com/en-us/azure/machine-learning/how-to-version-track-datasets

upvoted 28 times

chaudha4 4 years, 9 months ago

I agree. The example shown in the link below does exactly what is being asked in the question. https://docs.microsoft.com/en-us/azure/machine-learning/how-to-version-track-datasets#versioning-best-practiceYou must register the minimum number of datasets possible. D is not correct, because you will have to do this manually each month,?But B you can't select by (each) Month.

upvoted 2 times

chaudha4 4 years, 9 months ago

I agree. The example shown in the link below does exactly what is being asked in the question. https://docs.microsoft.com/en-us/azure/machine-learning/how-to-version-track-datasets#versioning-best-practice

upvoted 2 times

levm39 4 years, 8 months ago

You must register the minimum number of datasets possible. D is not correct, because you will have to do this manually each month,?But B you can't select by (each) Month.

upvoted 4 times

levm39 4 years, 8 months ago

You must register the minimum number of datasets possible. D is not correct, because you will have to do this manually each month,?

upvoted 4 times

YipingRuan 4 years, 7 months ago

But B you can't select by (each) Month.

upvoted 1 times

Shailen 4 years, 2 months ago

But D don't satisfy the last requirement that register minimal data set possible since each specific sales file need to register in option D. Given answer B seems correct as it fulfils all conditions.

upvoted 4 times

chevyli 3 years, 5 months ago

I guess you can by using module like Split or Filter data? You can specify the condition to get data before a particular month

upvoted 3 times

TheCyanideLancer 4 years, 1 month ago

Quick update, verified, correct ans is D. Cross checked in coursera and validated there.

upvoted 19 times

jmh106 Most Recent 7 months, 3 weeks ago

Opción B (CORRECTA) Un único dataset tabular Usa la ruta con comodín: sales/*/sales.csv Carga todos los meses automáticamente, incluso los futuros. Permite filtrar por mes/año en los experimentos usando una columna month. Cumple todos los requisitos con un solo dataset. □ Esta es exactamente la práctica recomendada por Azure ML.

upvoted 1 times

Lion007 2 years, 1 month ago

The Correct answer is: D Option D is the most appropriate choice because it allows for both the inclusion of all data to date for general training and the ability to use specific versions for experiments that require data up to a particular month. The "minimum number of datasets" can be interpreted as the minimum number of distinct dataset entities registered in the workspace. With versioning (Option D), you're still working with one dataset entity, but with multiple versions, which aligns with the requirement of minimal dataset registration. Justification: - Versioning in Azure Machine Learning allows you to handle the evolving data by creating new versions of the dataset each month, without increasing the number of dataset entities in the workspace. - By using version tags, you can manage and reference the appropriate data snapshot for experiments as needed. - This approach offers a balance between efficient data management and the ability to run experiments on specific subsets of the data as of a given date, thus meeting all the stated requirements.

upvoted 4 times

Kanwal001 2 years, 5 months ago

On exam 28/08/2023..

upvoted 4 times

Depayser 2 years, 9 months ago

Selected Answer: D

Option D

upvoted 1 times

Depayser 2 years, 9 months ago

Option DChatGPT agrees.

upvoted 1 times

phydev 2 years, 7 months ago

ChatGPT agrees.

upvoted 1 times

MarinaMijailovic 2 years, 9 months ago

A: *replaces* the the existing dataset -> can't directly filter data before the specific month B: captures all the sales data from different folders in *one dataset* -> can't can't directly filter data before the specific month C: requires registering multiple datasets D: satisfies all the requirements

upvoted 1 times

Yuriy_Ch 2 years, 11 months ago

Exactly this question was on exam 07/03/2023

upvoted 2 times

Yuriy_Ch 2 years, 11 months ago

Exactly this question was on exam 07/03/2023Is Answer B or D?

upvoted 2 times

Jit1981 2 years, 10 months ago

Is Answer B or D?

upvoted 2 times

mamau 3 years ago

B. Create a tabular dataset that references the datastore and specifies the path 'sales/*/sales.csv', register the dataset with the name sales_dataset and a tag named month indicating the month and year it was registered, and use this dataset for all experiments. This option meets all the requirements of the problem statement: ✑ The dataset loads all of the sales data to date into a structure that can be easily converted to a dataframe. ✑ You can create experiments that use only data that was created before a specific previous month, ignoring any data that was added after that month by filtering the dataset based on the "month" tag. ✑ The minimum number of datasets possible is registered (only one).

upvoted 3 times

phdykd 3 years ago

Option D satisfies the last requirement of registering the minimum number of datasets possible. While option B uses a single dataset that references the entire path 'sales/*/sales.csv', it still requires registering the dataset every month with a new tag indicating the month and year. In comparison, option D registers each month's sales data as a new version of the same dataset with a tag indicating the month and year. This allows you to only have to register one dataset instead of multiple datasets, minimizing the number of registered datasets. Option B does not satisfy the requirement of being able to create experiments that use only data that was created before a specific previous month as it only references the entire path and not individual files for each month

upvoted 2 times

Edriv 3 years, 2 months ago

Option C

upvoted 2 times

Arend78 3 years, 2 months ago

If I look at the explanation for the "correct" (?) answer B, it seems that they mean to ask "How to load CSVs form the appropriate folders using the least amount of lines?" In the explanation they use an asterix. Not a very clear question i.m.o.

upvoted 1 times

fvil 3 years, 3 months ago

On exam 07/11/2022

upvoted 1 times

victorafb 3 years, 3 months ago

on exam 16/10/2022 I've answer D

upvoted 3 times

ning 3 years, 9 months ago

Absolutely correct, one dataset with different versions. Versions are NOT the same as different dataset!

upvoted 2 times

JTWang 3 years, 10 months ago

on exam 04/22/2022

upvoted 1 times

azurelearner666 3 years, 10 months ago

Correct answer is D, Even this question is really very badly written to promote misunderstanding and confusion.

upvoted 3 times

DRAG DROP -

An organization uses Azure Machine Learning service and wants to expand their use of machine learning.

You have the following compute environments. The organization does not want to create another compute environment.

You need to determine which compute environment to use for the following scenarios.

Which compute types should you use? To answer, drag the appropriate compute environments to the correct scenarios. Each compute environment may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Select and Place:

Environment name	Compute type
nb_server	Compute Instance
aks_cluster	Azure Kubernetes Service
mlc_cluster	Machine Learning Compute

Environments

- nb_server
- aks_cluster
- mlc_cluster

Answer Area

Scenario	Environment
Run an Azure Machine Learning Designer training pipeline.	Environment
Deploying a web service from the Azure Machine Learning designer.	Environment

Suggested Answer:



Training targets	Automated ML	ML pipelines	Azure Machine Learning designer
Local computer	yes		
Azure Machine Learning compute cluster	yes & hyperparameter tuning	yes	yes
Azure Machine Learning compute instance	yes & hyperparameter tuning	yes	yes
Remote VM	yes & hyperparameter tuning	yes	
Azure Databricks	yes (SDK local mode only)	yes	
Azure Data Lake Analytics		yes	
Azure HDInsight		yes	
Azure Batch		yes	



Box 1: nb_server -

Box 2: mlc_cluster - With Azure Machine Learning, you can train your model on a variety of resources or environments, collectively referred to as compute targets. A compute target can be a local machine or a cloud resource, such as an Azure Machine Learning Compute, Azure HDInsight or a remote virtual machine. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target https://docs.microsoft.com/en-us/azure/machine-learning/how-to-set-up-training-targets

stonefl Highly Voted 4 years, 11 months ago

answer should be: mlc_cluster, aks_clustershould be compute instance and AKS. What is machine learning compute?The answer should be: mlc_cluster aks_cluster Here is why: https://docs.microsoft.com/en-us/azure/machine-learning/concept-designer#compute When using AML designer, the supported compute targets are - AML compute, for TRAINING - AKS, for DEPLOYMENTFor production grade model training, use an Azure Machine Learning compute cluster with multi-node scaling capabilities. For production grade model deployment, use Azure Kubernetes Service

cluster. Compute Type "Machine Learning Compute" is "Compute Clusters" under Compute tab. https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-instance

upvoted 108 times

stonefl Highly Voted 4 years, 11 months ago

answer should be: mlc_cluster, aks_cluster

upvoted 108 times

prashantjoge 4 years, 8 months ago

should be compute instance and AKS. What is machine learning compute?

upvoted 11 times

David_Tadeu 3 years, 10 months ago

The answer should be: mlc_cluster aks_cluster Here is why: https://docs.microsoft.com/en-us/azure/machine-learning/concept-designer#compute When using AML designer, the supported compute targets are - AML compute, for TRAINING - AKS, for DEPLOYMENT

upvoted 6 times

jitsblitz 4 years, 5 months ago

For production grade model training, use an Azure Machine Learning compute cluster with multi-node scaling capabilities. For production grade model deployment, use Azure Kubernetes Service cluster. Compute Type "Machine Learning Compute" is "Compute Clusters" under Compute tab. https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-instance

upvoted 5 times

kty 4 years, 11 months ago

https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target to use designer in training, it has to be " compute cluster "

upvoted 8 times

kty 4 years, 11 months ago

https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target to use designer in training, it has to be " compute cluster "As a web service that's used for real-time inference. Web service deployments use one of the following compute targets: Local computer Azure Machine Learning compute instance Azure Container Instances Azure Kubernetes Service Azure Functions

upvoted 8 times

kty 4 years, 11 months ago

As a web service that's used for real-time inference. Web service deployments use one of the following compute targets: Local computer Azure Machine Learning compute instance Azure Container Instances Azure Kubernetes Service Azure Functions

upvoted 1 times

Karthikat Most Recent 1 year, 10 months ago

on exam 3/25/2024

upvoted 5 times

Kanwal001 2 years, 5 months ago

On exam 28/08/2023..what is the answer?

upvoted 4 times

Kanwal001 2 years, 5 months ago

On exam 28/08/2023..

upvoted 4 times

braupi 2 years, 5 months ago

what is the answer?

upvoted 1 times

phydev 2 years, 7 months ago

This was on Exam today (20 July 2023). I chose: mlc_cluster, aks_cluster respectively.

upvoted 3 times

Dilesha 3 years ago

On Exam 17 Feb 2023

upvoted 4 times

phdykd 3 years ago

Azure Machine Learning compute for training, AKS for deployment. https://learn.microsoft.com/en-us/azure/machine-learning/concept-designer#compute

upvoted 2 times

phdykd 3 years ago

Azure Machine Learning compute for training, AKS for deployment. https://learn.microsoft.com/en-us/azure/machine-learning/concept-designer#computeAzure Machine Learning compute cluster Azure Machine Learning compute instance Azure Machine Learning Kubernetes for

designer these three computes can be used, I think in this case compute cluster is the right answer, what is machine learning compute? compute cluster or compute instance?

upvoted 2 times

meysa 3 years ago

Azure Machine Learning compute cluster Azure Machine Learning compute instance Azure Machine Learning Kubernetes for designer these three computes can be used, I think in this case compute cluster is the right answer, what is machine learning compute? compute cluster or compute instance?

upvoted 1 times

ning 3 years, 9 months ago

Question is very poorly written, I believe that mlc --> Azure Machine Learning Compute Instance, nb_server --> some external compute instance linked with Azure Machine Learning Studio????

upvoted 3 times

DingDongSingSong 3 years, 10 months ago

Answer is: 1. Azure ML compute instance/cluster for training using Azure ML designer. Ref Link:https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target#azure-machine-learning-compute-managed. The link also shows that Azure ML Designer also supports Kubernetes compute for training as well. So this answer is a little tricky 2. Deployment of web service via Azure ML designer is to Kubernetes Cluster Ref Link: https://github.com/solliancenet/azure-machine-learning-quickstarts/blob/master/aml-visual-interface/README.md

upvoted 2 times

TheCyanideLancer 4 years, 1 month ago

Agree with stonefl answer,the same question appeared in Coursera practice test and the explanation for this is as below - Machine Learning Compute Cluster supports integration with AML designer training pipeline, and Azure Kubernetes Service supports integration with AML Designer.

upvoted 1 times

ashii007 4 years, 1 month ago

this question is dated. AKS is now a validation option for designer option.

upvoted 1 times

JoshuaXu 4 years, 3 months ago

on exam 6 Nov 2021, but with more questions. be aware of the different steps: training, deploy, real-time inference, and databricks scenario...

upvoted 1 times

hargur 4 years, 4 months ago

on 19Oct2021

upvoted 1 times

kisskeo 4 years, 4 months ago

On Exam 01 Oct 2021

upvoted 1 times

mthombenindhl84 4 years, 5 months ago

on exam 11/9/2021

upvoted 1 times

snsnsnsn 4 years, 5 months ago

on exam 2/9/21

upvoted 2 times

dushmantha 4 years, 5 months ago

On exam 2021/08/31. my ans mcl and aks

upvoted 2 times

HOTSPOT -

You create an Azure Machine Learning compute target named ComputeOne by using the STANDARD_D1 virtual machine image.

ComputeOne is currently idle and has zero active nodes.

You define a Python variable named ws that references the Azure Machine Learning workspace. You run the following Python code:

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

Hot Area:

```
from azureml.core.compute import ComputeTarget, AmlCompute
from azureml.core.compute_target import ComputeTargetException
the_cluster_name = "ComputeOne"
try:
  the_cluster = ComputeTarget(workspace=ws, name=the_cluster_name)
  print('Step1')
except ComputeTargetException:
  config = AmlCompute.provisioning_configuration(vm_size='STANDARD_DS12_v2', max_nodes=4)
  the_cluster = ComputeTarget.create(ws, the_cluster_name, config)
  print('Step2')
```

Answer Area

	Yes	No
A new machine learning compute resource is created with a virtual machine size of STANDARD_DS12_v2 and a maximum of four nodes.	○	○
Any experiments configured to use `the_cluster` will run on ComputeOne.	○	○
The text **Step1** will be printed to the screen.	○	○

Suggested Answer:

Answer Area

	Yes	No
A new machine learning compute resource is created with a virtual machine size of STANDARD_DS12_v2 and a maximum of four nodes.	○	○
Any experiments configured to use `the_cluster` will run on ComputeOne.	○	○
The text **Step1** will be printed to the screen.	○	○

Answer Area

	Yes	No
A new machine learning compute resource is created with a virtual machine size of STANDARD_DS12_v2 and a maximum of four nodes.	○	○
Any experiments configured to use `the_cluster` will run on ComputeOne.	○	○
The text **Step1** will be printed to the screen.	○	○

Box 1: Yes - ComputeTargetException class: An exception related to failures when creating, interacting with, or configuring a compute target. This exception is commonly raised for failures attaching a compute target, missing headers, and unsupported configuration values. Create(workspace, name, provisioning_configuration) Provision a Compute object by specifying a compute type and related configuration. This method creates a new compute target rather than attaching an existing one.

Box 2: Yes -

Box 3: No - The line before print('Step1') will fail. Reference: https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.computetarget

dev2dev Highly Voted 5 years, 5 months ago

Correct answers are No, Yes, Yes

upvoted 103 times

dev2dev Highly Voted 5 years, 5 months ago

Correct answers are No, Yes, YesI completely agree, because the exception will not be generated and the code inside this block (exception) will not be executed.yes given answer is wrong and correct answer is NO Yes YesAgreed, the code given actually represents the compute creation, if there was no computeOne then the code will create one but as in this case there is already a computeOne, Exception block will not be executed

upvoted 103 times

trickerk 5 years, 1 month ago

I completely agree, because the exception will not be generated and the code inside this block (exception) will not be executed.

upvoted 3 times

azurecert2021 5 years, 1 month ago

yes given answer is wrong and correct answer is NO Yes Yes

upvoted 7 times

Bridgerton 4 years, 4 months ago

Agreed, the code given actually represents the compute creation, if there was no computeOne then the code will create one but as in this case there is already a computeOne, Exception block will not be executed

upvoted 7 times

JoshuaXu 4 years, 9 months ago

Tried the code and the answers "No, Yes, Yes" CONFIRMED

upvoted 11 times

smaha_18 Most Recent 8 months, 1 week ago

Correct answers are No, Yes, Yes. As per Dec 2025, SDK v2 version - Python Code from azure.ai.ml import MLClient from azure.ai.ml.entities import AmlCompute from azure.core.exceptions import ResourceNotFoundError compute_name = "ComputeOne" try: compute = ml_client.compute.get(compute_name) print("Step1") except ResourceNotFoundError: compute = AmlCompute(name=compute_name,

size="STANDARD_DS12_v2", min_instances=0, max_instances=4) ml_client.compute.begin_create_or_update(compute) print("Step2") but the logic remains the same

upvoted 1 times

f11c733 2 years, 2 months ago

The answer is incorrect. No, yes, yes is correct because Compute 1 already exists

upvoted 1 times

Nadine_nm 2 years, 12 months ago

In the first line of the code : try : ..... print('Step1') We are checking if the compute already exists or not. Since it has been created the print will work, and so the "exception" sequence won't be executed. the fact that the cluster is not running, will not delete it. Answer should ben then, No, Yes, Yes.

upvoted 1 times

jpalaci22 3 years, 5 months ago

Seen on the exam 20Feb2023

upvoted 4 times

MansoorDataScientist 3 years, 6 months ago

NO,YES,YES

upvoted 2 times

Arend78 3 years, 8 months ago

I agree with the majority saying "Correct answer = No, Yes, Yes" I believe "idle" does not mean offline or non-existent, just "not busy"

upvoted 2 times

azurelearner666 4 years, 4 months ago

Wrong, Should be Yes Yes Yes. My criteria for the first one is that the compute resource is created, that is the cluster, with a vm size of DS12_v2 and max 4 nodes. This compute resource is the cluster definition, not the VM's which are created while used. But, the cluster definition, the Machine Learning Compute Resource IS CREATED. So, Yes, Yes and Yes.

upvoted 1 times

DingDongSingSong 4 years, 4 months ago

Why is the first answer No, and the second answer Yes?

upvoted 1 times

adamwar 4 years, 9 months ago

Correct answer No, Yes, Yes You can test this yourself by making a cluster (cpu-cluster) which can scale to zero nodes then running the code below: from azureml.core.compute import ComputeTarget from azureml.core.compute_target import ComputeTargetException cluster_name = "cpu-cluster" try: the_cluster = ComputeTarget(ws, name=cluster_name) print("Cluster with zero nodes ready to use") except ComputeTargetException: print("Failed to reference cluster with zero nodes")

upvoted 3 times

hargur 4 years, 10 months ago

on 19Oct2021

upvoted 1 times

kisskeo 4 years, 10 months ago

On Exam 01 Oct 2021

upvoted 1 times

snsnsnsn 4 years, 11 months ago

on exam 2/9/21

upvoted 1 times

datamijn 5 years ago

on 2/8/2021

upvoted 2 times

Rosh4yuh 5 years, 1 month ago

on 17/7/2021 Answers are No, Yes, Yes

upvoted 5 times

ljljljlj 5 years, 1 month ago

On exam 2021/7/10

upvoted 2 times

HOTSPOT -

You are developing a deep learning model by using TensorFlow. You plan to run the model training workload on an Azure Machine Learning Compute Instance.

You must use CUDA-based model training.

You need to provision the Compute Instance.

Which two virtual machines sizes can you use? To answer, select the appropriate virtual machine sizes in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Virtual machine size

Search by name...

Name ↑	vCPUs	GPUs	RAM	Resource disk
BASIC_A0	1		0.75 GB	20 GB
STANDARD_D3_V2	4		14 GB	200 GB
STANDARD_E64_V3	64		432 GB	1,600 GB
STANDARD_M64LS	64		512 GB	2,000 GB
STANDARD_NC12	12	2	112 GB	680 GB
STANDARD_NC24	24	4	224 GB	1,440 GB

Suggested Answer:

Answer Area

Virtual machine size

Search by name...

Name ↑	vCPUs	GPUs	RAM	Resource disk
BASIC_A0	1		0.75 GB	20 GB
STANDARD_D3_V2	4		14 GB	200 GB
STANDARD_E64_V3	64		432 GB	1,600 GB
STANDARD_M64LS	64		512 GB	2,000 GB
STANDARD_NC12	12	2	112 GB	680 GB
STANDARD_NC24	24	4	224 GB	1,440 GB

Answer Area

Virtual machine size

Search by name...

Name ↑	vCPUs	GPUs	RAM	Resource disk
BASIC_A0	1		0.75 GB	20 GB
STANDARD_D3_V2	4		14 GB	200 GB
STANDARD_E64_V3	64		432 GB	1,600 GB
STANDARD_M64LS	64		512 GB	2,000 GB
STANDARD_NC12	12	2	112 GB	680 GB
STANDARD_NC24	24	4	224 GB	1,440 GB

CUDA is a parallel computing platform and programming model developed by Nvidia for general computing on its own GPUs (graphics processing units). CUDA enables developers to speed up compute-intensive applications by harnessing the power of GPUs for the parallelizable part of the computation. Reference: https://www.infoworld.com/article/3299703/what-is-cuda-parallel-programming-for-gpus.html

chaudha4 Highly Voted 4 years, 3 months ago

In case it is not obvious, you pick the ones that support GPU since CUDA requirement is mentioned in the question.

upvoted 24 times

jpalaci22 2 years, 5 months ago

Seen on the exam 20Feb2023

upvoted 3 times

jpalaci22 2 years, 5 months ago

Seen on the exam 20Feb2023Hello, How was your experience ?

upvoted 3 times

Oreo_Princess Most Recent 2 years, 2 months ago

Hello, How was your experience ?

upvoted 1 times

DodoScript 2 years, 12 months ago

Correct, you need GPU's for CUDA

upvoted 4 times

ljljljlj 4 years, 1 month ago

On exam 2021/7/10

upvoted 3 times

azurecert2021 4 years, 1 month ago

given answer is correct as per below link NC series support CUDA model training. https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target#compute-isolation

upvoted 3 times

shuvovertigo 4 years, 2 months ago

correct

upvoted 2 times

SaulG 4 years, 2 months ago

Answer is Correct!

upvoted 1 times

DRAG DROP -

You are analyzing a raw dataset that requires cleaning.

You must perform transformations and manipulations by using Azure Machine Learning Studio.

You need to identify the correct modules to perform the transformations.

Which modules should you choose? To answer, drag the appropriate modules to the correct scenarios. Each module may be used once, more than once, or not at all.

You may need to drag the split bar between panes or scroll to view content.

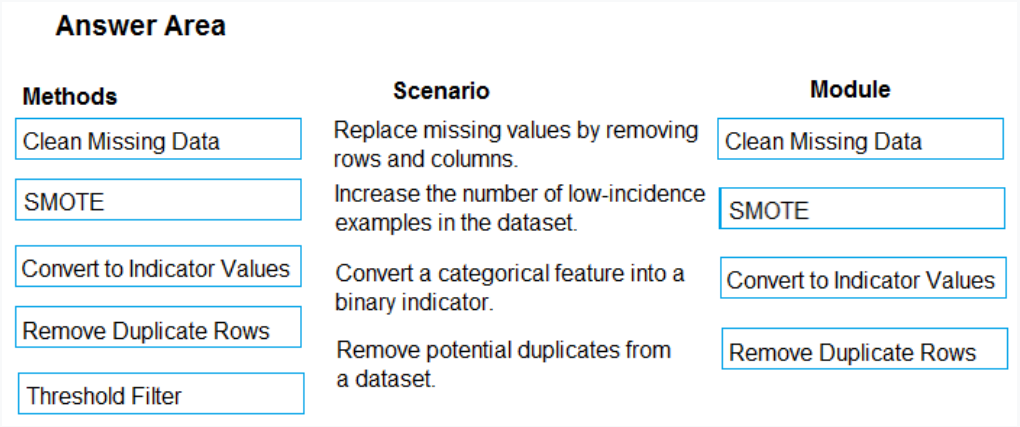
NOTE: Each correct selection is worth one point.

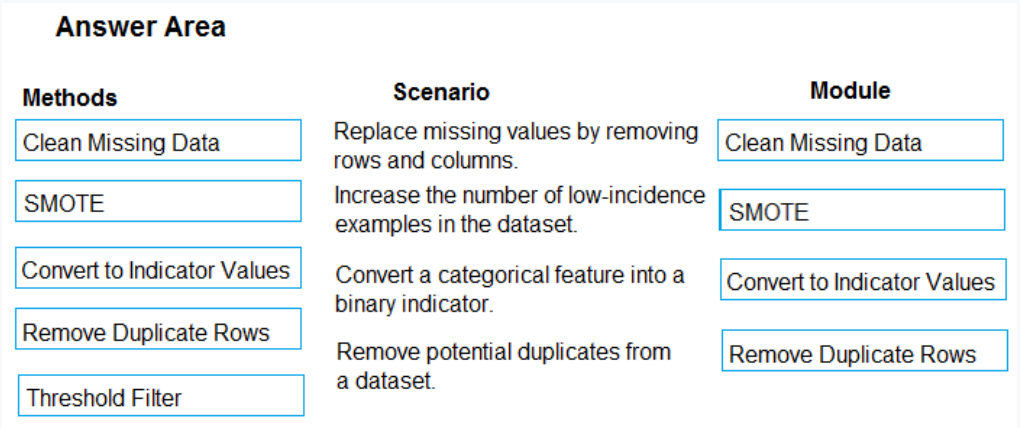
Select and Place:

Answer Area

Methods	Scenario	Module
Clean Missing Data	Replace missing values by removing rows and columns.	
SMOTE	Increase the number of low-incidence examples in the dataset.	
Convert to Indicator Values	Convert a categorical feature into a binary indicator.	
Remove Duplicate Rows	Remove potential duplicates from a dataset.	
Threshold Filter		

Suggested Answer:





Box 1: Clean Missing Data -

Box 2: SMOTE - Use the SMOTE module in Azure Machine Learning Studio to increase the number of underepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases. Box 3: Convert to Indicator Values Use the Convert to Indicator Values module in Azure Machine Learning Studio. The purpose of this module is to convert columns that contain categorical values into a series of binary indicator columns that can more easily be used as features in a machine learning model.

Box 4: Remove Duplicate Rows - Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/convert-to-indicator-values

Tehseen Highly Voted 3 years, 1 month ago

Seems true

upvoted 16 times

azurelearner666 Highly Voted 2 years, 4 months ago

Correct! Note that "Convert to indicator values" (re-encode Categorical values into binary columns) is also called One-Hot Encoding, https://docs.microsoft.com/en-us/learn/modules/prepare-data-for-machine-learning-azure-databricks/6-perform-data-encoding

upvoted 6 times

DodoScript Most Recent 1 year, 12 months ago

Answer is correct

upvoted 4 times

k1ngs1zed 2 years, 4 months ago

Correct!

upvoted 3 times

Question #28 *Topic 2*

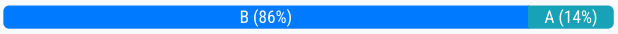
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are using Azure Machine Learning Studio to perform feature engineering on a dataset.
You need to normalize values to produce a feature column grouped into bins.
Solution: Apply an Entropy Minimum Description Length (MDL) binning mode.
Does the solution meet the goal?

A. Yes

B. No

Suggested Answer: *A*
Entropy MDL binning mode: This method requires that you select the column you want to predict and the column or columns that you want to group into bins. It then makes a pass over the data and attempts to determine the number of bins that minimizes the entropy. In other words, it chooses a number of bins that allows the data column to best predict the target column. It then returns the bin number associated with each row of your data in a column named quantized. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

Community vote distribution



GaborO Highly Voted 5 years, 10 months ago
MDL doesn't normalize values, so I think the correct answer is B.
upvoted 30 times

GaborO Highly Voted 5 years, 10 months ago
MDL doesn't normalize values, so I think the correct answer is B.agree with GaborO, it should be Quantile Normalization and not Entropy MDL "If you select the Quantiles binning mode, use the Quantile normalization option to determine how values are normalized prior to sorting into quantiles. Note that normalizing values transforms the values, but does not affect the final number of bins" https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins#bkmk_EffectsI agree: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins See the table "Module parameters" Name Range Type Default Description Binning mode List QuantizationMode Quantiles Choose a binning method Quantile normalization any BinningNormalization Choose the method for normalizing quantiles For BinningNormalization should be used Quantile Normalization.It is not just 'MDL', it is 'entropy MDL binning mode'.
upvoted 30 times

febriyanasn 5 years ago
agree with GaborO, it should be Quantile Normalization and not Entropy MDL "If you select the Quantiles binning mode, use the Quantile normalization option to determine how values are normalized prior to sorting into quantiles. Note that normalizing values transforms the values, but does not affect the final number of bins" https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins#bkmk_Effects
upvoted 4 times

trickerk 4 years, 7 months ago
I agree: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins See the table "Module parameters" Name Range Type Default Description Binning mode List QuantizationMode Quantiles Choose a binning method Quantile normalization any BinningNormalization Choose the method for normalizing quantiles For BinningNormalization should be used Quantile Normalization.
upvoted 4 times

meswapnilspal 5 years, 9 months ago
It is not just 'MDL', it is 'entropy MDL binning mode'.
upvoted 8 times

David_Tadeu 3 years, 10 months ago
According to https://docs.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/group-data-into-bins, you can specify the following binning modes: - Entropy MDL - Quantiles - Equal Width - Custom Edges - Equal Width with Custom Start and Stop From all of these, the only binning mode which supports normalization is Quantiles. In particular, Entropy MDL does NOT support normalization.

upvoted 8 times

astone42 Most Recent 1 year, 7 months ago

It's deprecated and out of scope after 16th January 2025.

upvoted 1 times

SanjayPatwardhan 1 year, 7 months ago

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

upvoted 1 times

NullVoider_0 2 years, 2 months ago

There is no use of entropy when normalizing data.

upvoted 1 times

InversaRadice 2 years, 2 months ago

MDL is correct

upvoted 1 times

Ahmed_Gehad 2 years, 6 months ago

The answer is B. No. Entropy Minimum Description Length (MDL) binning is a technique that can be used to group values into bins. However, it is not a normalization technique. Normalization is a technique that is used to scale values so that they have a similar range. In this case, the goal is to normalize values to produce a feature column grouped into bins. However, the solution of applying an Entropy MDL binning mode will not achieve this goal. Instead, you should use a normalization technique, such as min-max normalization or z-score normalization.

upvoted 1 times

pranav33 2 years, 8 months ago

B. No The solution described does not meet the goal of normalizing values to produce a feature column grouped into bins. Entropy Minimum Description Length (MDL) is a criterion used for model selection and not specifically for binning or normalization of data. MDL is typically used for tasks like feature selection or model complexity estimation, but not for creating bins. To achieve the goal of normalizing values and creating bins, other techniques such as equal-width binning or equal-frequency binning can be used.

upvoted 1 times

mamau 3 years ago

Correct answer B https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/group-data-into-bins

upvoted 1 times

phdykd 3 years ago

No. EMDL binning is used for feature selection, not for feature engineering.

upvoted 1 times

Mebyxu 3 years ago

Correct answer is B

upvoted 1 times

Edriv 3 years, 2 months ago

Option A (Yes)

upvoted 2 times

zehraoneexam 3 years, 11 months ago

It is true. Because quantile is unsupervised .

upvoted 1 times

synapse 3 years, 11 months ago

MDL does not normalize

upvoted 2 times

Maunik 4 years, 7 months ago

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins - Correct answer seems to be A. In in graph it mentions normalized data in to bins

upvoted 2 times

Haet 4 years, 10 months ago

The answer is Entropy MDL reason if you see the explanation it says they want feature column group into bins this is done by only Entropy MDL and not in Quantile Normalization.

upvoted 2 times

Haet 4 years, 10 months ago

The answer is Entropy MDL reason if you see the explanation it says they want feature column group into bins this is done by only Entropy MDL and not in Quantile Normalization.Incorrect. The question requirements states that the data has to be normalised and then binned. You cannot do Normalisation with entropy MDL, it must be quantile normalisation first.

upvoted 2 times

allanm 4 years, 9 months ago

Incorrect. The question requirements states that the data has to be normalised and then binned. You cannot do Normalisation with entropy MDL, it must be quantile normalisation first.

upvoted 2 times

111ssy 5 years, 5 months ago

Entropy MDL: This method requires that you select the column you want to predict and the column or columns that you want to group into bins. It then makes a pass over the data and attempts to determine the number of bins that minimizes the entropy. In other words, it chooses a number of bins that allows the data column to best predict the target column. It then returns the bin number associated with each row of your data in a column named quantized. https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

upvoted 3 times

Question #29 *Topic 2*

HOTSPOT -

You are preparing to use the Azure ML SDK to run an experiment and need to create compute. You run the following code:

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

Hot Area:

```
from azureml.core.compute import ComputeTarget, AmlCompute
from azureml.core.compute_target import ComputeTargetException
ws = Workspace.from_config()
cluster_name = 'aml-cluster'
try:
  training_compute = ComputeTarget(workspace=ws, name=cluster_name)
except ComputeTargetException:
  compute_config = AmlCompute.provisioning_configuration(vm_size='STANDARD_D2_V2', vm_priority='lowpriority',
max_nodes=4)
  training_compute = ComputeTarget.create(ws, cluster_name, compute_config)
  training_compute.wait_for_completion(show_output=True)
```

Answer Area

	Yes	No
If a compute cluster named aml-cluster already exists in the workspace, it will be deleted and replaced.	○	○
The `wait_for_completion()` method will not return until the aml-cluster compute has four active nodes.	○	○
If the code creates a new aml-cluster compute target, it may be preempted due to capacity constraints.	○	○
The aml-cluster compute target is deleted from the workspace after the training experiment completes.	○	○

Suggested Answer:

Answer Area

	Yes	No
If a compute cluster named aml-cluster already exists in the workspace, it will be deleted and replaced.	○	○
The `wait_for_completion()` method will not return until the aml-cluster compute has four active nodes.	○	○
If the code creates a new aml-cluster compute target, it may be preempted due to capacity constraints.	○	○
The aml-cluster compute target is deleted from the workspace after the training experiment completes.	○	○

Answer Area

	Yes	No
If a compute cluster named aml-cluster already exists in the workspace, it will be deleted and replaced.	○	○
The `wait_for_completion()` method will not return until the aml-cluster compute has four active nodes.	○	○
If the code creates a new aml-cluster compute target, it may be preempted due to capacity constraints.	○	○
The aml-cluster compute target is deleted from the workspace after the training experiment completes.	○	○

Box 1: No - If a compute cluster already exists it will be used.

Box 2: Yes - The wait_for_completion method waits for the current provisioning operation to finish on the cluster.

Box 3: Yes - Low Priority VMs use Azure's excess capacity and are thus cheaper but risk your run being pre-empted.

Box 4: No - Need to use training_compute.delete() to deprovision and delete the AmlCompute target. Reference: https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/training/train-on-amlcompute/train-on-amlcompute.ipynb https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.computetarget

iuolu Highly Voted 4 years, 9 months ago

The correct answer: No No Yes NoCorrect

upvoted 65 times

iuolu Highly Voted 4 years, 9 months ago

The correct answer: No No Yes No

upvoted 65 times

Edriv 3 years, 2 months ago

Correct

upvoted 2 times

reini 4 years, 9 months ago

second answer should be no, as the here not specified parameter min_node_count defaults to 'none' - see https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute(class)?view=azure-ml-pyHave you looked at other questions? Are the answers to other questions correct?Exactly. Also technically it defaults to zero, so only the cluster is created and the nodes are created/added whenever they are needed. But response is No, No, Yes, No

upvoted 13 times

reini 4 years, 9 months ago

second answer should be no, as the here not specified parameter min_node_count defaults to 'none' - see https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute(class)?view=azure-ml-py

upvoted 13 times

scott_klein_12345 4 years, 9 months ago

Have you looked at other questions? Are the answers to other questions correct?

upvoted 1 times

azurelearner666 3 years, 10 months ago

Exactly. Also technically it defaults to zero, so only the cluster is created and the nodes are created/added whenever they are needed. But response is No, No, Yes, No

upvoted 2 times

Ejire Most Recent 1 year, 8 months ago

Answer is No NoYes No. If you need voucher for your DP and AI exams contact me on+2348139103938

upvoted 1 times

kel_dp_100 2 years, 3 months ago

the second should be Yes, from GPT Yes, your understanding is correct. The wait_for_completion method, when called on a compute target, will wait until the specified compute target (in this case, the 'aml-cluster') has all its nodes provisioned and ready for use. The parameter show_output=True will also display detailed output while waiting, so you can monitor the progress. So, in your specific code, it will not return until the 'aml-cluster' has all four nodes provisioned and ready for use.

upvoted 1 times

MarinaMijailovic 2 years, 9 months ago

NO NO YES NO NO - If it doesn't exist an exception is raised, then a new compute target is created. It doesn't delete and replace an existing one. NO - The wait_for_completion() method will return when the compute target is in the 'Succeeded' or 'Failed' provisioning state. It doesn't specifically wait for all four nodes to become active. YES - The compute target is being created with 'lowpriority' VMs, which can be preempted if Azure needs the capacity. NO - This code snippet does not include any code to delete the compute target after the training experiment completes.

upvoted 5 times

Yuriy_Ch 2 years, 11 months ago

Exactly this question was on exam 07/March/2023

upvoted 2 times

jpalaci22 2 years, 12 months ago

Seen on the exam 20Feb2023

upvoted 2 times

turtle666 3 years, 10 months ago

training_compute created from ComputeTarget class, and wait_for_completion() not have min_node option, different from AmlCompute Class https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.computetarget?view=azure-ml-py#azureml-core-compute-computetarget-wait-for-completion

upvoted 1 times

Thornehead 3 years, 10 months ago

No, Yes, Yes and No. Read the question and coding carefully.

upvoted 3 times

kkkk_jjjj 3 years, 11 months ago

on exam 18/03/2022

upvoted 2 times

TheYazan 3 years, 11 months ago

On march-9-2022

upvoted 1 times

JoshuaXu 4 years, 3 months ago

on exam 6 Nov 2021

upvoted 1 times

hargur 4 years, 4 months ago

on 19Oct2021

upvoted 1 times

Rosh4yuh 4 years, 7 months ago

on 17/7/2021

upvoted 3 times

ljljljlj 4 years, 7 months ago

On exam 2021/7/10

upvoted 4 times

Meg04 4 years, 7 months ago

min_node_count | default value: None Minimum number of nodes to wait for before considering provisioning to be complete. This doesn't have to equal the minimum number of nodes that the compute was provisioned with, however it should not be greater than that.
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute(class)?view=azure-ml-py#wait-for-completion-show-output-false--min-node-count-none--timeout-in-minutes-25--is-delete-operation-false-

upvoted 1 times

azurecert2021 4 years, 7 months ago

yes agree with other correct answer is No No Yes No

upvoted 4 times

Question #30 *Topic 2*

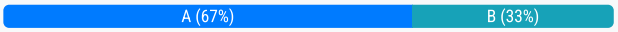
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are a data scientist using Azure Machine Learning Studio.
You need to normalize values to produce an output column into bins to predict a target column.
Solution: Apply a Quantiles normalization with a QuantileIndex normalization.
Does the solution meet the goal?

A. Yes

B. No

Suggested Answer: *B*
Use the Entropy MDL binning mode which has a target column. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

Community vote distribution



kty Highly Voted 5 years, 5 months ago
I think Answer is 'A' If you select the Quantiles binning mode, use the Quantile normalization option to determine how values are normalized prior to sorting into quantiles. Note that normalizing values transforms the values, but does not affect the final number of bins. For an example, see Effects of Different Normalization Methods. The following normalization types are supported: Percent: Values are normalized within the range [0,100] PQuantile: Values are normalized within the range [0,1] QuantileIndex: Values are normalized within the range [1,number of bins]
upvoted 29 times

l2azure Highly Voted 5 years, 4 months ago
Answer is correct, only the binning mode 'Entropy MDL' takes into account the target column. See documentation: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins"You need to normalize values to produce an output column into bins to predict a target column." =/= we need to take into consideration the target column when binningWrong, it is A. and the proper way to normalize the values is with Quantile normalization "If you select the Quantiles binning mode, use the Quantile normalization option to determine how values are normalized prior to sorting into quantiles. Note that normalizing values transforms the values, but does not affect the final number of bins" https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins#bkmk_Effects
upvoted 13 times

l2azure 5 years, 4 months ago
Answer is correct, only the binning mode 'Entropy MDL' takes into account the target column. See documentation: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins
upvoted 13 times

lcgcastro96 3 years, 2 months ago
"You need to normalize values to produce an output column into bins to predict a target column." =/= we need to take into consideration the target column when binning
upvoted 1 times

azurelearner666 4 years, 4 months ago
Wrong, it is A. and the proper way to normalize the values is with Quantile normalization "If you select the Quantiles binning mode, use the Quantile normalization option to determine how values are normalized prior to sorting into quantiles. Note that normalizing values transforms the values, but does not affect the final number of bins" https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins#bkmk_Effects
upvoted 3 times

ashvinn Most Recent 2 years ago
A is to go
upvoted 1 times

NullVoider_0 2 years, 8 months ago

The Quantiles normalization module in Azure ML Studio normalizes values based on quantiles into a specified number of bins. The QuantileIndex normalization module further converts those binned quantiles into indices values. Together these two normalization modules can take an input column and transform it into quantile indices, binning the original values. This quantile binning allows the feature to then be used for targets prediction, meeting that part of the stated goal as well.

upvoted 2 times

Ahmed_Gehad 3 years ago

The answer is B. No. Quantiles normalization is a technique that is used to normalize values by dividing the range of values into equal-sized intervals, or quantiles. QuantileIndex normalization is a technique that is used to normalize values by assigning each value to a quantile based on its rank. In this case, the goal is to normalize values to produce an output column into bins to predict a target column. However, the solution of applying a Quantiles normalization with a QuantileIndex normalization will not achieve this goal. Instead, you should use a normalization technique that is specifically designed for binning, such as Equal Width Binning or Equal Frequency Binning.

upvoted 2 times

phdykd 3 years, 5 months ago

B. No. Quantiles normalization is a normalization technique that transforms the values in a dataset to have a specified distribution, usually a normal distribution. It is not used to produce an output column into bins. Quantile Index normalization is a technique that transforms a dataset so that its values lie between 0 and 1, using the minimum and maximum values in the dataset. It is also not used to produce an output column into bins. To produce an output column into bins, one could use the Discretize module in Azure Machine Learning Studio.

upvoted 2 times

DingDongSingSong 4 years, 4 months ago

The question is poorly worded: You need to normalize values to produce an output column into bins to predict a target column. Why would you bin an output (target) column when you're projecting a target column? That just doesn't make sense.

upvoted 4 times

dija123 4 years, 8 months ago

I go with A

upvoted 2 times

v06ayxop1 4 years, 10 months ago

Entropy MDL is not yet available in new Azure ML but only in Azure ML classic. https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/group-data-into-bins#how-to-configure-group-data-into-bins

upvoted 3 times

jitsblitz 4 years, 11 months ago

"You need to normalize values to produce an output column into bins to predict a target column." Only Quantiles binning normalizes. You are not asked to use target column for binning. Dont get confused with just the mention of target column. All feature selection is to predict target column only. Entropy MDL uses Target column to determine number of bins.

upvoted 8 times

Askme101 5 years, 7 months ago

Yes answer is correct as 'a' as this requires a target column

upvoted 1 times

Question #31 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are creating a new experiment in Azure Machine Learning Studio.
One class has a much smaller number of observations than the other classes in the training set.
You need to select an appropriate data sampling strategy to compensate for the class imbalance.
Solution: You use the Scale and Reduce sampling mode.
Does the solution meet the goal?

A. Yes

B. No

Suggested Answer: *B*
Instead use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode. Note: SMOTE is used to increase the number of underepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases. Incorrect Answers: Common data tasks for the Scale and Reduce sampling mode include clipping, binning, and normalizing numerical values. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/data-transformation-scale-and-reduce

azurelearner666 Highly Voted 3 years, 10 months ago
Response is correct, B. The "Scale and Reduce sampling mode" will not be able to compensate for the class imbalance, so B (no) is the right Answer. The proper response should be SMOTE. More info on https://docs.microsoft.com/en-us/azure/machine-learning/component-reference/smote
upvoted 7 times

Tehseen Highly Voted 4 years, 7 months ago
Correct af
upvoted 6 times

windy610 Most Recent 2 years, 2 months ago
should use SMOTE
upvoted 1 times

nick234987 4 years, 4 months ago
The correct answer is B
upvoted 2 times

Question #32 *Topic 2*

You are analyzing a dataset by using Azure Machine Learning Studio.
You need to generate a statistical summary that contains the p-value and the unique count for each feature column.
Which two modules can you use? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. Computer Linear Correlation

B. Export Count Table

C. Execute Python Script

D. Convert to Indicator Values

E. Summarize Data

Suggested Answer: *B, E*
The Export Count Table module is provided for backward compatibility with experiments that use the Build Count Table (deprecated) and Count Featurizer (deprecated) modules. E: Summarize Data statistics are useful when you want to understand the characteristics of the complete dataset. For example, you might need to know: ✑ How many missing values are there in each column? ✑ How many unique values are there in a feature column? ✑ What is the mean and standard deviation for each column? ✑ The module calculates the important scores for each column, and returns a row of summary statistics for each variable (data column) provided as input. Incorrect Answers: A: The Compute Linear Correlation module in Azure Machine Learning Studio is used to compute a set of Pearson correlation coefficients for each possible pair of variables in the input dataset. C: With Python, you can perform tasks that aren't currently supported by existing Studio modules such as: Visualizing data using matplotlib Using Python libraries to enumerate datasets and models in your workspace Reading, loading, and manipulating data from sources not supported by the Import Data module D: The purpose of the Convert to Indicator Values module is to convert columns that contain categorical values into a series of binary indicator columns that can more easily be used as features in a machine learning model. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/export-count-table https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/summarize-data

Community vote distribution
CE (75%) AE (25%)

lucazav Highly Voted 4 years, 4 months ago
Summarize Data module output also has the following columns: P0.5, P1, P5, P95, P99, P99.5 Probably the person in charge of these tests was not very prepared and did not even check Microsoft's docs, so he confused "percentile" values with "p-" values. By the way, "p-values" of what!?Fully agree, sometimes the questions are a bit "off" (and that's polite...)Indeed, very confusing. Thanks for mentioning this!
upvoted 32 times

lucazav Highly Voted 4 years, 4 months ago
Summarize Data module output also has the following columns: P0.5, P1, P5, P95, P99, P99.5 Probably the person in charge of these tests was not very prepared and did not even check Microsoft's docs, so he confused "percentile" values with "p-" values. By the way, "p-values" of what!?
upvoted 32 times

azurelearner666 2 years, 10 months ago
Fully agree, sometimes the questions are a bit "off" (and that's polite...)
upvoted 2 times

Arend78 2 years, 2 months ago
Indeed, very confusing. Thanks for mentioning this!
upvoted 2 times

phdykd 2 years ago
C. Execute Python Script E. Summarize Data In Azure Machine Learning Studio, you can use the "Summarize Data" module to generate a statistical summary that includes unique count and basic statistics such as mean, standard deviation, minimum, maximum, and quartiles for each feature column. To calculate p-value, you would need to use a "Execute Python Script" module and write a custom code to perform the hypothesis test and compute the p-value. The "Export Count Table" module in Azure Machine Learning Studio only provides a count of the number of instances of each unique value in a single feature column, it doesn't provide statistical summary information like the mean, standard deviation, minimum, maximum, quartiles, or p-value. To get the p-value and a more comprehensive statistical summary, you would need to use the "Execute Python Script" module and write custom code as described in my previous answer.

upvoted 11 times

jmh106 Most Recent 7 months, 3 weeks ago

Qué ocurre en el examen DP-100 Microsoft usa “valor p” de forma laxa En muchos ítems, “valor p” = “salida estadística del módulo” El corrector espera que asocies: “valor p” → Compute Linear Correlation No evalúan estadística, evalúan reconocimiento de módulos Por eso, en el contexto de examen, la respuesta esperada es: A. Compute Linear Correlation E. Summarize Data Aunque conceptualmente sea discutible.

upvoted 1 times

MarinaMijailovic 1 year, 9 months ago

C. Execute Python Script E. Summarize Data The "Execute Python Script" module allows you to run your own Python script, which can be used to calculate any statistic, including unique counts and p-values. The "Summarize Data" module in Azure Machine Learning Studio provides a statistical summary of the input dataset, including count, mean, mode, standard deviation, minimum, and maximum. And like somebody mentioned, it should be percentile and not p-value.

upvoted 3 times

Edriv 2 years, 2 months ago

Options AD

upvoted 1 times

kty 3 years, 11 months ago

Answer is just 'C'

upvoted 4 times

saurabhk1 3 years, 11 months ago

Actually, the question is not clear about which type of p-value, it wants to calculate. Only module that can do it is "Execute a python script"

upvoted 4 times

Askme101 4 years, 1 month ago

it should say percentile values

upvoted 3 times

modschegiebsch 4 years, 7 months ago

Isn't "Execute a python script" actually the only option to calculate the p-value?

upvoted 7 times

Question #33 *Topic 2*

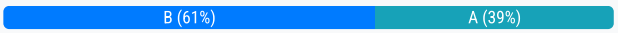
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are analyzing a numerical dataset which contains missing values in several columns.
You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set.
You need to analyze a full dataset to include all values.
Solution: Use the Last Observation Carried Forward (LOCF) method to impute the missing data points.
Does the solution meet the goal?

A. Yes

B. No

Suggested Answer: *B*
Instead use the Multiple Imputation by Chained Equations (MICE) method. Replace using MICE: For each missing value, this option assigns a new value, which is calculated by using a method described in the statistical literature as "Multivariate Imputation using Chained Equations" or "Multiple Imputation by Chained Equations". With a multiple imputation method, each variable with missing data is modeled conditionally using the other variables in the data before filling in the missing values. Note: Last observation carried forward (LOCF) is a method of imputing missing data in longitudinal studies. If a person drops out of a study before it ends, then his or her last observed score on the dependent variable is used for all subsequent (i.e., missing) observation points. LOCF is used to maintain the sample size and to reduce the bias caused by the attrition of participants in a study. Reference: https://methods.sagepub.com/reference/encyc-of-research-design/n211.xml https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/

Community vote distribution



timosi Highly Voted 4 years, 4 months ago
yes, it meets the goal to keep the dimensionality of the feature set. Is it the right approach? without context difficult to say, but if the observations are independent, the answer would be no.
upvoted 17 times

David_Tadeu Highly Voted 3 years, 4 months ago
ID | NAME | T1 | T2 | T3 | 1 | Joe | 37 | 89 | - | 2 | Pip | 39 | - | 80 | 3 | Guy | 89 | 99 | 45 | 4 | Bane | 33 | - | - | 5 | Sue | 77 | - | - | If we applied LOCF, it would turn into ID | NAME | T1 | T2 | T3 | 1 | Joe | 37 | 89 | 89 | 2 | Pip | 39 | 80 | 80 | 3 | Guy | 89 | 99 | 45 | 4 | Bane | 33 | 33 | 33 | 5 | Sue | 77 | 77 | 77 | Hence it would: - CLEAN THE MISSING DATA - MAINTAIN THE DIMENSIONALITY (if it would make sense or no to do it, it would be a different question). The LOCF method can be used when data are longitudinal (i.e. repeated measures have been taken per subject by time point). So, it would make sense to use it in a dataset where each row contains an Olympic athlete and the columns contain their distance attempts for long jump, but it would not make sense to use it in a dataset where each row contains an Olympic athlete and the columns contain their time attempts for 100m, 400m, and 1000m.
upvoted 9 times

ManuelHenriques Most Recent 1 year, 5 months ago
Yes, LOCF maintains numerical dimensionality
upvoted 1 times

84b1989 1 year, 7 months ago
Last Observation Carried Forward (LOCF): The LOCF method fills missing values by carrying forward the last observed value. This method is typically used in time-series data where the order of observations matters. However, it may not be appropriate for non-time-series data or datasets where the assumption of continuity (i.e., the last observed value is a good estimate for missing values) does not hold. Impact on Dimensionality: While LOCF does not reduce the dimensionality of the dataset, it may introduce bias or inaccuracies if the missing values are not related to the last observed values. Better Alternatives: For numerical datasets, other imputation methods such as mean, median, or mode imputation, or more advanced techniques like k-nearest neighbors (KNN) imputation, might be more appropriate and effective.
upvoted 3 times

Ahmed_Gehad 2 years ago
The Last Observation Carried Forward (LOCF) method is a technique that is used to impute missing values by carrying forward the last observed value for that variable. This can be useful for imputing missing values in time series data. However, the LOCF method can be problematic if there are a lot of missing values in a column. This is because the LOCF method will simply carry forward the last observed value, even if that value is an outlier.

This can skew the results of your analysis. In this case, the goal is to clean the missing values using an appropriate operation without affecting the dimensionality of the feature set. However, the LOCF method can affect the dimensionality of the feature set, as it will create new rows in the dataset to store the imputed values. Therefore, the LOCF method is not the best solution for this problem.

upvoted 6 times

PI_Team 2 years, 1 month ago

The correct answer is B. No. The Last Observation Carried Forward (LOCF) method is commonly used for imputing missing values in time series data, where the last observed value is carried forward to replace the missing values. However, this method may not be appropriate for non-time series data or when the missing values are not expected to have a linear or sequential relationship. To clean the missing values in a numerical dataset without affecting the dimensionality of the feature set, other methods such as mean imputation, median imputation, or regression imputation could be considered. These methods calculate the mean, median, or use regression models to estimate the missing values based on the available data without changing the dimensionality of the feature set. SaMYes, ChatGPT agrees.

upvoted 4 times

PI_Team 2 years, 1 month ago

Selected Answer: B

The correct answer is B. No. The Last Observation Carried Forward (LOCF) method is commonly used for imputing missing values in time series data, where the last observed value is carried forward to replace the missing values. However, this method may not be appropriate for non-time series data or when the missing values are not expected to have a linear or sequential relationship.

To clean the missing values in a numerical dataset without affecting the dimensionality of the feature set, other methods such as mean imputation, median imputation, or regression imputation could be considered. These methods calculate the mean, median, or use regression models to estimate the missing values based on the available data without changing the dimensionality of the feature set.

SaM

upvoted 4 times

phydev 2 years ago

Yes, ChatGPT agrees.

upvoted 1 times

DrewSeeSharp 2 years, 6 months ago

No, the solution does not meet the goal of analyzing a full dataset to include all values. While the LOCF method can be useful for imputing missing data points in time series data, it is not appropriate for all types of data and may not accurately reflect the true values of the missing data points. Additionally, imputing missing data using LOCF will not result in a full dataset as the missing values will still exist in the original dataset. Other imputation methods such as mean imputation, median imputation, or multiple imputation may be more appropriate depending on the characteristics of the dataset and the nature of the missing data.

upvoted 4 times

phdykd 2 years, 6 months ago

Yes, the solution meets the goal of analyzing a full dataset that includes all values by using the Last Observation Carried Forward (LOCF) method to impute the missing data points. With LOCF, the missing value in a given time series is filled with the most recent non-missing value in the same series. This method is simple to implement and can be useful in situations where missing values are few and the time series is relatively smooth. However, it is important to note that this method can introduce bias if the missing values are not missing at random and can lead to artificially smoothed data.

upvoted 3 times

Edriv 2 years, 8 months ago

Agree Yes

upvoted 1 times

Amyreads 2 years, 8 months ago

The question also suggests "you are analyzing the full data set" to impute the missing values and therefore MICE would be correct

upvoted 5 times

JTWang 2 years, 10 months ago

The question requirements need to analyze the full dataset and include all values to fill in missing values, so LOCF does not meet the subject requirements. LOCF only fill miss value by last value.As per your answer, The option should be A instead of B

upvoted 2 times

JTWang 2 years, 10 months ago

Selected Answer: B

The question requirements need to analyze the full dataset and include all values to fill in missing values, so LOCF does not meet the subject requirements.

LOCF only fill miss value by last value.

upvoted 2 times

silva_831 2 years, 9 months ago

As per your answer, The option should be A instead of B

upvoted 1 times

ning 3 years, 2 months ago

Absolutely possible to replace data with LOCF to keep the same number of rows and columns of the original dataset! That is the only requirements for the question!

upvoted 4 times

FU_User 3 years, 3 months ago

"You need to analyze a full dataset to include all values." This requirement is not fulfilled as only the last observation is used -> B

upvoted 2 times

azurecert2021 4 years, 1 month ago

not sure but i was not able to find LOCF on below link https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

upvoted 3 times

snac 4 years, 2 months ago

I think this should be a yes. LOCF will keep the dimensionality as much as MICE. And dimensionality was the only real requirement.

upvoted 4 times

prashantjoge 4 years, 2 months ago

the answer should be yes irrespective. How can mice reduce dimensionality?

upvoted 1 times

HOTSPOT -

You are creating a machine learning model in Python. The provided dataset contains several numerical columns and one text column. The text column represents a product's category. The product category will always be one of the following:

- ✑ Bikes
- ✑ Cars
- ✑ Vans
- ✑ Boats

You are building a regression model using the scikit-learn Python package.

You need to transform the text data to be compatible with the scikit-learn Python package.

How should you complete the code segment? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

```
from sklearn import linear_model
import [ pandas as df | numpy as df | scipy as df ]
dataset = df.read_csv("data\\ProductSales.csv")
ProductCategoryMapping = {"Bikes":1, "Cars":2, "Boats": 3,
"Vans": 4}
dataset['ProductCategoryMapping'] =
dataset['ProductCategory'].[ map[ProductCategoryMapping] | reduce[ProductCategoryMapping] | transpose[ProductCategoryMapping] ]
regr = linear_model.LinearRegression()
X_train = dataset[['ProductCategoryMapping', 'ProductSize',
'ProductCost']]
y_train = dataset[['Sales']]
regr.fit(X_train, y_train)
```

Suggested Answer:

Answer Area

```
from sklearn import linear_model
import [pandas as df | numpy as df | scipy as df]
dataset = df.read_csv("data\\ProductSales.csv")
ProductCategoryMapping = {"Bikes":1, "Cars":2, "Boats": 3,
"Vans": 4}
dataset['ProductCategoryMapping'] =
dataset['ProductCategory'].[map[ProductCategoryMapping] | reduce[ProductCategoryMapping] | transpose[ProductCategoryMapping]]
regr = linear_model.LinearRegression()
X_train = dataset[['ProductCategoryMapping', 'ProductSize',
'ProductCost']]
y_train = dataset[['Sales']]
regr.fit(X_train, y_train)
```

Answer Area

```
from sklearn import linear_model
import [pandas as df | numpy as df | scipy as df]
dataset = df.read_csv("data\\ProductSales.csv")
ProductCategoryMapping = {"Bikes":1, "Cars":2, "Boats": 3,
"Vans": 4}
dataset['ProductCategoryMapping'] =
dataset['ProductCategory'].[map[ProductCategoryMapping] | reduce[ProductCategoryMapping] | transpose[ProductCategoryMapping]]
regr = linear_model.LinearRegression()
X_train = dataset[['ProductCategoryMapping', 'ProductSize',
'ProductCost']]
y_train = dataset[['Sales']]
regr.fit(X_train, y_train)
```

Box 1: pandas as df - Pandas takes data (like a CSV or TSV file, or a SQL database) and creates a Python object with rows and columns called data frame that looks very similar to table in a statistical software (think Excel or SPSS for example. Box 2: transpose[ProductCategoryMapping] Reshape the data from the pandas Series to columns. Reference: https://datascienceplus.com/linear-regression-in-python/

dwdas Highly Voted 5 years, 9 months ago

The correct answer is .map(ProductCategoryMapping). I tested it with sample code. Using transpose generates and error.

upvoted 90 times

Cake Highly Voted 5 years, 1 month ago

Also, who imports pandas as df?

upvoted 32 times

Cake 5 years, 1 month ago

Also, who imports pandas as df?Exactly!lol...same question!

upvoted 32 times

shree12 1 year, 11 months ago

Exactly!

upvoted 1 times

HkIsCrazY 5 years ago

lol...same question!

upvoted 3 times

thisiston Most Recent 1 year, 9 months ago

from sklearn import linear_model import pandas as pd # Correct import for handling data frames # Load dataset dataset = pd.read_csv("data\\ProductSales.csv") # Map categories to numbers ProductCategoryMapping = {"Bikes": 1, "Cars": 2, "Boats": 3, "Vans": 4} # Apply the mapping dataset['ProductCategoryMapping'] = dataset['ProductCategory'].map(ProductCategoryMapping) # Initialize the regression model regr = linear_model.LinearRegression() # Specify the features and target variable X_train = dataset[['ProductCategoryMapping', 'ProductSize', 'ProductCost']] Y_train = dataset['Sales'] # Fit the model regr.fit(X_train, Y_train)

upvoted 2 times

sai384957324 1 year, 10 months ago

Pandas as df and map is correct

upvoted 1 times

Lion007 2 years, 1 month ago

1. Import the pandas library to handle data frames: import pandas as pd 2. Use the map function to convert the text categories into numerical data: dataset['ProductCategoryMapping'] = dataset['ProductCategory'].map(ProductCategoryMapping) To clear the confusion of some: Transpose is wrong. In a DataFrame, the transpose() function is used to transpose index and columns. Reflect the DataFrame over its main diagonal by writing rows as columns and vice-versa. Transposing a dataset does not convert categorical variables into a format suitable for regression analysis. This operation does not transform categorical data into a numerical format. In contrast, the 'map' function in pandas is used to match and transform values in a Series based on a given dictionary of correspondences. In the case of the product categories 'Bikes', 'Cars', 'Vans', and 'Boats', these categories are to be mapped to numerical values {1, 2, 3, 4}, as specified in the ProductCategoryMapping dictionary. The purpose of this is to transform the categorical data into a numerical format that can be understood by regression models in scikit-learn.

upvoted 4 times

lena123 3 years, 1 month ago

The answer is map(), Transpose doesn't transforms!!

upvoted 3 times

ashii007 4 years, 1 month ago

map() it is. https://pandas.pydata.org/docs/reference/api/pandas.Series.map.html

upvoted 2 times

ljljljlj 4 years, 7 months ago

On exam 2021/7/10

upvoted 5 times

trickerk 4 years, 7 months ago

"reduce" or "transpose" will generate an error while running the code, further no brackets are allowed in the code but just parenthesis, so need to use .map() instead .map[]. Correct answer: .map

upvoted 2 times

rishi_ram 4 years, 8 months ago

Definitely answer is .map(ProductCategoryMapping). Please run the experiment using transpose will generate error

upvoted 3 times

scott_klein_12345 4 years, 9 months ago

import pandas as pd salesData = pd.DataFrame({'Product': {0: 'Bikes', 1: 'Cars', 2: 'Vans'}, 2017: {0: 34, 1: 65, 2: 48}, 2018: {0: 25, 1: 76, 2: 55}}) productCategoryMapping = {'Bikes':1, 'Cars':2, 'Vans':3} salesData['Mapping'] = salesData['Product'].map(productCategoryMapping) print(salesData) This gives the following output - Product 2017 2018 Mapping 0 Bikes 34 25 1 1 Cars 65 76 2 2 Vans 48 55 3 So I think A) .map is correct answer

upvoted 3 times

wjrmffldrhrl 4 years, 11 months ago

Pandas.map() method need to lambdaOh... I'm wrong.. I just tested it and the map is correct.

upvoted 3 times

wjrmffldrhrl 4 years, 11 months ago

Pandas.map() method need to lambda

upvoted 3 times

wjrmffldrhrl 4 years, 11 months ago

Oh... I'm wrong.. I just tested it and the map is correct.

upvoted 3 times

timotheescheurer 5 years ago

The answer is pd.map ! I just tested it.

upvoted 3 times

wahaha 5 years, 1 month ago

The correct answer should be .map because Transpose only works for dataframes not particular column

upvoted 4 times

tttyyy 5 years, 2 months ago

Transpose is correct. As Product Catogory is categorical, it is unsuitable to be directly used in regression. Transpose translates it into 5 variables, each with value 0 and 1. Better for regression.Transpose is Wrong. The justification provided for the choice of 'transpose' reflects a misunderstanding. Transposing a dataset does not convert categorical variables into a format suitable for regression analysis. Transposing is a matrix operation that flips a table over its diagonal, turning rows into columns and vice versa. This operation does not transform categorical data into a numerical format. In contrast, the 'map' function in pandas is used to match and transform values in a Series based on a given dictionary of correspondences. In the case of the product categories 'Bikes', 'Cars', 'Vans', and 'Boats', these categories are to be mapped to numerical values {1, 2, 3, 4}, as specified in the ProductCategoryMapping dictionary. The purpose of this is to transform the categorical data into a numerical format that can be understood by regression models in scikit-learn.

upvoted 5 times

tttyyy 5 years, 2 months ago

Transpose is correct. As Product Catogory is categorical, it is unsuitable to be directly used in regression. Transpose translates it into 5 variables, each with value 0 and 1. Better for regression.

upvoted 5 times

Lion007 2 years, 1 month ago

Transpose is Wrong. The justification provided for the choice of 'transpose' reflects a misunderstanding. Transposing a dataset does not convert categorical variables into a format suitable for regression analysis. Transposing is a matrix operation that flips a table over its diagonal, turning rows into columns and vice versa. This operation does not transform categorical data into a numerical format. In contrast, the 'map' function in pandas is used to match and transform values in a Series based on a given dictionary of correspondences. In the case of the product categories 'Bikes', 'Cars', 'Vans', and 'Boats', these categories are to be mapped to numerical values {1, 2, 3, 4}, as specified in the ProductCategoryMapping dictionary. The purpose of this is to transform the categorical data into a numerical format that can be understood by regression models in scikit-learn.

upvoted 1 times

SN22 5 years, 4 months ago

the second answer should be map

upvoted 6 times

user11111 5 years, 5 months ago

The answer is Transpose, map is Used for substituting each value in a Series with another value https://www.w3resource.com/pandas/dataframe/dataframe-transpose.php https://www.w3resource.com/pandas/series/series-map.php

upvoted 4 times

user11111 5 years, 5 months ago

The answer is Transpose, map is Used for substituting each value in a Series with another value https://www.w3resource.com/pandas/dataframe/dataframe-transpose.php https://www.w3resource.com/pandas/series/series-map.phpThat's what you want to do.LOL... exactly! wew want to substitute the categorical text entries "Bikes", "Cars",... into another values 1,2,3... Even I would have went for One-Hot Encoding as that's the best way to do it (but not part of the question. For more info, see https://docs.microsoft.com/en-us/learn/modules/prepare-data-for-machine-learning-azure-databricks/6-perform-data-encoding)

upvoted 4 times

adamwar 4 years, 3 months ago

That's what you want to do.LOL... exactly! wew want to substitute the categorical text entries "Bikes", "Cars",... into another values 1,2,3... Even I would have went for One-Hot Encoding as that's the best way to do it (but not part of the question. For more info, see https://docs.microsoft.com/en-us/learn/modules/prepare-data-for-machine-learning-azure-databricks/6-perform-data-encoding)

upvoted 2 times

adamwar 4 years, 3 months ago

That's what you want to do.

upvoted 2 times

azurelearner666 3 years, 10 months ago

LOL... exactly! wew want to substitute the categorical text entries "Bikes", "Cars",... into another values 1,2,3... Even I would have went for One-Hot Encoding as that's the best way to do it (but not part of the question. For more info, see https://docs.microsoft.com/en-us/learn/modules/prepare-data-for-machine-learning-azure-databricks/6-perform-data-encoding)

upvoted 1 times

Question #35 *Topic 2*

You plan to deliver a hands-on workshop to several students. The workshop will focus on creating data visualizations using Python. Each student will use a device that has internet access.
Student devices are not configured for Python development. Students do not have administrator access to install software on their devices. Azure subscriptions are not available for students.
You need to ensure that students can run Python-based data visualization code.
Which Azure tool should you use?

A. Anaconda Data Science Platform

B. Azure BatchAI

C. Azure Notebooks

D. Azure Machine Learning Service

Suggested Answer: *C*
Reference: https://notebooks.azure.com/

Community vote distribution
C (100%)

chaudha4 Highly Voted 4 years, 3 months ago
Correct answer is Google Colab :)this option is not avaliable
upvoted 48 times

chaudha4 Highly Voted 4 years, 3 months ago
Correct answer is Google Colab :)
upvoted 48 times

hargur 3 years, 10 months ago
this option is not avaliable
upvoted 3 times

renuka1234 3 years, 8 months ago
I think anaconda is correct. You don't need admin to install.
upvoted 6 times

renuka1234 3 years, 8 months ago
I think anaconda is correct. You don't need admin to install.I agree. There is no Azure Subscription available, so how could the answer be D?Agree - AnacondaCorrection (Azure Notebook)
upvoted 6 times

michaelmorar 2 years, 8 months ago
I agree. There is no Azure Subscription available, so how could the answer be D?
upvoted 1 times

Edriv 2 years, 8 months ago
Agree - Anaconda
upvoted 1 times

Edriv 2 years, 8 months ago
Agree - AnacondaCorrection (Azure Notebook)
upvoted 1 times

Edriv 2 years, 8 months ago
Correction (Azure Notebook)
upvoted 1 times

phydev Most Recent 2 years ago
On exam 20 July 2023. I chose Azure Notebooks.
upvoted 2 times

krishna1818 2 years, 2 months ago
We can use Azure Notebooks

upvoted 3 times

azurelearner666 3 years, 4 months ago

C is deprecated, not working anymore. D is the only Azure option to do this at the moment. Even doubt that this question with responses is still in the exam... would be too bad for Microsoft's quality.

upvoted 2 times

Skyrocket 3 years, 7 months ago

Answer D is correct, C is deprecated.

upvoted 3 times

ljljljlj 4 years, 1 month ago

On exam 2021/7/10

upvoted 2 times

ljljljlj 4 years, 1 month ago

On exam 2021/7/10what is the correct answer ?if the question is still in exam, then it would have to be D. Microsoft would not leave the question on the exam if the correct answer has been retired.

upvoted 2 times

slash_nyk 4 years, 1 month ago

what is the correct answer ?

upvoted 1 times

slash_nyk 4 years, 1 month ago

what is the correct answer ?if the question is still in exam, then it would have to be D. Microsoft would not leave the question on the exam if the correct answer has been retired.

upvoted 1 times

thhvancouver 4 years ago

if the question is still in exam, then it would have to be D. Microsoft would not leave the question on the exam if the correct answer has been retired.

upvoted 1 times

Maunik 4 years, 1 month ago

Azure machine learning service: https://notebooks.azure.com/ The Azure Notebooks preview has ended. You can enjoy powerful, integrated Jupyter notebooks with the following products and services from Microsoft and GitHub. Notebooks in Visual Studio Code Azure Machine Learning GitHub Codespaces beta

upvoted 1 times

Tehseen 4 years, 1 month ago

The right answer is D as the others are deprecated and ML Service provides multi-user support on compute instance.I agree Azure notebook is already deprecated https://notebooks.azure.com/ If this question still exist, these are possible answer based from website (but from current answer list AML service should be the answer right now) - Notebooks in Visual Studio Code - GitHub Codespaces beta - Azure Machine Learning - Azure Lab Services - GitHub

upvoted 2 times

Tehseen 4 years, 1 month ago

The right answer is D as the others are deprecated and ML Service provides multi-user support on compute instance.

upvoted 2 times

SnowCheetah 4 years, 1 month ago

I agree Azure notebook is already deprecated https://notebooks.azure.com/ If this question still exist, these are possible answer based from website (but from current answer list AML service should be the answer right now) - Notebooks in Visual Studio Code - GitHub Codespaces beta - Azure Machine Learning - Azure Lab Services - GitHub

upvoted 2 times

andre999 4 years, 1 month ago

I am not sure what is the answer then. Is this question obsolete, does anyone know?

upvoted 1 times

andre999 4 years, 1 month ago

I am not sure what is the answer then. Is this question obsolete, does anyone know?Answer C is correct

upvoted 1 times

nick234987 3 years, 10 months ago

Answer C is correct

upvoted 1 times

chaudha4 4 years, 3 months ago

Azure notebook is retired and codespace is still beta.

upvoted 3 times

XiaoQiang 4 years, 10 months ago

Azure notebook will be retired in 2 days...

upvoted 4 times

XiaoQiang 4 years, 10 months ago

Azure notebook will be retired in 2 days...Azure should something like colab, the performance and offer are much more generous.Now the alternative is Github Codespaces: https://docs.microsoft.com/en-us/azure/notebooks/quickstart-export-jupyter-notebook-project#use-notebooks-in-github-codespaces

upvoted 4 times

nato16 4 years, 9 months ago

Azure should something like colab, the performance and offer are much more generous.

upvoted 5 times

lucazav 4 years, 10 months ago

Now the alternative is Github Codespaces: https://docs.microsoft.com/en-us/azure/notebooks/quickstart-export-jupyter-notebook-project#use-notebooks-in-github-codespaces

upvoted 7 times

Question #36 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are analyzing a numerical dataset which contains missing values in several columns.
You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set.
You need to analyze a full dataset to include all values.
Solution: Replace each missing value using the Multiple Imputation by Chained Equations (MICE) method.
Does the solution meet the goal?

A. Yes

B. No

Suggested Answer: *A*
Replace using MICE: For each missing value, this option assigns a new value, which is calculated by using a method described in the statistical literature as "Multivariate Imputation using Chained Equations" or "Multiple Imputation by Chained Equations". With a multiple imputation method, each variable with missing data is modeled conditionally using the other variables in the data before filling in the missing values. Note: Multivariate imputation by chained equations (MICE), sometimes called ɔ€fully conditional specificationɔ€ or ɔ€sequential regression multiple imputationɔ€ has emerged in the statistical literature as one principled method of addressing missing data. Creating multiple imputations, as opposed to single imputations, accounts for the statistical uncertainty in the imputations. In addition, the chained equations approach is very flexible and can handle variables of varying types (e.g., continuous or binary) as well as complexities such as bounds or survey skip patterns. Reference: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/ https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

Community vote distribution
A (100%)

methodidacte Highly Voted 6 years, 1 month ago
So, correct answer is A.
upvoted 8 times

gq_dp1 Most Recent 1 year, 11 months ago
MICE is obsolete. This is no longer the case on v2
upvoted 1 times

phydev 2 years, 7 months ago
ChatGPT says Yes.
upvoted 1 times

pranav33 2 years, 7 months ago
Agree with that
upvoted 1 times

David_Tadeu 3 years, 10 months ago
Related with this question https://www.examtopics.com/discussions/microsoft/view/48586-exam-dp-100-topic-1-question-33-discussion/ And both MICE and LOCF are correctLOCF does not take into consideration the full dataset
upvoted 2 times

David_Tadeu 3 years, 10 months ago
Selected Answer: A
Related with this question https://www.examtopics.com/discussions/microsoft/view/48586-exam-dp-100-topic-1-question-33-discussion/

And both MICE and LOCF are correct
upvoted 2 times

lcgcastro96 2 years, 8 months ago
LOCF does not take into consideration the full dataset
upvoted 2 times

Question #37 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are analyzing a numerical dataset which contains missing values in several columns.
You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set.
You need to analyze a full dataset to include all values.
Solution: Remove the entire column that contains the missing data point.
Does the solution meet the goal?

A. Yes

B. No

Suggested Answer: *B*
Use the Multiple Imputation by Chained Equations (MICE) method. Reference: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

Community vote distribution
B (100%)

pg13 Highly Voted 4 years, 2 months ago
correct answer is no
upvoted 5 times

pranav33 Most Recent 2 years, 1 month ago
No is correct
upvoted 1 times

TheCyanideLancer 3 years, 8 months ago
If you remove columns, dimensions will be affected, hence answer given is correct, which is no.
upvoted 3 times

Shanggavee 3 years, 10 months ago
the answer is no
upvoted 2 times

Shashi_mv 3 years, 10 months ago
the answer is Ýes. we cannot remove the entire column for a single missing data point.So it is No
upvoted 1 times

Shashi_mv 3 years, 10 months ago
the answer is Ýes. we cannot remove the entire column for a single missing data point.
upvoted 1 times

nick234987 3 years, 10 months ago
So it is No
upvoted 4 times

Question #38 *Topic 2*

You are creating a new experiment in Azure Machine Learning Studio. You have a small dataset that has missing values in many columns. The data does not require the application of predictors for each column. You plan to use the Clean Missing Data.
You need to select a data cleaning method.
Which method should you use?

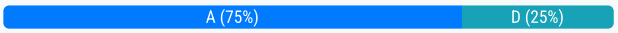
A. Replace using Probabilistic PCA

B. Normalization

C. Synthetic Minority Oversampling Technique (SMOTE)

D. Replace using MICE

Suggested Answer: *A*
Replace using Probabilistic PCA: Compared to other options, such as Multiple Imputation using Chained Equations (MICE), this option has the advantage of not requiring the application of predictors for each column. Instead, it approximates the covariance for the full dataset. Therefore, it might offer better performance for datasets that have missing values in many columns. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

Community vote distribution


ajaysdr Most Recent 1 year, 7 months ago
The MICE (Multiple Imputation by Chained Equations) method is effective for imputing missing values by considering the relationships between different columns, making it suitable for datasets with many missing values
upvoted 1 times

raidenstrike1945 1 year, 8 months ago
CoPilot game me this ans: D. Replace using MICE (Multiple Imputation by Chained Equations) MICE is an effective imputation technique that can handle multiple columns with missing values by using regression models to iteratively impute the missing data, making it suitable for your needs.
upvoted 1 times

Hisayuki 1 year, 9 months ago
The point is "The data does not require the application of predictors for each column." So, it means reducing the dimension and use the PCA - Primary Component Analysis
upvoted 3 times

PI_Team 2 years, 1 month ago
Question is outdated in my opinion. In Clean Missing Data, you can see only: Replace with meanmedian/mode/ and remove enitre row/column https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/clean-missing-data?view=azureml-api-2
upvoted 4 times

phdykd 2 years, 1 month ago
A is the answer. It is in classic version.
upvoted 1 times

krishna1818 2 years, 2 months ago
As a predictor is not required we can use PPCA method
upvoted 2 times

ajay0011 2 years, 4 months ago
Answer is PPCA. MICE is wrong totally check documentation.
upvoted 1 times

phdykd 2 years, 6 months ago
D. Replace using MICE (Multiple Imputation by Chained Equations) is a method that should be used to clean missing data in this scenario. It is commonly used when the data has missing values and the aim is to impute the missing values while preserving the relationships among variables in the data. A. Replace using Probabilistic PCA (Principal Component Analysis) is not the most suitable method for cleaning missing data in this scenario, as it is typically used for dimensionality reduction and feature extraction, rather than imputing missing values. The method of choice for cleaning missing data in this scenario is D. Replace using MICE (Multiple Imputation by Chained Equations), as it is commonly used for imputing missing values while preserving the relationships among variables in the data.

upvoted 4 times

ruggerofreddi 3 years, 2 months ago

PCA is for dimentionality reduction: it diagonalize the covariance matrix (being simmetric for the spectral theorem u can always diagonalize it) and than cuts off the dimensions with small eigenvalues/variance... I am not aware of any variant of this algoritm to impute missing values. do you have any reference? thank you

upvoted 1 times

ruggerofreddi 3 years, 2 months ago

PCA is for dimentionality reduction: it diagonalize the covariance matrix (being simmetric for the spectral theorem u can always diagonalize it) and than cuts off the dimensions with small eigenvalues/variance... I am not aware of any variant of this algoritm to impute missing values. do you have any reference? thank youOnly did it once in uni, but PCA is a legit method for imputing missing values. If I remember well the whole idea was that you generate the missing values through a linear regression using the features z generated by the PCA process. Either way, I might be very wrong and this link seems to explain better than I do: https://stats.stackexchange.com/a/43125

upvoted 1 times

lewitt 2 years, 11 months ago

Only did it once in uni, but PCA is a legit method for imputing missing values. If I remember well the whole idea was that you generate the missing values through a linear regression using the features z generated by the PCA process. Either way, I might be very wrong and this link seems to explain better than I do: https://stats.stackexchange.com/a/43125

upvoted 1 times

ning 3 years, 3 months ago

MICE vs PPCA, this is not so easy to answer in practice, for exam purpose, I agree with A

upvoted 1 times

MohammadKhubeb 3 years, 6 months ago

A, is the correct answer. For dimension reduction. PCA algo is significantly used.

upvoted 1 times

adamwar 3 years, 9 months ago

What does "application of predictors" for each column mean?I have the same question, could someone plz explain?I think this just means, that MICE uses the last value before a NULL value to calculate a good representive for this NULL values, but in case there are a lot of NULL values this technique isn´t a good solution and therefor Probabilistic PCA (which doesn´t need the predessesor values) is the better choice.You're describing Last Observation Carried Forward not MICE. Application of predictors reference makes no sense here with respect to data cleansingIt means you will not need to used all the other columns to predict (or replace) the missing values in one column. Basically it's saying that you will not apply a method like linear regression using all the other columns to fill the missing column. That's what MICE do, so they said that to you eliminate that option

upvoted 2 times

adamwar 3 years, 9 months ago

What does "application of predictors" for each column mean?

upvoted 2 times

Samuela 3 years, 9 months ago

I have the same question, could someone plz explain?

upvoted 1 times

Samuela 3 years, 9 months ago

I have the same question, could someone plz explain?I think this just means, that MICE uses the last value before a NULL value to calculate a good representive for this NULL values, but in case there are a lot of NULL values this technique isn´t a good solution and therefor Probabilistic PCA (which doesn´t need the predessesor values) is the better choice.You're describing Last Observation Carried Forward not MICE. Application of predictors reference makes no sense here with respect to data cleansing

upvoted 1 times

Sichlis 3 years, 8 months ago

I think this just means, that MICE uses the last value before a NULL value to calculate a good representive for this NULL values, but in case there are a lot of NULL values this technique isn´t a good solution and therefor Probabilistic PCA (which doesn´t need the predessesor values) is the better choice.You're describing Last Observation Carried Forward not MICE. Application of predictors reference makes no sense here with respect to data cleansing

upvoted 3 times

Sichlis 3 years, 8 months ago

I think this just means, that MICE uses the last value before a NULL value to calculate a good representive for this NULL values, but in case there are a lot of NULL values this technique isn´t a good solution and therefor Probabilistic PCA (which doesn´t need the predessesor values) is the better choice.

upvoted 3 times

DingDongSingSong 3 years, 4 months ago

You're describing Last Observation Carried Forward not MICE. Application of predictors reference makes no sense here with respect to data cleansing

upvoted 1 times

Padilha 2 years, 7 months ago

It means you will not need to used all the other columns to predict (or replace) the missing values in one column. Basically it's saying that you will not apply a method like linear regression using all the other columns to fill the missing column. That's what MICE do, so they said that to you eliminate that option

upvoted 1 times

Vipuls 4 years, 7 months ago

Yes, given Answer is right

upvoted 4 times

Andrexx 4 years, 9 months ago

Agree with the answer

upvoted 3 times

Question #39 *Topic 2*

You use Azure Machine Learning Studio to build a machine learning experiment.
You need to divide data into two distinct datasets.
Which module should you use?

A. Split Data

B. Load Trained Model

C. Assign Data to Clusters

D. Group Data into Bins

Suggested Answer: *D*
The Group Data into Bins module supports multiple options for binning data. You can customize how the bin edges are set and how values are apportioned into the bins. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

Community vote distribution
A (85%) D (15%)

sagar21692 Highly Voted 5 years, 9 months ago
Correct Answer is a) Split Data https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data
upvoted 112 times

syed_ahmed Highly Voted 4 years, 7 months ago
There are so many errors in the solution answers. I request the admins to please update the correct answers as they are very misleading.sadly, probs not gonna happen lol
upvoted 39 times

syed_ahmed 4 years, 7 months ago
There are so many errors in the solution answers. I request the admins to please update the correct answers as they are very misleading.
upvoted 39 times

bibabrian 2 years, 10 months ago
sadly, probs not gonna happen lol
upvoted 1 times

dporwal04 Most Recent 2 years, 2 months ago
correct ans is A
upvoted 3 times

daviskv74 2 years, 4 months ago
Chat GPT says A is correct
upvoted 3 times

Ahmed_Gehad 2 years, 6 months ago
The Split Data module in Azure Machine Learning Studio can be used to divide a dataset into two distinct datasets. The module allows you to specify the fraction of the dataset that you want to include in each of the two datasets. The other modules you mentioned are not used to divide data into two distinct datasets. The Load Trained Model module is used to load a trained machine learning model into Azure Machine Learning Studio. The Assign Data to Clusters module is used to assign data points to clusters. The Group Data into Bins module is used to group data points into bins.
upvoted 1 times

MohammadKhubeb 2 years, 9 months ago
Answer should be D, because in the BINS the data get easily visualize. Let say we have 100 rows, and two values - Male and Female. so, through BINS we can easily see the ration these two.
upvoted 2 times

Peeking 2 years, 12 months ago
Split Data
upvoted 2 times

jlopezfelizzola 3 years, 5 months ago
Correct answer is A
upvoted 3 times

ning 3 years, 9 months ago

This is A for sure

upvoted 2 times

Thornehead 3 years, 10 months ago

Answer is surely A, because question is asked about "Two Distinct Datasets". Group binning module only groups like a separator in excel while keeping the data in the same dateset.

upvoted 3 times

zehraoneexam 3 years, 11 months ago

"Use the Split Data component to divide a dataset into two distinct sets."

upvoted 1 times

akshays1909 3 years, 11 months ago

To get "Distinct" data, Split Data works fine.

upvoted 1 times

c18 3 years, 12 months ago

I recently took the exam and it's DCan you know your answer is right? If you choose D and your score is 1000/1000 , it means this choice is correct. How can you know all the rihgt answer when you took the exam?

upvoted 3 times

c18 3 years, 12 months ago

Selected Answer: D

I recently took the exam and it's D

upvoted 3 times

Norasit 3 years ago

Can you know your answer is right? If you choose D and your score is 1000/1000 , it means this choice is correct. How can you know all the rihgt answer when you took the exam?

upvoted 10 times

RainbowJuly 4 years, 1 month ago

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data

upvoted 3 times

Skyrocket 4 years, 1 month ago

Split Data is the correct answer. I request the admins to please update the correct answers as they are very misleading.

upvoted 1 times

greengarden 4 years, 2 months ago

Correct Answer is Split Data

upvoted 1 times

adamwar 4 years, 3 months ago

Who writes the answers for these things?

upvoted 6 times

Question #40 *Topic 2*

You are a lead data scientist for a project that tracks the health and migration of birds. You create a multi-class image classification deep learning model that uses a set of labeled bird photographs collected by experts.
You have 100,000 photographs of birds. All photographs use the JPG format and are stored in an Azure blob container in an Azure subscription.
You need to access the bird photograph files in the Azure blob container from the Azure Machine Learning service workspace that will be used for deep learning model training. You must minimize data movement.
What should you do?

A. Create an Azure Data Lake store and move the bird photographs to the store.

B. Create an Azure Cosmos DB database and attach the Azure Blob containing bird photographs storage to the database.

C. Create and register a dataset by using TabularDataset class that references the Azure blob storage containing bird photographs.

D. Register the Azure blob storage containing the bird photographs as a datastore in Azure Machine Learning service.

E. Copy the bird photographs to the blob datastore that was created with your Azure Machine Learning service workspace.

Suggested Answer: *D*
We recommend creating a datastore for an Azure Blob container. When you create a workspace, an Azure blob container and an Azure file share are automatically registered to the workspace. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-access-data

Community vote distribution
D (100%)

G14th Highly Voted 4 years, 9 months ago
Correct!
upvoted 13 times

Karthikat Most Recent 1 year, 10 months ago
on exam 3/25/2024
upvoted 2 times

Mal42 2 years, 6 months ago
On exam 18 Aug 2023
upvoted 2 times

phydev 2 years, 7 months ago
On exam 20 July 2023.
upvoted 1 times

ajay0011 2 years, 10 months ago
correct
upvoted 1 times

Yuriy_Ch 2 years, 11 months ago
Same or similar was on exam 07/March/2023
upvoted 3 times

practia 3 years ago
The option with the lowest movement.
upvoted 1 times

Edriv 3 years, 2 months ago
Are "D" and "E" the same?
upvoted 1 times

victorafb 3 years, 3 months ago
On exam 16/10/22 i've anwser D
upvoted 1 times

racnaoamo 3 years, 9 months ago
on exam 18-5-22
upvoted 1 times

JTWang 3 years, 10 months ago

on exam 04/22/2022

upvoted 1 times

synapse 3 years, 11 months ago

D.. corrct answer given

upvoted 1 times

hargur 4 years, 4 months ago

on 19Oct2021

upvoted 1 times

kisskeo 4 years, 4 months ago

On Exam 01 Oct 2021

upvoted 1 times

mthombenindhl84 4 years, 5 months ago

on exam 11/9/2021

upvoted 1 times

dushmantha 4 years, 5 months ago

On exam 2021/08/31

upvoted 1 times

ljljljlj 4 years, 7 months ago

On exam 2021/7/10

upvoted 2 times

Question #41 *Topic 2*

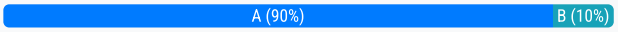
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are analyzing a numerical dataset which contains missing values in several columns.
You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set.
You need to analyze a full dataset to include all values.
Solution: Calculate the column median value and use the median value as the replacement for any missing value in the column.
Does the solution meet the goal?

A. Yes

B. No

Suggested Answer: *B*
Use the Multiple Imputation by Chained Equations (MICE) method. Reference: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/ https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

Community vote distribution



Aleem Highly Voted 5 years, 3 months ago
"You need to analyze a full dataset" just means you can't drop the rows or the columns. Replacing missing data with the median may increase the cardinality but dimensionality is only increased by adding new feature columns. Median replacement is a valid method in this case. The answer should be "Yes".I guess because they mentioned "appropriate" which is vague but it might be the case
upvoted 34 times

Aleem Highly Voted 5 years, 3 months ago
"You need to analyze a full dataset" just means you can't drop the rows or the columns. Replacing missing data with the median may increase the cardinality but dimensionality is only increased by adding new feature columns. Median replacement is a valid method in this case. The answer should be "Yes".
upvoted 34 times

TheYazan 4 years, 5 months ago
I guess because they mentioned "appropriate" which is vague but it might be the case
upvoted 2 times

amarques 7 years ago
In the reference you mentioned it's not explicit that you can only use MICE. There's any reason because we cannot use median? Thank youI think the answer should ONLY be the Mean. Since there are missing values in many columns and MICE uses other columns to calculate the value. Trying to calculate values where the other columns also have missing values could alter the result.You can also use Median for numerical values, as stated in the exercice.
upvoted 14 times

amarques 7 years ago
In the reference you mentioned it's not explicit that you can only use MICE. There's any reason because we cannot use median? Thank you
upvoted 14 times

NaishinMatiri 5 years, 4 months ago
I think the answer should ONLY be the Mean. Since there are missing values in many columns and MICE uses other columns to calculate the value. Trying to calculate values where the other columns also have missing values could alter the result.
upvoted 4 times

NaishinMatiri 5 years, 4 months ago
I think the answer should ONLY be the Mean. Since there are missing values in many columns and MICE uses other columns to calculate the value. Trying to calculate values where the other columns also have missing values could alter the result.You can also use Median for numerical values, as stated in the exercice.
upvoted 4 times

andre999 5 years, 1 month ago
You can also use Median for numerical values, as stated in the exercice.

upvoted 3 times

jefimija Most Recent 1 year, 10 months ago

if values are missing in the entire column, how do you calculate the median?

upvoted 1 times

haby 2 years, 8 months ago

B for sure. Median is only available if the columns/rows are numeric or ordinal, if they are categorical or even string, we have to use Mode or other methods.

upvoted 1 times

haby 2 years, 8 months ago

B for sure. Median is only available if the columns/rows are numeric or ordinal, if they are categorical or even string, we have to use Mode or other methods.The question says it's a numerical dataset.

upvoted 1 times

Anjiiiiiii 2 years, 2 months ago

The question says it's a numerical dataset.

upvoted 2 times

Ahmed_Gehad 3 years ago

The solution meets the goal because it uses an appropriate operation to clean the missing values without affecting the dimensionality of the feature set. The median value is a good choice for imputing missing values because it is not affected by outliers. Additionally, the median value does not affect the dimensionality of the feature set because it is a single value.

upvoted 1 times

snegnik 3 years, 2 months ago

I think B is True. Median replacement "analyze" only one column, but MICE method use all columns for one. Also we can use Probablistic PCA mothod.

upvoted 2 times

MarinaMijailovic 3 years, 3 months ago

If columns are numeric then yes, otherwise no. The problem is we dont have that information.

upvoted 1 times

swatidorge1010 3 years, 3 months ago

Please change answer or justify

upvoted 1 times

swatidorge1010 3 years, 3 months ago

Selected Answer: B

Answer : B As its not mentioned all columns have numeric values.

upvoted 2 times

swatidorge1010 3 years, 3 months ago

Answer : B As its not mentioned all columns have numeric values.It is mentioned "You are analyzing a numerical dataset"

upvoted 2 times

snegnik 3 years, 2 months ago

It is mentioned "You are analyzing a numerical dataset"

upvoted 2 times

ajay0011 3 years, 4 months ago

Yes is the answer, please change it

upvoted 1 times

Edriv 3 years, 8 months ago

Agree A

upvoted 1 times

jlopezfelizzola 3 years, 11 months ago

Should be A

upvoted 3 times

FU_User 4 years, 3 months ago

It is correct as it doesn't add or drop columns. Also "You need to analyze a full dataset" can either mean that the algorithm should take the full existing dataset into account when replacing values (which this one does) or that no existing data should be deleted (median does not delete anything)

upvoted 3 times

VEDPRASAD 4 years, 4 months ago

In an existing project we did median, with clustered dataset. so median works

upvoted 2 times

David_Tadeu 4 years, 4 months ago

Applying the mentioned method to the following dataset 1 | 5 | - | - | 7 | 3 | - | 0 | 2 | 2 | 7 | 4 | 2 | 6 | 9 | - | 2 | - | 3 | - | - | - | 7 | - | would lead to 1 | 4 | 4 | 4 | 7 | 3 | 2 | 0 | 2 | 2 | 7 | 4 | 2 | 6 | 9 | 4 | 2 | 4 | 3 | 5 | 5 | 5 | 7 | 5 | Hence - missing data cleaned - dimensionality preservedoops the entry (1,2) is meant to be a 5 in the dataset below

upvoted 4 times

David_Tadeu 4 years, 4 months ago

Selected Answer: A

Applying the mentioned method to the following dataset

1 | 5 | - | - | 7 | 3 | - | 0 | 2 | 2 | 7 | 4 | 2 | 6 | 9 | - | 2 | - | 3 | - | - | - | 7 | - |

would lead to

1 | 4 | 4 | 4 | 7 | 3 | 2 | 0 | 2 | 2 | 7 | 4 | 2 | 6 | 9 | 4 | 2 | 4 | 3 | 5 | 5 | 5 | 7 | 5 |

Hence - missing data cleaned - dimensionality preserved

upvoted 4 times

David_Tadeu 4 years, 4 months ago

oops the entry (1,2) is meant to be a 5 in the dataset below

upvoted 1 times

synapse 4 years, 5 months ago

Copying a good answer: "You need to analyze a full dataset" just means you can't drop the rows or the columns. Replacing missing data with the median may increase the cardinality but dimensionality is only increased by adding new feature columns. Median replacement is a valid method in this case. The answer should be "Yes"

upvoted 4 times

mikosann 4 years, 11 months ago

I think the answer is correct. Filling missing values with mean/median is a highly used method. It works on each column separately and independently. And it only replaces the missing values, doesn't add any new columns or new rows to the dataset which means it doesn't effect dimensions. Taking the mean/median of the column and replacing missing values is one of the beginner data science topics. MICE can be a better method but this doesn't mean the answer is wrong.I mean the solution is correct. Not the answer.

upvoted 4 times

mikosann 4 years, 11 months ago

I think the answer is correct. Filling missing values with mean/median is a highly used method. It works on each column separately and independently. And it only replaces the missing values, doesn't add any new columns or new rows to the dataset which means it doesn't effect dimensions. Taking the mean/median of the column and replacing missing values is one of the beginner data science topics. MICE can be a better method but this doesn't mean the answer is wrong.

upvoted 4 times

mikosann 4 years, 11 months ago

I mean the solution is correct. Not the answer.

upvoted 3 times

You create an Azure Machine Learning workspace.

You must create a custom role named DataScientist that meets the following requirements:

✑ Role members must not be able to delete the workspace.

✑ Role members must not be able to create, update, or delete compute resources in the workspace.

✑ Role members must not be able to add new users to the workspace.

You need to create a JSON file for the DataScientist role in the Azure Machine Learning workspace.

The custom role must enforce the restrictions specified by the IT Operations team.

Which JSON code segment should you use?

A.

```
{
"Name": "DataScientist",
     "IsCustom": true,
     "Description": "Project Data Scientist role",
     "Actions": ["*"],
     "NotActions":[
     "Microsoft.MachineLearningServices/workspaces/*/delete",
     "Microsoft.MachineLearningServices/workspaces/computes/*/write",
     "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
     "Microsoft.Authorization/*/write"
],
"AssignableScopes": [
  "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
  ]
}
```

B.

```
{
"Name": "DataScientist",
"IsCustom": true,
"Description": "Project Data Scientist role",
"Actions": ["*"],
"NotActions":[],
"AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
]
}
```

C.

```
{
"Name": "DataScientist",
"IsCustom": true,
"Description": "Project Data Scientist role",
"Actions": ["Microsoft.MachineLearningServices/workspaces/*/delete",
     "Microsoft.MachineLearningServices/workspaces/computes/*/write",
     "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
     "Microsoft.Authorization/*/write"
],
"NotActions":[],
"AssignableScopes":[
     "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
]
}
```

D.

```
{
"Name": "DataScientist",
"IsCustom": true,
"Description": "Project Data Scientist role",
"Actions": [],
"NotActions":["*"],
"AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
]
}
```

Suggested Answer:
The following custom role can do everything in the workspace except for the following actions: ✑ It can't create or update a compute resource. ✑ It can't delete a compute resource. ✑ It can't add, delete, or alter role assignments. ✑ It can't delete the workspace. To create a custom role, first construct a role definition JSON file that specifies the permission and scope for the role. The following example defines a custom role named "Data Scientist Custom" scoped at a specific workspace level: data_scientist_custom_role.json : { "Name": "Data Scientist Custom", "IsCustom": true, "Description": "Can run experiment but can't create or delete compute.", "Actions": ["*"], "NotActions": ["Microsoft.MachineLearningServices/workspaces/*/delete", "Microsoft.MachineLearningServices/workspaces/write", "Microsoft.MachineLearningServices/workspaces/computes/*/write", "Microsoft.MachineLearningServices/workspaces/computes/*/delete", "Microsoft.Authorization/*/write"], "AssignableScopes": ["/subscriptions/ /resourceGroups/ /providers/Microsoft.MachineLearningServices/workspaces/ "] } Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-assign-roles

MattAnya Highly Voted 3 years, 1 month ago
On exam 03 Jan2023
upvoted 9 times

ljljljlj Highly Voted 4 years, 7 months ago
On exam 2021/7/10can you please share your experience. How many questions were there from this dump in the exam ?
upvoted 8 times

ljljljlj 4 years, 7 months ago
On exam 2021/7/10
upvoted 8 times

slash_nyk 4 years, 7 months ago
can you please share your experience. How many questions were there from this dump in the exam ?
upvoted 8 times

Karthikat Most Recent 1 year, 10 months ago
on exam 3/25/2024
upvoted 2 times

NullVoider_0 2 years ago
On exam 12-02-2024.
upvoted 2 times

Awooga 2 years ago
On exam 2024-02-06
upvoted 1 times

Mal42 2 years, 6 months ago
On exam 18 Aug 2023
upvoted 3 times

phydev 2 years, 7 months ago
On exam 20 July 2023.
upvoted 2 times

henry_123 2 years, 8 months ago
On exam 2023-05-27
upvoted 2 times

Edriv 3 years, 2 months ago
Option C
upvoted 3 times

therealola 3 years, 8 months ago
Similar question on exam 18-06-22
upvoted 2 times

racnaoamo 3 years, 9 months ago
on exam 18-5-22
upvoted 2 times

kkkk_jjjj 3 years, 11 months ago
similar question on exam 18/03/2022
upvoted 2 times

TheYazan 3 years, 11 months ago

Similar question on march-9-2022

upvoted 2 times

kisskeo 4 years, 4 months ago

On Exam 01 Oct 2021

upvoted 2 times

snsnsnsn 4 years, 5 months ago

on exam 2/9/21

upvoted 2 times

dushmantha 4 years, 5 months ago

On exam 2021/08/31 (similar question)

upvoted 2 times

datamijn 4 years, 6 months ago

on 2/8/2021

upvoted 4 times

Question #43 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are a data scientist using Azure Machine Learning Studio.
You need to normalize values to produce an output column into bins to predict a target column.
Solution: Apply an Equal Width with Custom Start and Stop binning mode.
Does the solution meet the goal?

A. Yes

B. No

Suggested Answer: *B*
Use the Entropy MDL binning mode which has a target column. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

Community vote distribution
B (100%)

dija123 Highly Voted 3 years, 8 months ago
I believe Quantile normalization is the correct answer
upvoted 8 times

chaudha4 Highly Voted 4 years, 3 months ago
Entropy MDL is not available in designer. The answer applies only to studio(classic). If asked in Designer context, I think the answer is actually Yes.
upvoted 6 times

Hisayuki Most Recent 1 year, 9 months ago
Should use Quantiles. The quantile method assigns values to bins based on percentile ranks, because you need to normalize, while Equal Width puts the data into each bin at the same interval between starting and ending values.
upvoted 2 times

Edriv 2 years, 7 months ago
Option A
upvoted 1 times

JTWang 3 years, 4 months ago
Group Data into Bins https://docs.microsoft.com/en-us/azure/machine-learning/component-reference/group-data-into-bins
upvoted 1 times

SnowCheetah 4 years, 1 month ago
"Equal Width with Custom Start and Stop binning mode" only do binning but with goal it's require normalized as well as binning problem. Thus that's incorrect @Askme101 provide possible answer using MDL is correct one https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins
upvoted 6 times

Aleem 4 years, 3 months ago
Entropy MDL may be the best answer, but the question is whether this solution meets the goal stated. The solution stated may not be optimal but does bin the values.
upvoted 5 times

Askme101 4 years, 7 months ago
Entropy MDL is the correct answer see below https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins
upvoted 3 times

concernedCitizen 5 years, 1 month ago
'A' is correct per the link: "Equal Width with Custom Start and Stop: This method is like the Equal Width option, but you can specify both lower and upper bin boundaries."Correct answer is B. Bcoz u have target column.
upvoted 3 times

concernedCitizen 5 years, 1 month ago

'A' is correct per the link: "Equal Width with Custom Start and Stop: This method is like the Equal Width option, but you can specify both lower and upper bin boundaries."

upvoted 3 times

Rickii 5 years, 1 month ago

Correct answer is B. Bcoz u have target column.

upvoted 5 times

juandante 5 years, 3 months ago

Answer B

upvoted 2 times

methodidacte 5 years, 7 months ago

"Studio" is now called "Designer"

upvoted 1 times

Question #44 *Topic 2*

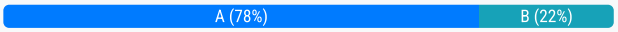
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are a data scientist using Azure Machine Learning Studio.
You need to normalize values to produce an output column into bins to predict a target column.
Solution: Apply a Quantiles binning mode with a PQuantile normalization.
Does the solution meet the goal?

A. Yes

B. No

Suggested Answer: *B*
Use the Entropy MDL binning mode which has a target column. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

Community vote distribution



modschegiebsch Highly Voted 6 years, 1 month ago
Answer is B, because Quantile binning is not supervised. The binning is independent of the target column. You cannot use the output to predict the target directly.
upvoted 22 times

kty Highly Voted 5 years, 4 months ago
If you select the Quantiles binning mode, use the Quantile normalization option to determine how values are normalized prior to sorting into quantiles. Note that normalizing values transform the values, but does not affect the final number of bins

Entropy MDL: This method requires that you select the column you want to predict and the column or columns that you want to group into bins. It then makes a pass over the data and attempts to determine the number of bins that minimizes the entropy. In other words, it chooses a number of bins that allows the data column to best predict the target column

I think the answer is 'A'
upvoted 15 times

kty 5 years, 4 months ago
If you select the Quantiles binning mode, use the Quantile normalization option to determine how values are normalized prior to sorting into quantiles. Note that normalizing values transform the values, but does not affect the final number of bins Entropy MDL: This method requires that you select the column you want to predict and the column or columns that you want to group into bins. It then makes a pass over the data and attempts to determine the number of bins that minimizes the entropy. In other words, it chooses a number of bins that allows the data column to best predict the target column I think the answer is 'A'Totally agree with you
upvoted 15 times

dija123 4 years, 8 months ago
Totally agree with you
upvoted 2 times

FactCheckr4 Most Recent 2 years ago
Why the Solution Doesn't Meet the Goal: Quantiles Binning vs. PQuantile Normalization: While quantiles binning directly addresses the goal of creating bins for normalization, PQuantile normalization is not designed specifically for binning data into discrete categories. PQuantile normalization is more about adjusting data distributions rather than creating discrete bins. To achieve the goal of normalizing values into bins, you should use Quantiles Binning directly. PQuantile normalization does not achieve this goal effectively because its purpose is to normalize data distributions rather than to bin data into quantile-based categories. Thus, the solution of applying a Quantiles binning mode with PQuantile normalization does not fully meet the goal of binning values for predicting a target column.
upvoted 1 times

NullVoider_0 2 years, 8 months ago

Using Quantiles binning mode with PQuantile normalization in Azure Machine Learning Studio is an appropriate solution for normalizing values and transforming them into bins, which can aid in the prediction of a target column in a machine learning model. This method is effective for creating evenly distributed bins based on the data's distribution, which can be beneficial for various predictive modeling tasks.

upvoted 1 times

PI_Team 2 years, 8 months ago

I need to correct my previous comment: The solution of using Quantiles binning mode with PQuantile normalization in Azure Machine Learning Studio is valid. Quantiles binning discretizes data based on percentile ranks, and the PQuantile normalization option within this mode normalizes values within a [0,1] range before sorting into quantiles. This is specific to preparing data for quantile binning and is not general data normalization. Therefore, this approach aligns with Azure Machine Learning Studio's capabilities for data preparation in quantile binning, meeting the objective of the task.

upvoted 3 times

PI_Team 3 years, 1 month ago

The correct answer to the question is (B) No. The given solution of applying a Quantiles binning mode with a PQuantile normalization does not meet the goal of normalizing values to produce an output column into bins for predicting a target column. While Quantiles binning and PQuantile normalization are useful techniques in their own right, they are not directly applicable for producing bins to predict a target column. To achieve the goal of normalizing values into bins to predict a target column, you would need to use appropriate techniques such as binning based on specific ranges or intervals, or other normalization methods tailored for your specific data and problem. SaMI need to correct my previous comment: The solution of using Quantiles binning mode with PQuantile normalization in Azure Machine Learning Studio is valid. Quantiles binning discretizes data based on percentile ranks, and the PQuantile normalization option within this mode normalizes values within a [0,1] range before sorting into quantiles. This is specific to preparing data for quantile binning and is not general data normalization. Therefore, this approach aligns with Azure Machine Learning Studio's capabilities for data preparation in quantile binning, meeting the objective of the task.

upvoted 1 times

PI_Team 3 years, 1 month ago

Selected Answer: B

The correct answer to the question is (B) No.

The given solution of applying a Quantiles binning mode with a PQuantile normalization does not meet the goal of normalizing values to produce an output column into bins for predicting a target column.

While Quantiles binning and PQuantile normalization are useful techniques in their own right, they are not directly applicable for producing bins to predict a target column.

To achieve the goal of normalizing values into bins to predict a target column, you would need to use appropriate techniques such as binning based on specific ranges or intervals, or other normalization methods tailored for your specific data and problem.

SaM

upvoted 1 times

striver 4 years, 2 months ago

Answer is A. Normalization makes the value fall in range [0, 1] and that's what PQuantile does too.

upvoted 1 times

ning 4 years, 3 months ago

Yes, this is the answer, see the link https://docs.microsoft.com/en-us/azure/machine-learning/component-reference/group-data-into-bins "Entropy MDL mode is defined in Studio (classic) and there's no corresponding open source package which can be leveraged to support in Designer yet." "If you select the Quantiles binning mode, use the Quantile normalization option to determine how values are normalized before sorting into quantiles. "

upvoted 3 times

TheCyanideLancer 4 years, 6 months ago

as per question, I believe, if target column is mentioned then ans mdl if feature column is mentioned then ans is PQuantile

upvoted 2 times

Tushazz 4 years, 7 months ago

Yes should be answer.

upvoted 1 times

chaudha4 5 years, 3 months ago

Entropy MDL is not available in designer. The answer applies only to studio(classic).

upvoted 7 times

Gonza967 6 years, 7 months ago

Answer is B

upvoted 2 times

HOTSPOT -

You are evaluating a Python NumPy array that contains six data points defined as follows: data = [10, 20, 30, 40, 50, 60]

You must generate the following output by using the k-fold algorithm implantation in the Python Scikit-learn machine learning library: train: [10 40 50 60], test: [20 30] train: [20 30 40 60], test: [10 50] train: [10 20 30 50], test: [40 60]

You need to implement a cross-validation to generate the output.

How should you complete the code segment? To answer, select the appropriate code segment in the dialog box in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

```
from numpy import array
from sklearn.model_selection import [ K-Means | k-fold | CrossValidation | ModelSelection ]
data – array([10, 20, 30, 40, 50, 60])
kfold – Kfold(n_splits- [ 1 | 2 | 3 | 6 ] ,shuffle – True, random_state-1)
for train, test in kFold, split( [ data | k-fold | array | train, test ] ):
print('train: %s, test: %5' % (data[train], data[test])
```

Suggested Answer:

Answer Area

```
from numpy import array
from sklearn.model_selection import [ k-fold ]
data = array([10, 20, 30, 40, 50, 60])
kfold = Kfold(n_splits=[ 3 ],shuffle = True, random_state=1)
for train, test in kFold, split( [ data ] ):
print('train: %s, test: %5' % (data[train], data[test])
```

(K-Means / k-fold / CrossValidation / ModelSelection; 1 / 2 / 3 / 6; data / k-fold / array / train, test)

Answer Area

```
from numpy import array
from sklearn.model_selection import [ k-fold ]
data = array([10, 20, 30, 40, 50, 60])
kfold = Kfold(n_splits=[ 3 ],shuffle = True, random_state=1)
for train, test in kFold, split( [ data ] ):
print('train: %s, test: %5' % (data[train], data[test])
```

(K-Means / k-fold / CrossValidation / ModelSelection; 1 / 2 / 3 / 6; data / k-fold / array / train, test)

Box 1: k-fold -

Box 2: 3 - K-Folds cross-validator provides train/test indices to split data in train/test sets. Split dataset into k consecutive folds (without shuffling by default). The parameter n_splits (int, default=3) is the number of folds. Must be at least 2.

Box 3: data - Example: Example: >>> >>> from sklearn.model_selection import KFold >>> X = np.array([[1, 2], [3, 4], [1, 2], [3, 4]]) >>> y = np.array([1, 2, 3, 4]) >>> kf = KFold(n_splits=2) >>> kf.get_n_splits(X) >>> print(kf) KFold(n_splits=2, random_state=None, shuffle=False) >>> for train_index, test_index in kf.split(X): ... print("TRAIN:", train_index, "TEST:", test_index) ... X_train, X_test = X[train_index], X[test_index] ... y_train, y_test = y[train_index], y[test_index] TRAIN: [2 3] TEST: [0 1] TRAIN: [0 1] TEST: [2 3] Reference: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html

podval Highly Voted 5 years, 7 months ago

Proper syntax: from sklearn.model_selection import KFold

upvoted 23 times

podval Highly Voted 5 years, 7 months ago

Proper syntax: from sklearn.model_selection import KFoldIf the actual question is written with 'k-fold' instead of 'Kfold', that's just stupid.

upvoted 23 times

David_Tadeu 3 years, 9 months ago

If the actual question is written with 'k-fold' instead of 'Kfold', that's just stupid.

upvoted 4 times

ljljljlj 4 years, 7 months ago

On exam 2021/7/10

upvoted 5 times

Matt2000 Most Recent 2 years ago

from sklearn.model_selection import KFold from numpy import array import numpy as np data = array([10,20,30,40,50,60]) k_fold = KFold(n_splits=3, shuffle=True,random_state=1) for train, test in k_fold, np.split(data): print(f'train: {train}, test: {test}')

upvoted 2 times

Matt2000 2 years ago

"-" shoud be read as "="

upvoted 1 times

Hisayuki 2 years, 3 months ago

You're gonna create three set of Train and Test dataset with Shuffling. So, the n_splits should be 3 in kfold. - train: [10 40 50 60], test: [20 30] - train: [20 30 40 60], test: [10 50] - train: [10 20 30 50], test: [40 60]

upvoted 3 times

ning 3 years, 9 months ago

Might be a typo, but overall is correct

upvoted 2 times

Question #46 *Topic 2*

You are with a time series dataset in Azure Machine Learning Studio.
You need to split your dataset into training and testing subsets by using the Split Data module.
Which splitting mode should you use?

A. Recommender Split

B. Regular Expression Split

C. Relative Expression Split

D. Split Rows with the Randomized split parameter set to true

Suggested Answer: *D*
Split Rows: Use this option if you just want to divide the data into two parts. You can specify the percentage of data to put in each split, but by default, the data is divided 50-50. Incorrect Answers: B: Regular Expression Split: Choose this option when you want to divide your dataset by testing a single column for a value. C: Relative Expression Split: Use this option whenever you want to apply a condition to a number column.
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data

Community vote distribution


jameswoo Highly Voted 5 years, 9 months ago
i think it is C. time-series data means you should split the data by date, otherwise you may have information leaking.
upvoted 58 times

jameswoo Highly Voted 5 years, 9 months ago
i think it is C. time-series data means you should split the data by date, otherwise you may have information leaking.Nope, my bad... definitely C. Relative Expression split cause it is a time series dataset.I think the Answer is correct. Since you just want to split into test and training Data. You are NOT trying to SPLIT into CATEGORIES. or dates
upvoted 58 times

human_ai 4 years, 6 months ago
Nope, my bad... definitely C. Relative Expression split cause it is a time series dataset.
upvoted 5 times

human_ai 4 years, 6 months ago
I think the Answer is correct. Since you just want to split into test and training Data. You are NOT trying to SPLIT into CATEGORIES. or dates
upvoted 7 times

saegeb2000 5 years, 7 months ago
This should be C: Relative Expression Split: Use this option whenever you want to apply a condition to a number column. The number can be a date/time field, a column that contains age or dollar amounts, or even a percentage. For example, you might want to divide your dataset based on the cost of the items, group people by age ranges, or separate data by a calendar date. https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/split-data
upvoted 29 times

84b1989 Most Recent 1 year, 7 months ago
Explanation: When working with a time series dataset in Azure Machine Learning Studio, the Split Rows mode with the Randomized split parameter set to true is the most appropriate choice for splitting the dataset into training and testing subsets. Here's why: Time Series Data Considerations: Time series data has a temporal order, and splitting it randomly ensures that the training and testing subsets are representative of the entire dataset without breaking the time sequence. This helps in maintaining the integrity of the data for model evaluation. Split Rows Mode: The Split Rows mode allows you to specify a fraction of the dataset to be used for training and testing. For example, you can allocate 70% of the data for training and 30% for testing. Randomized Split Parameter: Setting the Randomized split parameter to true ensures that the data is shuffled before splitting, which is crucial for time series data to avoid bias and ensure that the model generalizes well.
upvoted 2 times

Lion007 2 years, 1 month ago
WRONG. The Correct answer is: C The correct method for splitting a time series dataset should consider the sequential nature of the data. The options available in the Split Data component in Azure ML are: 1. Split Rows: This mode is used to simply divide the data into two parts. This mode is generally used when the sequence of data is not a concern. 2. Regular Expression Split: This mode is for dividing the dataset based on a pattern in a

text field such as analyzing sentiment. 3. Relative Expression Split: This mode applies to conditions on a number column, which could include date/time fields. For time series data, where the sequence and continuity of data points are important, neither randomization (as in Split Rows with randomization) nor pattern-based splits (as in Regular Expression Split) are appropriate. Instead, the Relative Expression Split, which can handle conditions on date/time fields, is suited for time series data, allowing the dataset to be divided without disrupting the sequence. Therefore, the correct answer should be C. Relative Expression Split.

upvoted 2 times

NullVoider_0 2 years, 2 months ago

When splitting a time series dataset in Azure Machine Learning Studio, you should use the "Split Rows" option with the "Randomized split" parameter set to false to ensure that the temporal order of the data is preserved. This approach is crucial for maintaining the integrity of time series data in training and testing subsets.

upvoted 3 times

dporwal04 2 years, 2 months ago

use any tool like search, bard, chatgpt or any other tool but ans is C

upvoted 1 times

ymj_000 2 years, 3 months ago

I think the answer is D because Randomized split is preferred option when you're creating training and test datasets. See https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/split-data?view=azureml-api-2. This is very sneaky by mentioning this is a time series data which makes me think the answer should be Relative Expression Split.

upvoted 2 times

PI_Team 2 years, 7 months ago

For splitting a time series dataset in Azure Machine Learning Studio, the appropriate splitting mode to use is the "Relative Expression Split" (C). The Relative Expression Split mode allows you to split the dataset based on conditions applied to a number column. This number column can be a date/time field, age, dollar amounts, percentages, or any other numerical value. It provides flexibility in defining the splitting criteria based on these numeric conditions. In the context of a time series dataset, you can use the Relative Expression Split mode to split the dataset based on conditions related to the time component, such as dividing data by calendar date, time periods, or specific ranges of dates.

upvoted 2 times

RamundiGR 3 years ago

it clearly C!! you can check on https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/split-data

upvoted 1 times

RamundiGR 3 years ago

why the moderator does not bother to correct those answers?

upvoted 2 times

NachoPrendes 3 years ago

I think D is the correct one to choose random dates belonging to all years in two datasets

upvoted 1 times

Edriv 3 years, 2 months ago

why not B?

upvoted 1 times

KIshor1212 3 years, 2 months ago

Calendar year A common scenario is to divide a dataset by years. The following expression selects all rows where the values in the column Year are greater than 2010.

upvoted 1 times

PremPatrick 3 years, 3 months ago

C should be correct

upvoted 2 times

fvil 3 years, 3 months ago

Question about Split Data module and differences between Regular Expression and Relative Expression appears on exam 07/11/2022

upvoted 1 times

azurelearner666 3 years, 10 months ago

Selected Answer: C Without any shadow of a doubt ;) In machine learning, train/test split splits the data randomly, as there's no dependence from one observation to the other. That's not the case with time series data. Here, you'll want to use values at the rear of the dataset for testing and everything else for training. Example: Select first 10 years for training and 2 years for testing. https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/split-data (check the Relative Expression Split section)

upvoted 1 times

sam844 3 years, 10 months ago

C is the correct choice. It is time series data so it has to be split by date which is only Relative Expression

upvoted 1 times

HOTSPOT -

You are preparing to build a deep learning convolutional neural network model for image classification. You create a script to train the model using CUDA devices.

You must submit an experiment that runs this script in the Azure Machine Learning workspace.

The following compute resources are available:

✑ a Microsoft Surface device on which Microsoft Office has been installed. Corporate IT policies prevent the installation of additional software

✑ a Compute Instance named ds-workstation in the workspace with 2 CPUs and 8 GB of memory

✑ an Azure Machine Learning compute target named cpu-cluster with eight CPU-based nodes

✑ an Azure Machine Learning compute target named gpu-cluster with four CPU and GPU-based nodes

You need to specify the compute resources to be used for running the code to submit the experiment, and for running the script in order to minimize model training time.

Which resources should the data scientist use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Resource type	Option
Run code to submit the experiment	the Microsoft Surface device the ds-workstation compute instance the cpu-cluster compute target the gpu-cluster compute target
Run the training script	the ds-workstation compute instance the cpu-cluster compute target the gpu-cluster compute target the Microsoft Surface device

Suggested Answer:

Answer Area

Resource type	Option
Run code to submit the experiment	the ds-workstation compute instance
Run the training script	the gpu-cluster compute target

Answer Area

Resource type	Option
Run code to submit the experiment	the ds-workstation compute instance
Run the training script	the gpu-cluster compute target

Box 1: the ds-workstation compute instance A workstation notebook instance is good enough to run experiments. Box 2: the gpu-cluster compute target Just as GPUs revolutionized deep learning through unprecedented training and inferencing performance, RAPIDS enables traditional machine learning practitioners to unlock game-changing performance with GPUs. With RAPIDS on Azure Machine Learning service, users can accelerate the entire machine learning pipeline, including data processing, training and inferencing, with GPUs from the NC_v3, NC_v2, ND or ND_v2 families. Users can unlock performance gains of more than 20X (with 4 GPUs), slashing training times from hours to minutes and dramatically reducing time-to-insight. Reference: https://azure.microsoft.com/sv-se/blog/azure-machine-learning-service-now-supports-nvidia-s-rapids/

dev2dev Highly Voted 2 years, 11 months ago

answers are correct, for submitting job we can use surface book but due to restrictions we cant install sdk/pythonThanks for your sharing

upvoted 32 times

dev2dev Highly Voted 2 years, 11 months ago

answers are correct, for submitting job we can use surface book but due to restrictions we cant install sdk/python

upvoted 32 times

aoddy 2 years, 10 months ago

Thanks for your sharing

upvoted 3 times

ning Most Recent 1 year, 9 months ago

The answer should be correct ... But the question is poorly written ... I believe it means, you can write python codes on the ds computer instance to submit experiment with script steps against gpu computer instance as target

upvoted 3 times

hargur 2 years, 4 months ago

on 19Oct2021

upvoted 3 times

kisskeo 2 years, 4 months ago

On Exam 01 Oct 2021

upvoted 1 times

snsnsnsn 2 years, 5 months ago

got similar question on 2/9/21

upvoted 1 times

VJPrakash 2 years, 6 months ago

on exam in August 2021

upvoted 3 times

Question #48 *Topic 2*

You create an Azure Machine Learning workspace. You are preparing a local Python environment on a laptop computer. You want to use the laptop to connect to the workspace and run experiments.
You create the following config.json file.
{
"workspace_name" : "ml-workspace"
}
You must use the Azure Machine Learning SDK to interact with data and experiments in the workspace.
You need to configure the config.json file to connect to the workspace from the Python environment.
Which two additional parameters must you add to the config.json file in order to connect to the workspace? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. login

B. resource_group

C. subscription_id

D. key

E. region

Suggested Answer: *B, C*
To use the same workspace in multiple environments, create a JSON configuration file. The configuration file saves your subscription (subscription_id), resource (resource_group), and workspace name so that it can be easily loaded. The following sample shows how to create a workspace. from azureml.core import Workspace ws = Workspace.create(name='myworkspace', subscription_id=' ', resource_group='myresourcegroup', create_resource_group=True, location='eastus2') Reference: https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace

Community vote distribution
BC (100%)

udiramana Most Recent 1 year, 11 months ago
given answer is correct
upvoted 4 times

fvil 2 years, 9 months ago
Appeared on exam 07/11/2022
upvoted 2 times

kkkk_jjjj 3 years, 5 months ago
on exam 18/03/2022
upvoted 1 times

synapse 3 years, 5 months ago
given answer is correct
upvoted 3 times

hargur 3 years, 10 months ago
on 19Oct2021
upvoted 2 times

kisskeo 3 years, 10 months ago
On Exam 01 Oct 2021
upvoted 2 times

mthombenindhl84 3 years, 11 months ago
on exam 11/9/2021
upvoted 2 times

dushmantha 3 years, 11 months ago
On exam 2021/08/31

upvoted 2 times

HOTSPOT -

You are performing a classification task in Azure Machine Learning Studio.

You must prepare balanced testing and training samples based on a provided data set.

You need to split the data with a 0.75:0.25 ratio.

Which value should you use for each parameter? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Parameter	Value
Splitting mode	Split rows Recommender Split Regular Expression Split Relative Expression Split
Fraction of rows in the first output dataset	0.75 0.25 0.5 1
Randomized split	True False
Stratified split	True False

Suggested Answer:

Answer Area

Parameter	Value
Splitting mode	Split rows
Fraction of rows in the first output dataset	0.75
Randomized split	True
Stratified split	False

Answer Area

Parameter	Value
Splitting mode	Split rows
Fraction of rows in the first output dataset	0.75
Randomized split	True
Stratified split	False

Box 1: Split rows - Use the Split Rows option if you just want to divide the data into two parts. You can specify the percentage of data to put in each split, but by default, the data is divided 50-50. You can also randomize the selection of rows in each group, and use stratified sampling. In stratified sampling, you must select a single column of data for which you want values to be apportioned equally among the two result datasets.

Box 2: 0.75 - If you specify a number as a percentage, or if you use a string that contains the "%" character, the value is interpreted as a percentage. All percentage values must be within the range (0, 100), not including the values 0 and 100.

Box 3: Yes - To ensure splits are balanced.

Box 4: No - If you use the option for a stratified split, the output datasets can be further divided by subgroups, by selecting a strata column.
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data

Yong2020 Highly Voted 5 years, 3 months ago

Stratified split should be true in order to be balancedI agreeTotally agreeI agree https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/split-dataWould tend to go with this answer too, cause it doesn't make sense that randomize split could ensure testing and training samples are balanced, since it perform "randomize selection", saying there might be possibility that some targeted values might happen to have more/less than the other.

upvoted 66 times

Yong2020 Highly Voted 5 years, 3 months ago

Stratified split should be true in order to be balanced

upvoted 66 times

kty 4 years, 4 months ago

I agree

upvoted 3 times

dija123 3 years, 8 months ago

Totally agree

upvoted 2 times

SnowCheetah 4 years, 1 month ago

I agree https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/split-data

upvoted 3 times

JUEI 5 years ago

Would tend to go with this answer too, cause it doesn't make sense that randomize split could ensure testing and training samples are balanced, since it perform "randomize selection", saying there might be possibility that some targeted values might happen to have more/less than the other.

upvoted 6 times

Andrexx 4 years, 9 months ago

In my opinion, stratified split should be true. Based on this: "Stratified split: Set this option to True to ensure that the two output datasets contain a representative sample of the values in the strata column or stratification key column. With stratified sampling, the data is divided such that each output dataset gets roughly the same percentage of each target value. For example, you might want to ensure that your training and testing sets are roughly balanced with regard to the outcome, or with regard to some other column such as gender." https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data-using-split-rows

upvoted 13 times

smaha_18 Most Recent 8 months ago

Stratified split = True Reason: With stratified sampling, the data is divided such that each output dataset gets roughly the same percentage of each target value. For example, you might want to ensure that your training and testing sets are roughly balanced with regard to the outcome Reference: https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/split-data-using-split-rows

upvoted 1 times

PI_Team 2 years, 1 month ago

When performing data splitting, especially in scenarios where there is an imbalance in the distribution of certain categories or groups within a column, using a strata column can be beneficial. By grouping the rows based on the strata column, it helps ensure that each subset (e.g., training or testing) maintains a similar representation of the different categories or groups present in the strata column. so Stratified split - TRUE SaM

upvoted 5 times

MohammadKhubeb 3 years, 6 months ago

Stratified split should be FALSE, because it is not the IMBALANCE DATASET. Please refer this: Chawla, Nitesh V., et al. "SMOTE: synthetic minority over-sampling technique." Journal of Artificial Intelligence Research 16 (2002): 321-357.

upvoted 3 times

azure1000 4 years ago

In a classification setting, it is often chosen to ensure that the train and test sets have approximately the same percentage of samples of each target class as the complete set. As a result, if the data set has a large amount of each class, stratified sampling is pretty much the same as random sampling. But if one class isn't much represented in the data set, which may be the case in your dataset since you plan to oversample the minority class, then stratified sampling may yield a different target class distribution in the train and test sets than what random sampling may yield.

upvoted 1 times

Lucario95 4 years, 2 months ago

I think Stratified split should be set to True for balancing the 2 subsets, no info about Random Splitting True/False though...
upvoted 3 times

Lucario95 4 years, 2 months ago
I think Stratified split should be set to True for balancing the 2 subsets, no info about Random Splitting True/False though...Normally you don't need to stratify if the dataset is balanced. It's the same in sklearn
upvoted 3 times

Padilha 2 years, 7 months ago
Normally you don't need to stratify if the dataset is balanced. It's the same in sklearn
upvoted 1 times

rishi_ram 4 years, 2 months ago
Based on below requirement I would say Stratified Split should be false Additional requirements for stratified sampling: The strata column can contain only nominal or categorical data. If the column contains continuous numeric data, an error message is raised. A column with too many unique values is not a good candidate for stratification. You might try collapsing some categories or grouping values beforehand.
upvoted 1 times

Tusharsp 4 years, 3 months ago
Contrary to all comments here, stratified split should be False. For it to be set to true, we need to select the column for which data needs to be stratified. There is no mention of any column in the question. In this scenario just saying select it to True does not make sense. Also for people following documentation, here is the extract. Stratified split: Set this option to True to ensure that the two output datasets contain a representative sample of the values in the strata column or stratification key column. With stratified sampling, the data is divided such that each output dataset gets roughly the same percentage of each target value. For example, you might want to ensure that your training and testing sets are roughly balanced with regard to the outcome or to some other column (such as gender).
upvoted 10 times

poons 4 years, 4 months ago
Since, the dataset is balanced, Stratified split = False might work.
upvoted 3 times

poons 4 years, 4 months ago
Since, the dataset is balanced, Stratified split = False might work.Didnt say it is balanced... said you have to balance it
upvoted 3 times

PremPatrick 2 years, 9 months ago
Didnt say it is balanced... said you have to balance it
upvoted 1 times

saurabhk1 4 years, 5 months ago
I too think Stratified should be set to True.
upvoted 2 times

Paa_Kwesi 4 years, 8 months ago
stratified split should be set to True With stratified sampling, the data is divided such that each output dataset gets roughly the same percentage of each target value. For example, you might want to ensure that your training and testing sets are roughly balanced with regard to the outcome, or with regard to some other column such as gender. https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data-using-split-rows
upvoted 6 times

Paa_Kwesi 4 years, 8 months ago
stratified split should be set to True With stratified sampling, the data is divided such that each output dataset gets roughly the same percentage of each target value. For example, you might want to ensure that your training and testing sets are roughly balanced with regard to the outcome, or with regard to some other column such as gender. https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data-using-split-rowsYou are absolutely right. For "balanced" set -> "Stratified" must be true
upvoted 6 times

DanielGP 4 years, 6 months ago
You are absolutely right. For "balanced" set -> "Stratified" must be true
upvoted 2 times

Pucha 4 years, 9 months ago
Stratified should be true
upvoted 2 times

podval 5 years, 1 month ago
Stratified means UNbalanced as it keeps the ratio between classes. That is why it needs to be set to False.Correct. Stratified should be False.where did you get the meaning of stratified is unbalanced? strata means layers/groups. in ML we do stratify data to make sure the data balance is

maintained.

upvoted 4 times

podval 5 years, 1 month ago

Stratified means UNbalanced as it keeps the ratio between classes. That is why it needs to be set to False.

upvoted 4 times

Gitty 5 years ago

Correct. Stratified should be False.

upvoted 2 times

dev2dev 4 years, 5 months ago

where did you get the meaning of stratified is unbalanced? strata means layers/groups. in ML we do stratify data to make sure the data balance is maintained.

upvoted 3 times

kath3624 5 years, 1 month ago

because of balanced testing and training samples, Statisfied should be set to True. https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data-using-split-rows

upvoted 8 times

jsnels86 5 years, 2 months ago

Stratified split is correct here

upvoted 4 times

Question #50 *Topic 2*

You create an Azure Machine Learning compute resource to train models. The compute resource is configured as follows:

✑ Minimum nodes: 2

✑ Maximum nodes: 4

You must decrease the minimum number of nodes and increase the maximum number of nodes to the following values:

✑ Minimum nodes: 0

✑ Maximum nodes: 8

You need to reconfigure the compute resource.

What are three possible ways to achieve this goal? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

A. Use the Azure Machine Learning studio.

B. Run the update method of the AmlCompute class in the Python SDK.

C. Use the Azure portal.

D. Use the Azure Machine Learning designer.

E. Run the refresh_state() method of the BatchCompute class in the Python SDK.

Suggested Answer: *A, B, C*

A: You can manage assets and resources in the Azure Machine Learning studio. B: The update(min_nodes=None, max_nodes=None, idle_seconds_before_scaledown=None) of the AmlCompute class updates the ScaleSettings for this AmlCompute target. C: To change the nodes in the cluster, use the UI for your cluster in the Azure portal. Reference: https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute(class)

Community vote distribution

ABC (100%)

Haet Highly Voted 4 years, 9 months ago

Answer is correct A,B,C. I have verified itCan you elaborate on "C". There is no access to the cluster from Azure portal. Where do you see it ?Azure portal is unlikely to do this and refresh_state() method of the BatchCompute also seems to be incorrect. In designer there is an option to create a new CC with the specifications we require, but not sure if it can be considered as reconfiguration as we're simply creating another resource.

upvoted 23 times

Haet Highly Voted 4 years, 9 months ago

Answer is correct A,B,C. I have verified it

upvoted 23 times

chaudha4 4 years, 9 months ago

Can you elaborate on "C". There is no access to the cluster from Azure portal. Where do you see it ?Azure portal is unlikely to do this and refresh_state() method of the BatchCompute also seems to be incorrect. In designer there is an option to create a new CC with the specifications we require, but not sure if it can be considered as reconfiguration as we're simply creating another resource.

upvoted 3 times

chaudha4 4 years, 9 months ago

Can you elaborate on "C". There is no access to the cluster from Azure portal. Where do you see it ?

upvoted 3 times

SaudMeethal 4 years, 9 months ago

Azure portal is unlikely to do this and refresh_state() method of the BatchCompute also seems to be incorrect. In designer there is an option to create a new CC with the specifications we require, but not sure if it can be considered as reconfiguration as we're simply creating another resource.

upvoted 4 times

levm39 4 years, 8 months ago

Answer is A,B C was possible in the past, they removed the option to manage compute instance for ML from the portal. E is incorrect

upvoted 17 times

smaha_18 Most Recent 8 months ago

Azure ML Studio → ML-focused management (Already inside Azure Portal: Azure ML studio → Manage → Compute → Compute clusters → Edit cluster settings) Azure Portal → Resource-level management (Azure Portal→ Azure Machine Learning → Launch Azure ML studio → Manage →

Compute → Compute clusters → Edit cluster settings) Python SDK → Programmatic management

upvoted 1 times

NullVoider_0 2 years ago

On exam 12-02-2024.

upvoted 1 times

Awooga 2 years ago

On exam 2024-02-06

upvoted 1 times

phdykd 3 years ago

A. Use the Azure Machine Learning studio. C. Use the Azure portal. D. Use the Azure Machine Learning designer.

upvoted 2 times

therealola 3 years, 8 months ago

On exam 18-06-22

upvoted 1 times

ning 3 years, 9 months ago

Only know A and B ...

upvoted 2 times

JTWang 3 years, 10 months ago

on exam 04/22/2022

upvoted 2 times

synapse 3 years, 11 months ago

A B C. Given answer is correct

upvoted 1 times

piyu18 4 years, 1 month ago

I think its correct, as we can access ml compute from cli and we can connect with cli from azure portal

upvoted 2 times

VJPrakash 4 years, 6 months ago

on exam in August 2021

upvoted 1 times

slash_nyk 4 years, 7 months ago

How you can do if from Machine Learning Studio ?

upvoted 2 times

scipio 4 years, 9 months ago

Cannot be E. BatchCompute is an attach resource, and refresh_state() is "This method updates the properties based on the current state of the corresponding cloud object. This is primarily used for manual polling of compute state." Just a refresh. I agree though that C is strange, except to use the portal to access the ML Studio. https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.batchcompute?view=azure-ml-py#refresh-state--

upvoted 4 times

chaudha4 4 years, 10 months ago

The correct answer is A, B and E. You cannot do it from portal.

upvoted 7 times

dev2dev 4 years, 11 months ago

I dont see from where we can resize the culuster size from portal except that from the ml workspace we click on studio link and change from there

upvoted 5 times

dev2dev 4 years, 11 months ago

I dont see from where we can resize the culuster size from portal except that from the workspace we click on studio link

upvoted 2 times

Question #51 *Topic 2*

HOTSPOT -

You have a dataset that contains 2,000 rows. You are building a machine learning classification model by using Azure Learning Studio. You add a Partition and

Sample module to the experiment.

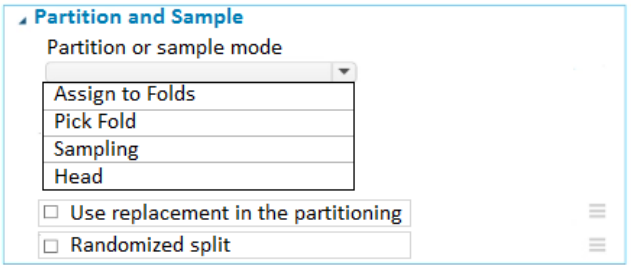
You need to configure the module. You must meet the following requirements:

- ✑ Divide the data into subsets
- ✑ Assign the rows into folds using a round-robin method
- ✑ Allow rows in the dataset to be reused

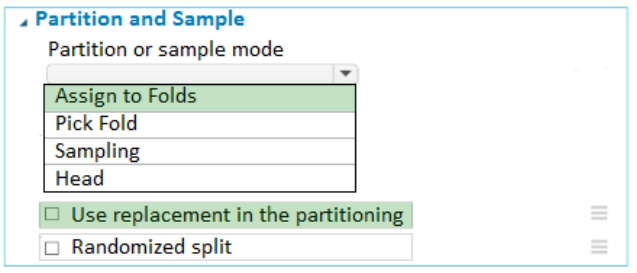
How should you configure the module? To answer, select the appropriate options in the dialog box in the answer area.

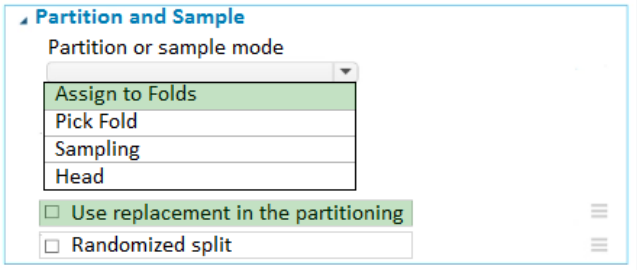
NOTE: Each correct selection is worth one point.

Hot Area:



Suggested Answer:





Use the Split data into partitions option when you want to divide the dataset into subsets of the data. This option is also useful when you want to create a custom number of folds for cross-validation, or to split rows into several groups. 1. Add the Partition and Sample module to your experiment in Studio (classic), and connect the dataset. 2. For Partition or sample mode, select Assign to Folds. 3. Use replacement in the partitioning: Select this option if you want the sampled row to be put back into the pool of rows for potential reuse. As a result, the same row might be assigned to several folds. 4. If you do not use replacement (the default option), the sampled row is not put back into the pool of rows for potential reuse. As a result, each row can be assigned to only one fold. 5. Randomized split: Select this option if you want rows to be randomly assigned to folds. If you do not select this option, rows are assigned to folds using the round-robin method. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample

phdykd 2 years ago

A. Assign to Folds: To divide the data into subsets and assign the rows into folds using a round-robin method. E. Use replacement in the partitioning: To allow the same row to be assigned to multiple folds.

upvoted 3 times

phdykd 2 years ago

A. Assign to Folds: To divide the data into subsets and assign the rows into folds using a round-robin method. C. Sampling: To allow rows in the dataset to be reused. E. Use replacement in the partitioning: To allow the same row to be assigned to multiple folds.

upvoted 3 times

phdykd 2 years ago

A. Assign to Folds: To divide the data into subsets and assign the rows into folds using a round-robin method. C. Sampling: To allow rows in the dataset to be reused. E. Use replacement in the partitioning: To allow the same row to be assigned to multiple folds.you are wrong given answer is correct, i have checked the module in designer

upvoted 3 times

ajay0011 Most Recent 1 year, 10 months ago

you are wrong given answer is correct, i have checked the module in designer

upvoted 2 times

synapse 2 years, 11 months ago

The given answer is correct.

upvoted 4 times

MohammadKhubeb 3 years ago

It is Randomized split because of Round Robin.You mean Randomized_split is unchecked because of Round Robin

upvoted 1 times

MohammadKhubeb 3 years ago

It is Randomized split because of Round Robin.

upvoted 1 times

David_Tadeu 2 years, 10 months ago

You mean Randomized_split is unchecked because of Round Robin

upvoted 6 times

Narendra05 3 years, 7 months ago

The Data to divide into fold is already mentioned that option is correct .1. Assign to Folds 2. use Replacement in the partition - Checked 3. Randomized split- uncheckedRandomized split: Select this option if you want rows to be randomly assigned to folds. If you do not select this option, rows are assigned to folds using the round-robin method.

upvoted 2 times

Narendra05 3 years, 7 months ago

The Data to divide into fold is already mentioned that option is correct .

upvoted 2 times

parwa Highly Voted 3 years, 7 months ago

1. Assign to Folds 2. use Replacement in the partition - Checked 3. Randomized split- unchecked

upvoted 9 times

parwa 3 years, 7 months ago

1. Assign to Folds 2. use Replacement in the partition - Checked 3. Randomized split- uncheckedRandomized split: Select this option if you want rows to be randomly assigned to folds. If you do not select this option, rows are assigned to folds using the round-robin method.

upvoted 9 times

YipingRuan Highly Voted 3 years, 6 months ago

Randomized split: Select this option if you want rows to be randomly assigned to folds. If you do not select this option, rows are assigned to folds using the round-robin method.

upvoted 10 times

Question #52 *Topic 2*

You create a new Azure subscription. No resources are provisioned in the subscription.
You need to create an Azure Machine Learning workspace.
What are three possible ways to achieve this goal? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. Run Python code that uses the Azure ML SDK library and calls the Workspace.create method with name, subscription_id, and resource_group parameters.

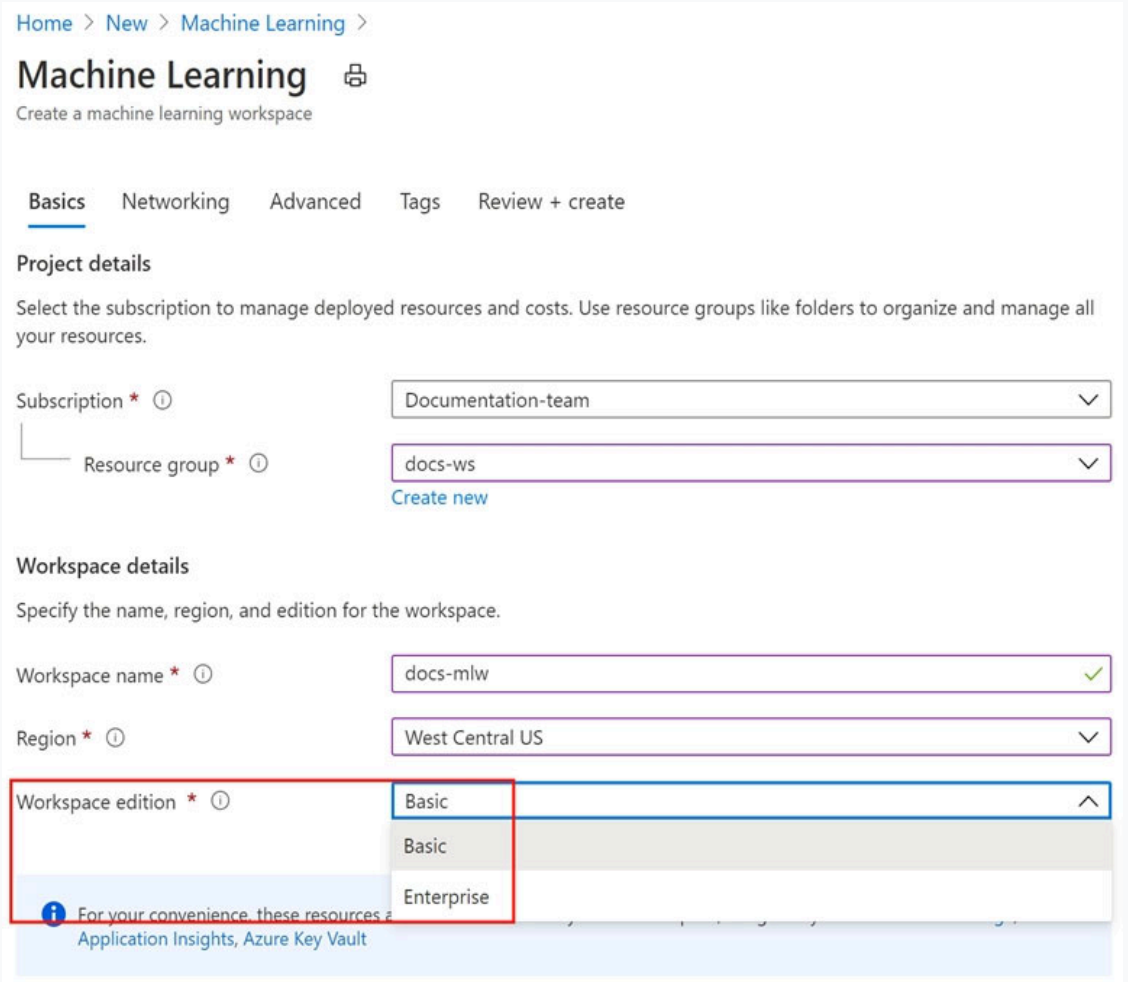
B. Navigate to Azure Machine Learning studio and create a workspace.

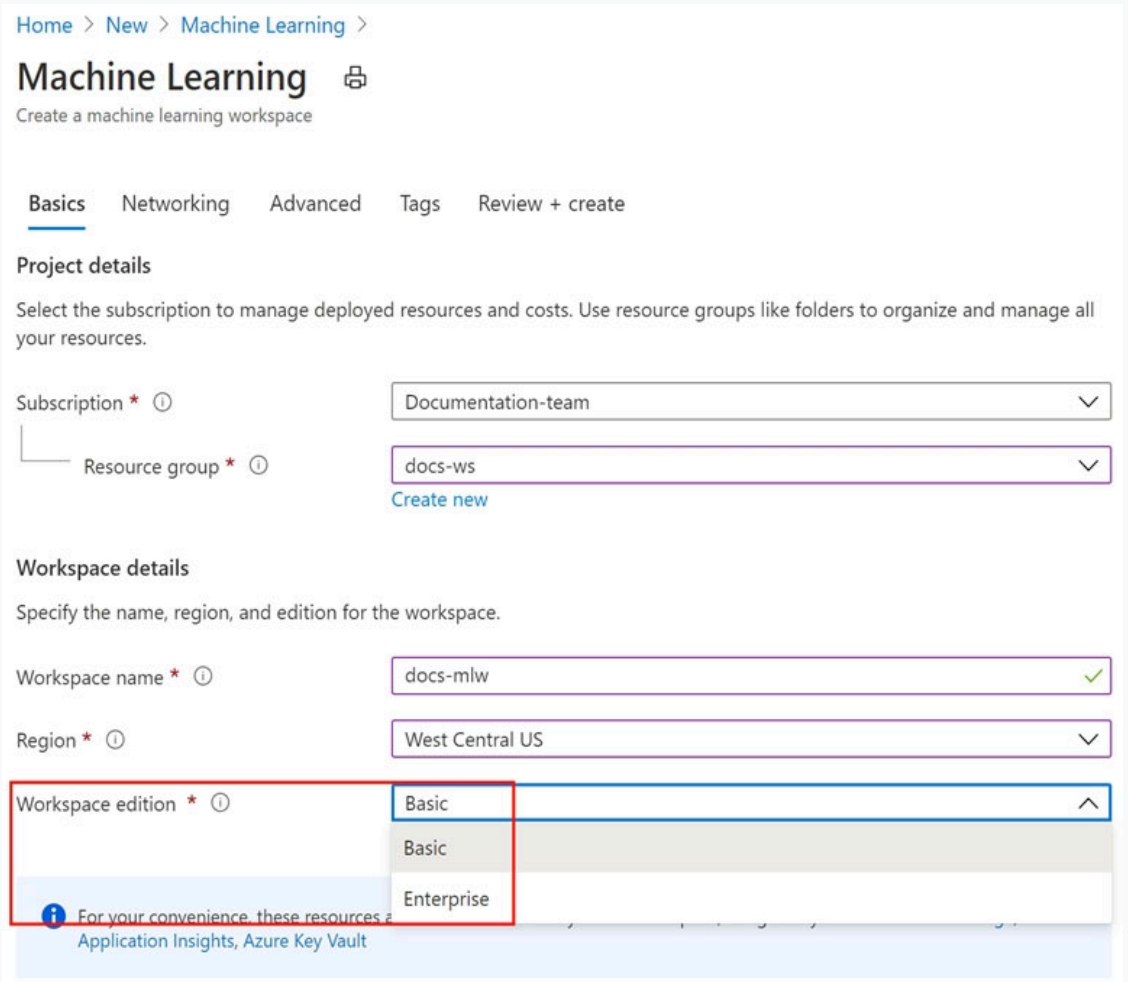
C. Use the Azure Command Line Interface (CLI) with the Azure Machine Learning extension to call the az group create function with --name and --location parameters, and then the az ml workspace create function, specifying ג€"w and ג€"g parameters for the workspace name and resource group.

D. Navigate to Azure Machine Learning studio and create a workspace.

E. Run Python code that uses the Azure ML SDK library and calls the Workspace.get method with name, subscription_id, and resource_group parameters.

Suggested Answer: *B, C, D*



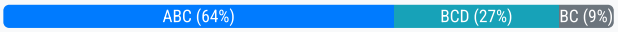


B: You can create a workspace in the Azure Machine Learning studio C: You can create a workspace for Azure Machine Learning with Azure CLI Install the machine learning extension. Create a resource group: az group create --name --location To create a new workspace where the services are automatically created, use the following command: az ml workspace create -w -g

D: You can create and manage Azure Machine Learning workspaces in the Azure portal. 1. Sign in to the Azure portal by using the credentials for your Azure subscription. 2. In the upper-left corner of Azure portal, select + Create a resource. 3. Use the search bar to find Machine Learning. 4. Select Machine Learning. 5. In the Machine Learning pane, select Create to begin.

Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-workspace-template https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace-cli https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace

Community vote distribution



ljljljlj Highly Voted 4 years, 7 months ago

On exam 2021/7/10 My answers: ABC

Correct options: A . Run Python code that uses the Azure ML SDK library and calls the Workspace.create method with name, subscription_id, resource_group, and location parameters. B . Use an Azure Resource Management template that includes a Microsoft.MachineLearningServices/ workspaces resource and its dependencies. C . Use the Azure Command Line Interface (CLI) with the Azure Machine Learning extension to call the az group create function with –name and –location parameters, and then the az ml workspace create function, specifying Cw and Cg parameters for the workspace name and resource group. D . Navigate to Azure Machine Learning studio and create a workspace. E . Run Python code that uses the Azure ML SDK library and calls the Workspace.get method with name, subscription_id, and resource_group parameters.

upvoted 39 times

ljljljlj Highly Voted 4 years, 7 months ago

On exam 2021/7/10 My answers: ABC Correct options: A . Run Python code that uses the Azure ML SDK library and calls the Workspace.create method with name, subscription_id, resource_group, and location parameters. B . Use an Azure Resource Management template that includes a Microsoft.MachineLearningServices/ workspaces resource and its dependencies. C . Use the Azure Command Line Interface (CLI) with the Azure Machine Learning extension to call the az group create function with –name and –location parameters, and then the az ml workspace create function, specifying Cw and Cg parameters for the workspace name and resource group. D . Navigate to Azure Machine Learning studio and create a workspace. E . Run Python code that uses the Azure ML SDK library and calls the Workspace.get method with name, subscription_id, and resource_group parameters.if there are no resources provisioned in the subscription, you cannot directly use Python code that calls Workspace.create method with the resource_group parameter. The resource_group parameter requires an existing resource group to be specified. So we can't use the "A" option. My Answer:BCDMy Answer:BCD A is wrong at 2022/10 The subscription_id and resource_group is not needed. D is correct. You can go to http://ml.azure.com to create a new workspace! python code: ws_basic = Workspace(name=basic_workspace_name, location="eastus", display_name="Basic workspace-example", description="This example shows how to create a basic workspace", hbi_workspace=False, tags=dict(purpose="demo"),) ml_client.workspaces.begin_create(ws_basic)

upvoted 39 times

PI_Team 2 years, 7 months ago

if there are no resources provisioned in the subscription, you cannot directly use Python code that calls Workspace.create method with the resource_group parameter. The resource_group parameter requires an existing resource group to be specified. So we can't use the "A" option. My Answer:BCD

upvoted 1 times

JTWang 3 years, 4 months ago

My Answer:BCD A is wrong at 2022/10 The subscription_id and resource_group is not needed. D is correct. You can go to http://ml.azure.com to create a new workspace! python code: ws_basic = Workspace(name=basic_workspace_name, location="eastus", display_name="Basic workspace-example", description="This example shows how to create a basic workspace", hbi_workspace=False, tags=dict(purpose="demo"),) ml_client.workspaces.begin_create(ws_basic)

upvoted 7 times

allanm 4 years, 9 months ago

I'm confused - Why is B and D the same answer option?B or D are Azure Machine Learning ServiceAlso noticed A & E are the same answers as well. What's going on?Try to look at Workspace.____ A. Workspace.create E. Workspace.getThe D option should be : Navigate to the Azure portal and create a workspace.then answer is: CD

upvoted 14 times

allanm 4 years, 9 months ago

I'm confused - Why is B and D the same answer option?

upvoted 14 times

Edriv 3 years, 1 month ago

B or D are Azure Machine Learning Service

upvoted 1 times

allanm 4 years, 9 months ago

Also noticed A & E are the same answers as well. What's going on?Try to look at Workspace.____ A. Workspace.create E. Workspace.get

upvoted 8 times

allanm 4 years, 9 months ago

Also noticed A & E are the same answers as well. What's going on?

upvoted 8 times

Norasit 3 years ago

Try to look at Workspace.____ A. Workspace.create E. Workspace.get

upvoted 1 times

andre999 4 years, 8 months ago

The D option should be : Navigate to the Azure portal and create a workspace.then answer is: CD

upvoted 11 times

andre999 4 years, 8 months ago

The D option should be : Navigate to the Azure portal and create a workspace.

upvoted 11 times

ppchar 4 years, 7 months ago

then answer is: CD

upvoted 2 times

NullVoider_0 Most Recent 2 years, 2 months ago

Option A uses the SDK to programmatically create the workspace Option B leverages the visual interface to create the workspace Option C uses CLI commands to automatically create both resource group and workspace

upvoted 2 times

Nadine_nm 2 years, 5 months ago

Answers should be ABD : https://learn.microsoft.com/en-us/azure/machine-learning/concept-workspace?view=azureml-api-2 Workspace management task | Portal | Studio | Python SDK |Azure CLI | VS Code Create a workspace | ✓ | ✓ | ✓ | ✓ | ✓

upvoted 2 times

phdykd 2 years, 7 months ago

BCD. az group create --name --location ws_basic = ml_client.workspaces.begin_create(ws_basic).result() print(ws_basic)

upvoted 1 times

krishna1818 2 years, 8 months ago

C is a definite option, and B also seems right

upvoted 1 times

Dilesha 3 years ago

On Exam 17 Feb 2023Thanks Dilesha. If by any chance you come back again, do you remember the differences between option B and D?The D option should be : Navigate to the Azure portal and create a workspace. And it's not correct. Correct answers are :A B C

upvoted 6 times

Dilesha 3 years ago

On Exam 17 Feb 2023

upvoted 6 times

AzureJobsTillRetire 2 years, 12 months ago

Thanks Dilesha. If by any chance you come back again, do you remember the differences between option B and D?The D option should be : Navigate to the Azure portal and create a workspace. And it's not correct. Correct answers are :A B C

upvoted 3 times

AzureJobsTillRetire 2 years, 12 months ago

Thanks Dilesha. If by any chance you come back again, do you remember the differences between option B and D?

upvoted 3 times

hammamse 2 years, 10 months ago

The D option should be : Navigate to the Azure portal and create a workspace. And it's not correct. Correct answers are :A B C

upvoted 2 times

phdykd 3 years ago

A. Run Python code that uses the Azure ML SDK library and calls the Workspace.create method with name, subscription_id, and resource_group parameters. B. Navigate to Azure Machine Learning studio and create a workspace. C. Use the Azure Command Line Interface (CLI) with the Azure Machine Learning extension to call the az group create function with --name and --location parameters, and then the az ml workspace create function, specifying "w" and "g" parameters for the workspace name and resource group. Option D. Navigate to Azure Machine Learning studio and create a workspace. is a correct answer, but it is a duplicate of option B. So, the correct answers are A, B, and C

upvoted 3 times

AzureJobsTillRetire 3 years ago

I think, but not 100% sure that the code for creation of a workspcace is workspaces.begin_create and not Workspaces.create as described in option A. example: ml_client.workspaces.begin_create(ws) https://learn.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace?tabs=pythonSorry my bad. Workspace.create is correct. from azureml.core import Workspace ws = Workspace.create(name='myworkspace', subscription_id=' ', resource_group='myresourcegroup', create_resource_group=True, location='eastus2') https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace?view=azure-ml-py#azureml-core-workspace-workspace-createIt seems that ABCD are all correct.

upvoted 1 times

AzureJobsTillRetire 3 years ago

Selected Answer: BCD

I think, but not 100% sure that the code for creation of a workspcace is workspaces.begin_create and not Workspaces.create as described in option A. example: ml_client.workspaces.begin_create(ws) https://learn.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace?tabs=python

upvoted 1 times

AzureJobsTillRetire 3 years ago

Sorry my bad. Workspace.create is correct. from azureml.core import Workspace ws = Workspace.create(name='myworkspace', subscription_id=' ', resource_group='myresourcegroup', create_resource_group=True, location='eastus2') https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace?view=azure-ml-py#azureml-core-workspace-workspace-create

upvoted 3 times

AzureJobsTillRetire 3 years ago

It seems that ABCD are all correct.

upvoted 1 times

MansoorDataScientist 3 years ago

A cannot consider before creating a workspace, yes we can create through different methods of CICD but it's not considered here

upvoted 1 times

JTWang 3 years, 4 months ago

If D is create workspace by Azure Machine Learning Studio, it is true! https://ml.azure.com/ You can go to https://ml.azure.com/ to create a new workspace.

upvoted 2 times

gursimran_s 3 years, 8 months ago

Why not ACD?

upvoted 2 times

zb99 3 years, 9 months ago

Selected Answer: BCD

A is wrong: Cannot create a workspace this way because resource group does not yet exist.

upvoted 2 times

zb99 3 years, 9 months ago

A is wrong: Cannot create a workspace this way because resource group does not yet exist.You can create a Workspace this way, but you need to include this argument: create_resource_group=True. This is why option A is wrong, there is no previous resource group and it isn't created automatically on the Worspace.create method.`create_resource_group=True` by default according to the docs (https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace?view=azure-ml-py#azureml-core-workspace-workspace-create). So A so be a valid option too

upvoted 2 times

casiopa 3 years, 2 months ago

You can create a Workspace this way, but you need to include this argument: create_resource_group=True. This is why option A is wrong, there is no previous resource group and it isn't created automatically on the Worspace.create method.`create_resource_group=True` by default according to the docs (https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace?view=azure-ml-py#azureml-core-workspace-workspace-create). So A so be a valid option too

upvoted 1 times

casiopa 3 years, 2 months ago

You can create a Workspace this way, but you need to include this argument: create_resource_group=True. This is why option A is wrong, there is no previous resource group and it isn't created automatically on the Worspace.create method.

upvoted 1 times

dsaz 3 years, 2 months ago

`create_resource_group=True` by default according to the docs (https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace?view=azure-ml-py#azureml-core-workspace-workspace-create). So A so be a valid option too

upvoted 3 times

ning 3 years, 9 months ago

With the correct options, ABC no doubt!

upvoted 1 times

DingDongSingSong 3 years, 10 months ago

Option D is a repeat

upvoted 3 times

Thornehead 3 years, 10 months ago

Use Azure portal Python SDK or Azure CLI Use Resource Manager template https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace?tabs=python

upvoted 4 times

Thornehead 3 years, 10 months ago

Use Azure portal or Python SDK Use Azure CLI Use Resource Manager template This is from the https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace?tabs=python

upvoted 1 times

Question #53 *Topic 2*

HOTSPOT -

You create an Azure Machine Learning workspace and set up a development environment. You plan to train a deep neural network (DNN) by using the

Tensorflow framework and by using estimators to submit training scripts.

You must optimize computation speed for training runs.

You need to choose the appropriate estimator to use as well as the appropriate training compute target configuration.

Which values should you use? To answer, select the appropriate options in the answer area.

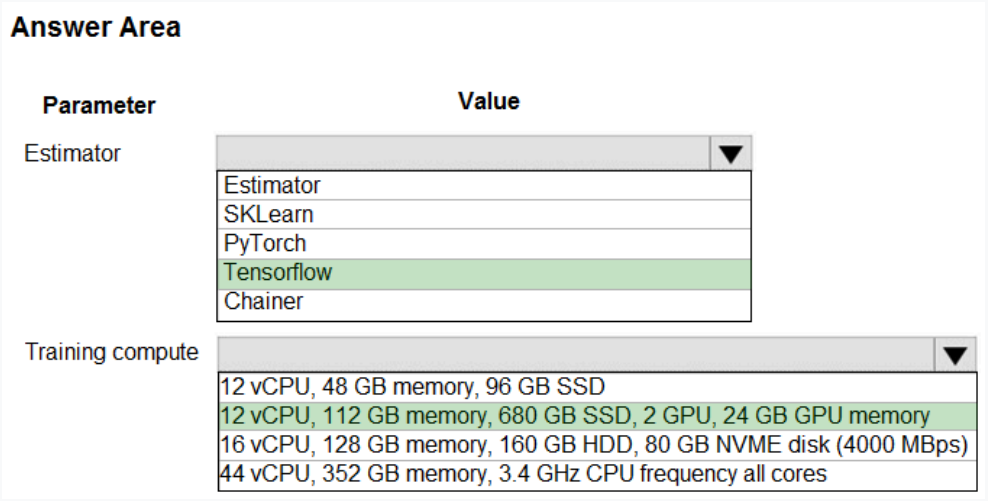
NOTE: Each correct selection is worth one point.

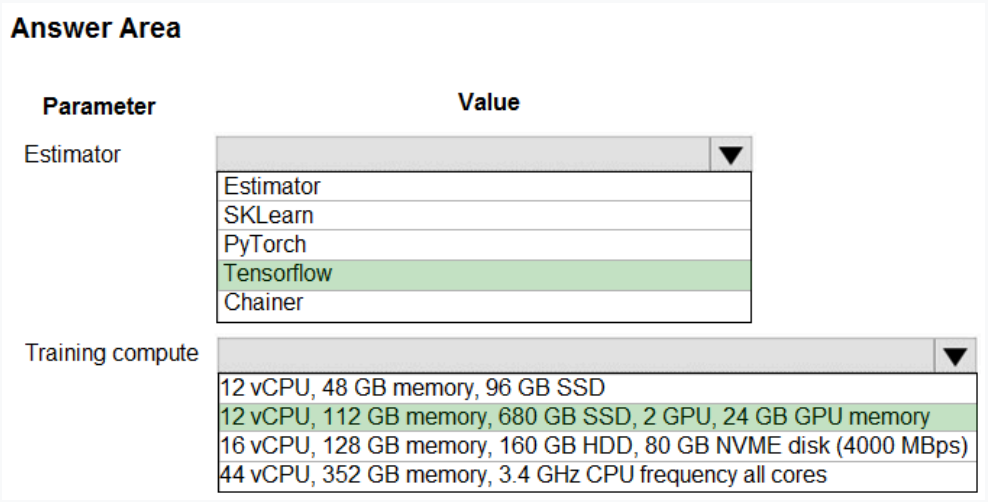
Hot Area:

Answer Area

Parameter	Value
Estimator	Estimator SKLearn PyTorch Tensorflow Chainer
Training compute	12 vCPU, 48 GB memory, 96 GB SSD 12 vCPU, 112 GB memory, 680 GB SSD, 2 GPU, 24 GB GPU memory 16 vCPU, 128 GB memory, 160 GB HDD, 80 GB NVME disk (4000 MBps) 44 vCPU, 352 GB memory, 3.4 GHz CPU frequency all cores

Suggested Answer:





Box 1: Tensorflow - TensorFlow represents an estimator for training in TensorFlow experiments. Box 2: 12 vCPU, 112 GB memory..,2 GPU,.. Use GPUs for the deep neural network. Reference: https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.dnn

MarinaMijailovic 2 years, 2 months ago

How do you know which training compute to choose? There are so many details. How to know in general what are the best options?TensorFlow will normally require GPU for faster computation times, the only machine there with GPU is the solution chosen one

upvoted 1 times

MarinaMijailovic 2 years, 2 months ago

How do you know which training compute to choose? There are so many details. How to know in general what are the best options?

upvoted 1 times

lcgcastro96 Highly Voted 2 years, 2 months ago

TensorFlow will normally require GPU for faster computation times, the only machine there with GPU is the solution chosen one

upvoted 5 times

PremPatrick 2 years, 9 months ago

Why not Pytorch as option?

upvoted 1 times

PremPatrick 2 years, 9 months ago

Why not Pytorch as option?Because it is written "by using the Tensorflow framework"

upvoted 1 times

RicSpd Highly Voted 2 years, 8 months ago

Because it is written "by using the Tensorflow framework"

upvoted 6 times

AjoseO 3 years, 6 months ago

Correct! DNN frameworks include TensorFlow and Pytorch, and works best with GPU!

upvoted 4 times

renuka1234 3 years, 8 months ago

Correct

upvoted 4 times

Question #54 *Topic 2*

HOTSPOT -

You have an Azure Machine Learning workspace named workspace1 that is accessible from a public endpoint. The workspace contains an Azure Blob storage datastore named store1 that represents a blob container in an Azure storage account named account1. You configure workspace1 and account1 to be accessible by using private endpoints in the same virtual network.

You must be able to access the contents of store1 by using the Azure Machine Learning SDK for Python. You must be able to preview the contents of store1 by using Azure Machine Learning studio.

You need to configure store1.

What should you do? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Requirement	Action
Access the contents of store1 by using the Azure Machine Learning SDK for Python.	Set store1 as the default datastore. Disable data validation for store1. Update authentication for store1. Regenerate the keys of account1.
Preview the contents of store1 by using Azure Machine Learning studio.	Set store1 as the default datastore. Disable data validation for store1. Update authentication for store1. Regenerate the keys of account1.

Suggested Answer:

Answer Area

Requirement	Action
Access the contents of store1 by using the Azure Machine Learning SDK for Python.	Regenerate the keys of account1.
Preview the contents of store1 by using Azure Machine Learning studio.	Update authentication for store1.

Answer Area

Requirement	Action
Access the contents of store1 by using the Azure Machine Learning SDK for Python.	Regenerate the keys of account1.
Preview the contents of store1 by using Azure Machine Learning studio.	Update authentication for store1.

Box 1: Regenerate the keys of account1. Azure Blob Storage support authentication through Account key or SAS token. To authenticate your access to the underlying storage service, you can provide either your account key, shared access signatures (SAS) tokens, or service principal Box 2: Update the authentication for store1. For Azure Machine Learning studio users, several features rely on the ability to read data from a dataset; such as dataset previews, profiles and automated machine learning. For these features to work with storage behind virtual networks, use a workspace managed identity in the studio to allow Azure Machine Learning to access the storage account from outside the virtual network. Note: Some of the studio's features are disabled by default in a virtual network. To re-enable these features, you must enable managed identity for storage accounts you intend to use in the studio. The following operations are disabled by default in a virtual network: ✑ Preview data in the studio. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-access-data

roo123 Highly Voted 2 years, 6 months ago

The answer should be 1. Update authentication 2.Set store1 as the default datastore

upvoted 6 times

MelMac Most Recent 1 year, 6 months ago

To access the contents of store1 using the Azure ML Python SDK, the correct action is to update authentication for store1. This ensures that the datastore is properly authenticated and accessible within the Azure Machine Learning workspace. To allow the preview of the contents of store1 using Azure ML studio, you need to set store1 as the default datastore. This ensures that the datastore is readily accessible and can be used for various operations within the Azure ML studio.

upvoted 1 times

jefimija 1 year, 10 months ago

1. set store1 as the default datastore 2. update the authentication

upvoted 1 times

NullVoider_0 2 years, 8 months ago

You should just update the authentication for store 1 for both cases. The key reason this meets both requirements is that the original problem stated that workspace1 and its linked storage account are configured with private endpoints for access within the virtual network.

upvoted 2 times

kel_dp_100 2 years, 9 months ago

second should be Set store1 as the default datastore:

upvoted 4 times

ning 4 years, 3 months ago

If you are in the same vnet, and the key or SAS never changed, why would you want to regenerate the key???

upvoted 2 times

ning 4 years, 3 months ago

If you are in the same vnet, and the key or SAS never changed, why would you want to regenerate the key???The workspace was accessible from a public endpoint, and you reconfigure it for it to be accessible by using private endpoints, and you will need to regenerate the key.

upvoted 2 times

AzureJobsTillRetire 3 years, 6 months ago

The workspace was accessible from a public endpoint, and you reconfigure it for it to be accessible by using private endpoints, and you will need to regenerate the key.

upvoted 3 times

azurelearner666 4 years, 4 months ago

seems correct

upvoted 3 times

HOTSPOT -

You are using an Azure Machine Learning workspace. You set up an environment for model testing and an environment for production.

The compute target for testing must minimize cost and deployment efforts. The compute target for production must provide fast response time, autoscaling of the deployed service, and support real-time inferencing.

You need to configure compute targets for model testing and production.

Which compute targets should you use? To answer, select the appropriate options in the answer area.

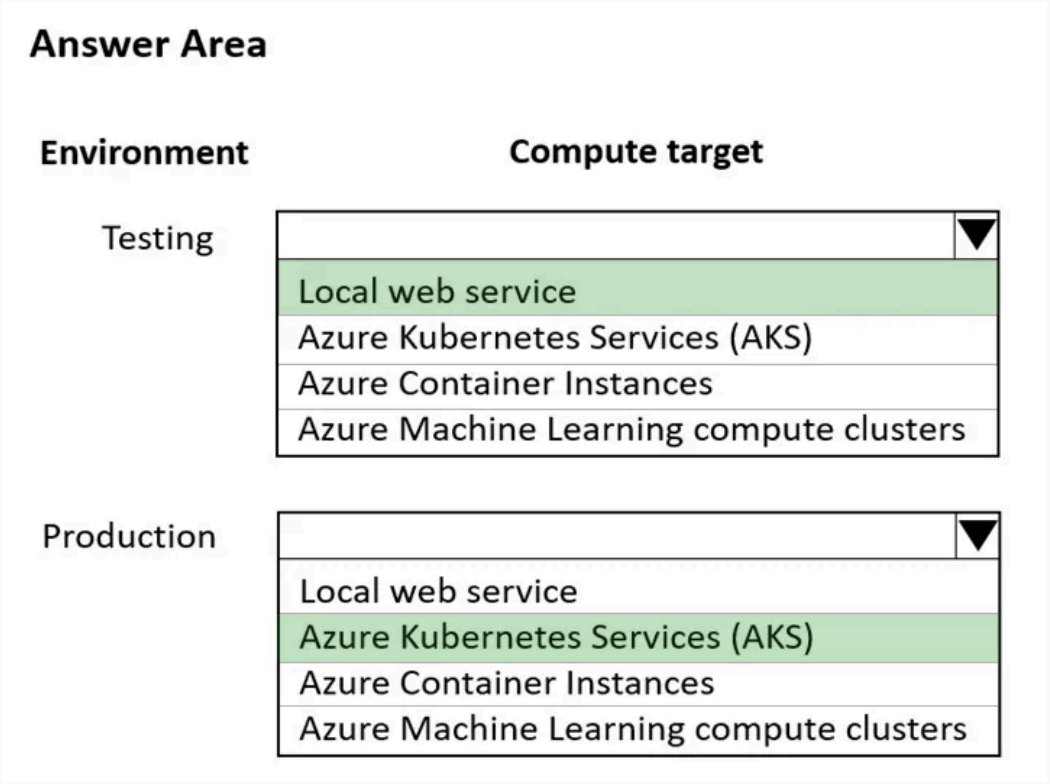
NOTE: Each correct selection is worth one point.

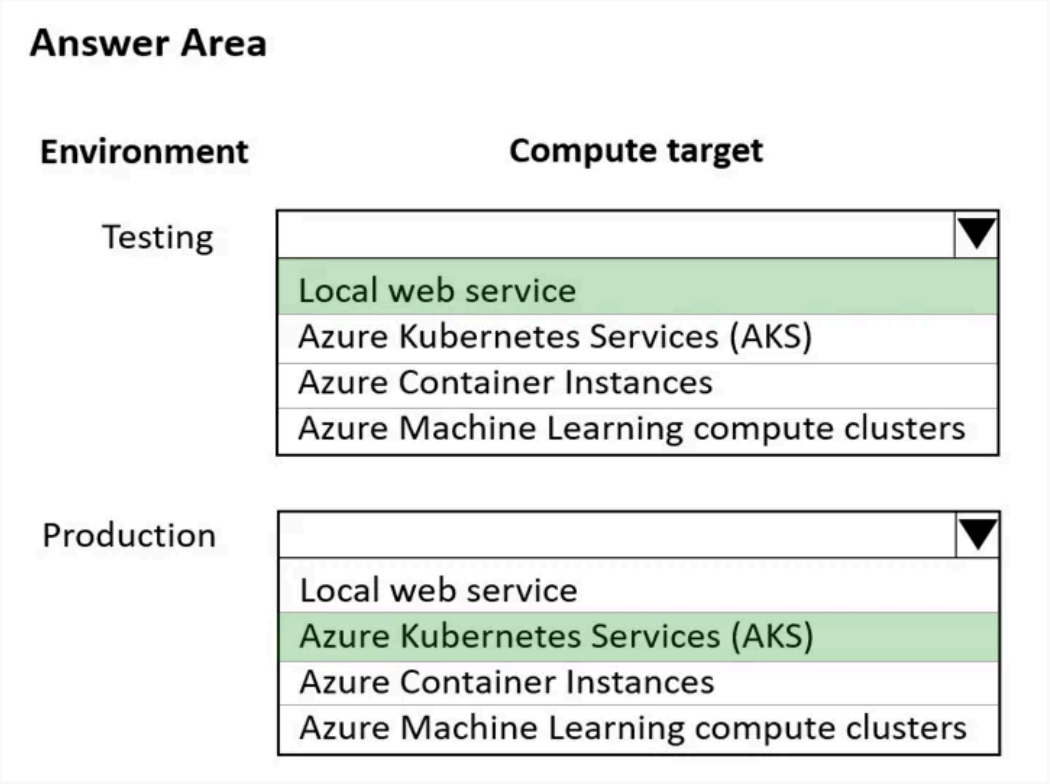
Hot Area:

Answer Area

Environment	Compute target
Testing	Local web service Azure Kubernetes Services (AKS) Azure Container Instances Azure Machine Learning compute clusters
Production	Local web service Azure Kubernetes Services (AKS) Azure Container Instances Azure Machine Learning compute clusters

Suggested Answer:





Box 1: Local web service - The Local web service compute target is used for testing/debugging. Use it for limited testing and troubleshooting. Hardware acceleration depends on use of libraries in the local system. Box 2: Azure Kubernetes Service (AKS) Azure Kubernetes Service (AKS) is used for Real-time inference. Recommended for production workloads. Use it for high-scale production deployments. Provides fast response time and autoscaling of the deployed service Reference: https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target

michaelmorar Highly Voted 3 years, 8 months ago

"You need to configure compute targets for model testing and production." Suggests that we don't need a target for training and evaluation, ONLY to test the deployed model. Therefore: 'Local Web Service' for Test, and AKS for PROD.

upvoted 13 times

sanctafrax Most Recent 2 years, 1 month ago

Question states it is for model testing. Therefore ACI is not needed, Local Webservice should be fine. AKS is required for prod

upvoted 1 times

daviskv74 2 years, 10 months ago

Chat GPT says for testing, ACI suitablei agress ACI for testing and for Production AKS

upvoted 4 times

daviskv74 2 years, 10 months ago

Chat GPT says for testing, ACI suitable

upvoted 4 times

prabhjot 2 years, 6 months ago

i agress ACI for testing and for Production AKS

upvoted 1 times

Yuriy_Ch 3 years, 5 months ago

Exactly this question was on exam 07/March/2023

upvoted 4 times

Edriv 3 years, 7 months ago

https://www.techtarget.com/searchcloudcomputing/definition/Azure-Container-Instances#:~:text=Azure%20Container%20Instances%20is%20a%20service%20that%20enables,having%20to%20provision%20or%20manage%20any%20underlying%20infrastructure.

upvoted 1 times

PremPatrick 3 years, 9 months ago

LocalWebservice for testing (not sure though!) https://learn.microsoft.com/en-us/azure/machine-learning/concept-compute-target

upvoted 2 times

FlexingD 3 years, 9 months ago

ACI for testing + 1

upvoted 4 times

reddragondms 3 years, 10 months ago

Local web services and azure container instances are listed under compute targets for inference, not training targets. Wouldn't the order for "minimize cost and deployment efforts" for testing be first "Local computer" then "Azure ML compute cluster"? https://learn.microsoft.com/en-us/azure/machine-learning/concept-compute-target

upvoted 1 times

claps92 3 years, 11 months ago

ACI for testing (minimizes the deployment effort)

upvoted 4 times

klowqw 3 years, 11 months ago

ACI for testingLocal for testing

upvoted 2 times

klowqw 3 years, 11 months ago

ACI for testing

upvoted 2 times

klowqw 3 years, 11 months ago

Local for testing

upvoted 1 times

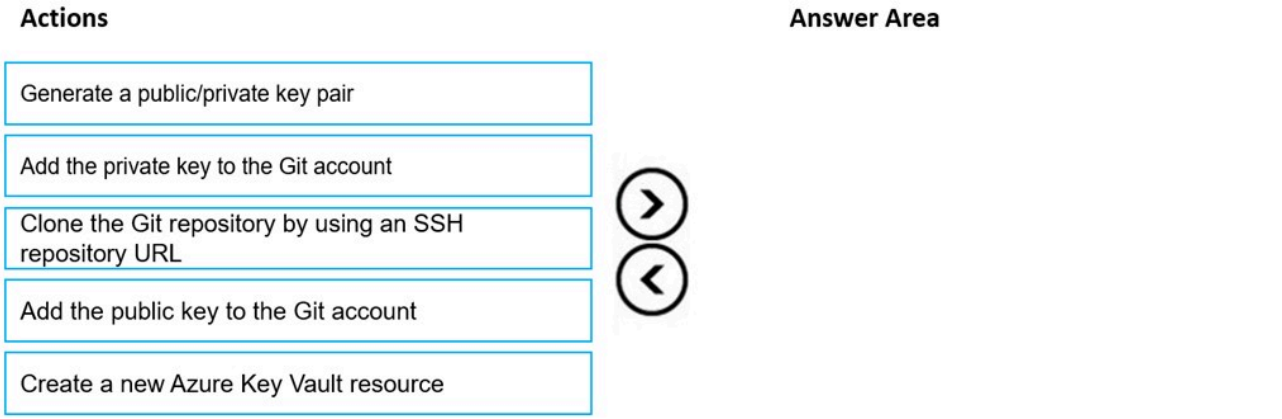
DRAG DROP -

You are using a Git repository to track work in an Azure Machine Learning workspace.

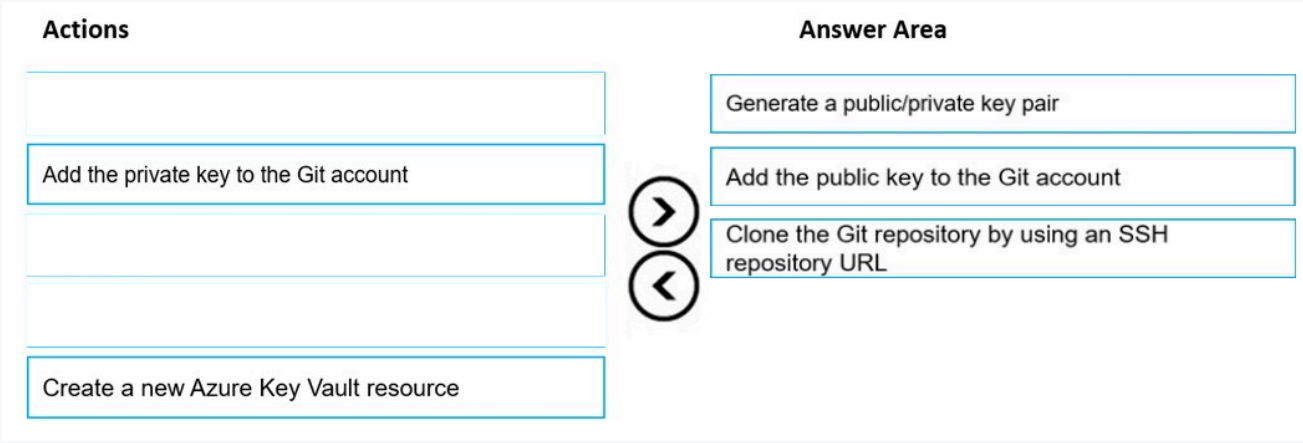
You need to authenticate a Git account by using SSH.

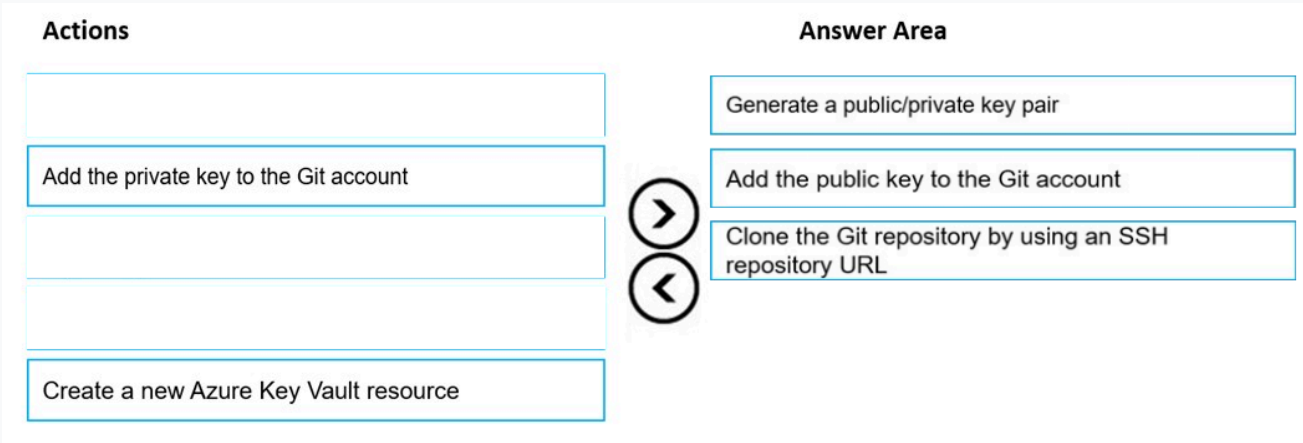
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Select and Place:



Suggested Answer:





Authenticate your Git Account with SSH: Step 1: Generating a public/private key pair

Generate a new SSH key - 1. Open the terminal window in the Azure Machine Learning Notebook Tab. 2. Paste the text below, substituting in your email address. ssh-keygen -t rsa -b 4096 -C "your_email@example.com" This creates a new ssh key, using the provided email as a label. > Generating public/private rsa key pair. Step 2: Add the public key to the Git Account In your terminal window, copy the contents of your public key file. Step 3: Clone the Git repository by using an SSH repository URL 1. Copy the SSH Git clone URL from the Git repo. 2. Paste the url into the git clone command below, to use your SSH Git repo URL. This will look something like: git clone git@example.com:GitUser/azureml-example.git Cloning into 'azureml-example'. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/concept-train-model-git-integration

silva_831 Highly Voted 3 years, 3 months ago

The given answer is correct.

upvoted 7 times

Kanwal001 Highly Voted 2 years, 5 months ago

On exam 28 Aug 2023

upvoted 5 times

thisiston Most Recent 1 year, 9 months ago

1. Generate a public/private key pair. This is the initial step in setting up SSH authentication. The key pair consists of a private key, which you keep secure on your machine, and a public key that you will share. 2. Add the public key to the Git account. Once the key pair is generated, the public key must be added to your Git account settings. This allows the Git server to recognize and authenticate your SSH requests. 3. Clone the Git repository by using an SSH repository URL. After your public key is accepted by the Git server, you can securely clone the repository using SSH, which will utilize your private key for authentication.

upvoted 1 times

Karthikat 1 year, 10 months ago

on exam 3/25/2024

upvoted 1 times

NullVoider_0 2 years ago

On exam 12-02-2024.

upvoted 1 times

Awooga 2 years ago

On exam 2024-02-06

upvoted 1 times

Mal42 2 years, 5 months ago

On exam 18 Aug 2023

upvoted 2 times

phydev 2 years, 7 months ago

On exam 20 July 2023.

upvoted 2 times

Yuriy_Ch 2 years, 11 months ago

Exactly this question was on exam 07/March/2023

upvoted 5 times

Yuriy_Ch 2 years, 11 months ago

Exactly this question was on exam 07/March/2023roughly how many questions from this site appeared in your exam if I may ask?

upvoted 5 times

daviduzo 2 years, 7 months ago

roughly how many questions from this site appeared in your exam if I may ask?

upvoted 1 times

RamundiGR 3 years ago

answer looks good

upvoted 4 times

michaelmorar 3 years, 3 months ago

Never share a private key :) Answer is correct.

upvoted 3 times

fvil 3 years, 3 months ago

Appeared on exam 07/11/2022

upvoted 2 times

Question #57 *Topic 2*

You use Azure Machine Learning to train a model based on a dataset named dataset1.
You define a dataset monitor and create a dataset named dataset2 that contains new data.
You need to compare dataset1 and dataset2 by using the Azure Machine Learning SDK for Python.
Which method of the DataDriftDetector class should you use?

A. run

B. get

C. backfill

D. update

Suggested Answer: *C*
A backfill run is used to see how data changes over time. Reference: https://docs.microsoft.com/en-us/python/api/azureml-datadrift/azureml.datadrift.datadriftdetector.datadriftdetector

Community vote distribution



84b1989 Most Recent 1 year, 7 months ago
To compare dataset1 and dataset2 for data drift using the Azure Machine Learning SDK for Python, you should use the run method of the DataDriftDetector class. Here's why: Purpose of the run method: The run method is used to execute the data drift detection process. It compares the baseline dataset (dataset1) with the target dataset (dataset2) and calculates metrics such as drift magnitude and feature-level drift. Steps to use the run method: Define a DataDriftDetector object, specifying the baseline dataset (dataset1) and the target dataset (dataset2). Call the run method to perform the comparison and generate the data drift report. Output: The run method produces a detailed report that highlights any significant differences between the two datasets, helping you identify potential data drift.
upvoted 2 times

viineet 1 year, 8 months ago
option A is correct, given answer is in correct as backfill use to generate report on datadrift in historical dataset over time
upvoted 3 times

labriji 2 years, 3 months ago
given answer is correct, have a look at the code provided in the official doc: https://learn.microsoft.com/en-us/azure/machine-learning/v1/how-to-monitor-datasets?view=azureml-api-1&tabs=python#:~:text=SDK%20documentation.-,Create%20dataset%20monitor,-Create%20a%20dataset
upvoted 4 times

reddragondms 2 years, 10 months ago
A DataDriftDetector object represents a data drift job definition that can be used to run three job run types: an adhoc run for analyzing a specific day's worth of data; see the run method. a scheduled run in a pipeline; see the enable_schedule method. a backfill run to see how data changes over time; see the backfill method.
upvoted 2 times

klowqw 2 years, 11 months ago
correct
upvoted 3 times

Question #58 *Topic 2*

You use an Azure Machine Learning workspace.
You have a trained model that must be deployed as a web service. Users must authenticate by using Azure Active Directory.
What should you do?

A. Deploy the model to Azure Kubernetes Service (AKS). During deployment, set the token_auth_enabled parameter of the target configuration object to true

B. Deploy the model to Azure Container Instances. During deployment, set the auth_enabled parameter of the target configuration object to true

C. Deploy the model to Azure Container Instances. During deployment, set the token_auth_enabled parameter of the target configuration object to true

D. Deploy the model to Azure Kubernetes Service (AKS). During deployment, set the auth.enabled parameter of the target configuration object to true

Suggested Answer: *A*
To control token authentication, use the token_auth_enabled parameter when you create or update a deployment Token authentication is disabled by default when you deploy to Azure Kubernetes Service. Note: The model deployments created by Azure Machine Learning can be configured to use one of two authentication methods: key-based: A static key is used to authenticate to the web service. token-based: A temporary token must be obtained from the Azure Machine Learning workspace (using Azure Active Directory) and used to authenticate to the web service. Incorrect Answers: C: Token authentication isn't supported when you deploy to Azure Container Instances. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-authenticate-web-service

Community vote distribution
A (100%)

madrid99 Highly Voted 3 years, 3 months ago
https://learn.microsoft.com/en-us/azure/machine-learning/v1/how-to-authenticate-web-service So there's a token_auth_enabled parmeter
upvoted 8 times

evangelist Most Recent 1 year, 8 months ago
Explanation: Azure Kubernetes Service (AKS): AKS provides enterprise-grade security, governance, and increased scalability. It is ideal for production deployments. token_auth_enabled parameter: This parameter, when set to true, enables token-based authentication using Azure Active Directory for the deployed web service. The other options do not correctly combine the deployment target and configuration parameters for enabling Azure Active Directory authentication: B and C (Azure Container Instances): While you can deploy models to Azure Container Instances (ACI), they do not support the token_auth_enabled parameter for Azure AD authentication. D (auth.enabled parameter): This is not a recognized parameter for setting up Azure AD authentication in the deployment configuration. Thus, the correct deployment strategy involves using AKS and setting token_auth_enabled to true during the deployment process
upvoted 4 times

daviskv74 2 years, 4 months ago
GPT says D is correct
upvoted 1 times

Nadine_nm 2 years, 5 months ago
the 2 parameters exists in the documentation, there is a slight difference in their definitions : token_auth_enabled = Whether or not to enable Token auth for this Webservice. If this is enabled, users can access this Webservice by fetching access token using their Azure Active Directory credentials auth_enabled = Whether or not to enable key auth for this Webservice. Defaults to True. since we are using Azure Active Directory, answer A is correct
upvoted 3 times

Mal42 2 years, 5 months ago
On exam 18 Aug 2023
upvoted 3 times

Mikku123 2 years, 6 months ago
Its identity based for AAD , not credential based SAS(token). so "D" is the correct answer!
upvoted 2 times

phydev 2 years, 7 months ago

On exam 20 July 2023.

upvoted 1 times

ajay0011 2 years, 10 months ago

I think D is correct. # Define workspace and model details workspace = Workspace.from_config() model = Model(workspace, 'model_name') aks_service_name = 'my-aks-service' deployment_config = AksWebservice.deploy_configuration(cpu_cores=1, memory_gb=1, auth_enabled=True # Enable Azure Active Directory authentication) env = Environment.from_conda_specification(name='env_name', file_path='./conda_dependencies.yml') auth = InteractiveLoginAuthentication() aks_service = Model.deploy(workspace=workspace, name=aks_service_name, models=[model], inference_config=inference_config, deployment_config=deployment_config, environment=env, auth=auth, overwrite=True)

upvoted 2 times

JTWang 3 years, 4 months ago

I can't find token_auth_enabled parameter. In ref i got parameter "auth_mode". https://learn.microsoft.com/en-us/azure/machine-learning/how-to-authenticate-online-endpoint You can set the authentication type when you create an online endpoint. Set the auth_mode to key or aml_token depending on which one you want to use. The default value is key.I didn't find token_auth-enabled as well according to attached link.

upvoted 3 times

JTWang 3 years, 4 months ago

I can't find token_auth_enabled parameter. In ref i got parameter "auth_mode".

https://learn.microsoft.com/en-us/azure/machine-learning/how-to-authenticate-online-endpoint

You can set the authentication type when you create an online endpoint. Set the auth_mode to key or aml_token depending on which one you want to use. The default value is key.

upvoted 3 times

silva_831 3 years, 3 months ago

I didn't find token_auth-enabled as well according to attached link.

upvoted 1 times

klowqw 3 years, 5 months ago

correctCorrect is A: From (v1 SDK) https://learn.microsoft.com/en-us/azure/machine-learning/v1/how-to-authenticate-web-service # Create the config aks_config = AksWebservice.deploy_configuration() # Enable token auth and disable (key) auth on the webservice aks_config = AksWebservice.deploy_configuration(token_auth_enabled=True, auth_enabled=False)

upvoted 3 times

klowqw 3 years, 5 months ago

correct

upvoted 3 times

manualrg 3 years ago

Correct is A: From (v1 SDK) https://learn.microsoft.com/en-us/azure/machine-learning/v1/how-to-authenticate-web-service # Create the config aks_config = AksWebservice.deploy_configuration() # Enable token auth and disable (key) auth on the webservice aks_config = AksWebservice.deploy_configuration(token_auth_enabled=True, auth_enabled=False)

upvoted 2 times

HOTSPOT -

You are the owner of an Azure Machine Learning workspace.

You must prevent the creation or deletion of compute resources by using a custom role. You must allow all other operations inside the workspace.

You need to configure the custom role.

How should you complete the configuration? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

```
{
 "Name": "Data Scientist Custom",
 "IsCustom": true
 "Description": "Description"
 "Actions": [
        [ Microsoft.MachineLearningServices/workspaces/*/read
          Microsoft.MachineLearningServices/workspaces/computes/*/write
          Microsoft.MachineLearningServices/workspaces/delete ]
        [ Microsoft.MachineLearningServices/workspaces/*/write
          Microsoft.MachineLearningServices/workspaces/computes/*/write
          Microsoft.MachineLearningServices/workspaces/delete ]
 ],
 "NotActions": [
        [ Microsoft.MachineLearningServices/workspaces/*/read
          Microsoft.MachineLearningServices/workspaces/*/write
          Microsoft.MachineLearningServices/workspaces/computes/*/delete ]
        [ Microsoft.MachineLearningServices/workspaces/*/read
          Microsoft.MachineLearningServices/workspaces/*/write
          Microsoft.MachineLearningServices/workspaces/computes/*/write ]
 ],
 "AssignableScopes": [
   "/subscriptions/<subscription_id>"
 ]
}
```

Suggested Answer:

Answer Area

```
{
 "Name": "Data Scientist Custom",
 "IsCustom": true
 "Description": "Description"
 "Actions": [
```

- Microsoft.MachineLearningServices/workspaces/*/read
- Microsoft.MachineLearningServices/workspaces/computes/*/write
- Microsoft.MachineLearningServices/workspaces/delete

- Microsoft.MachineLearningServices/workspaces/*/write
- Microsoft.MachineLearningServices/workspaces/computes/*/write
- Microsoft.MachineLearningServices/workspaces/delete

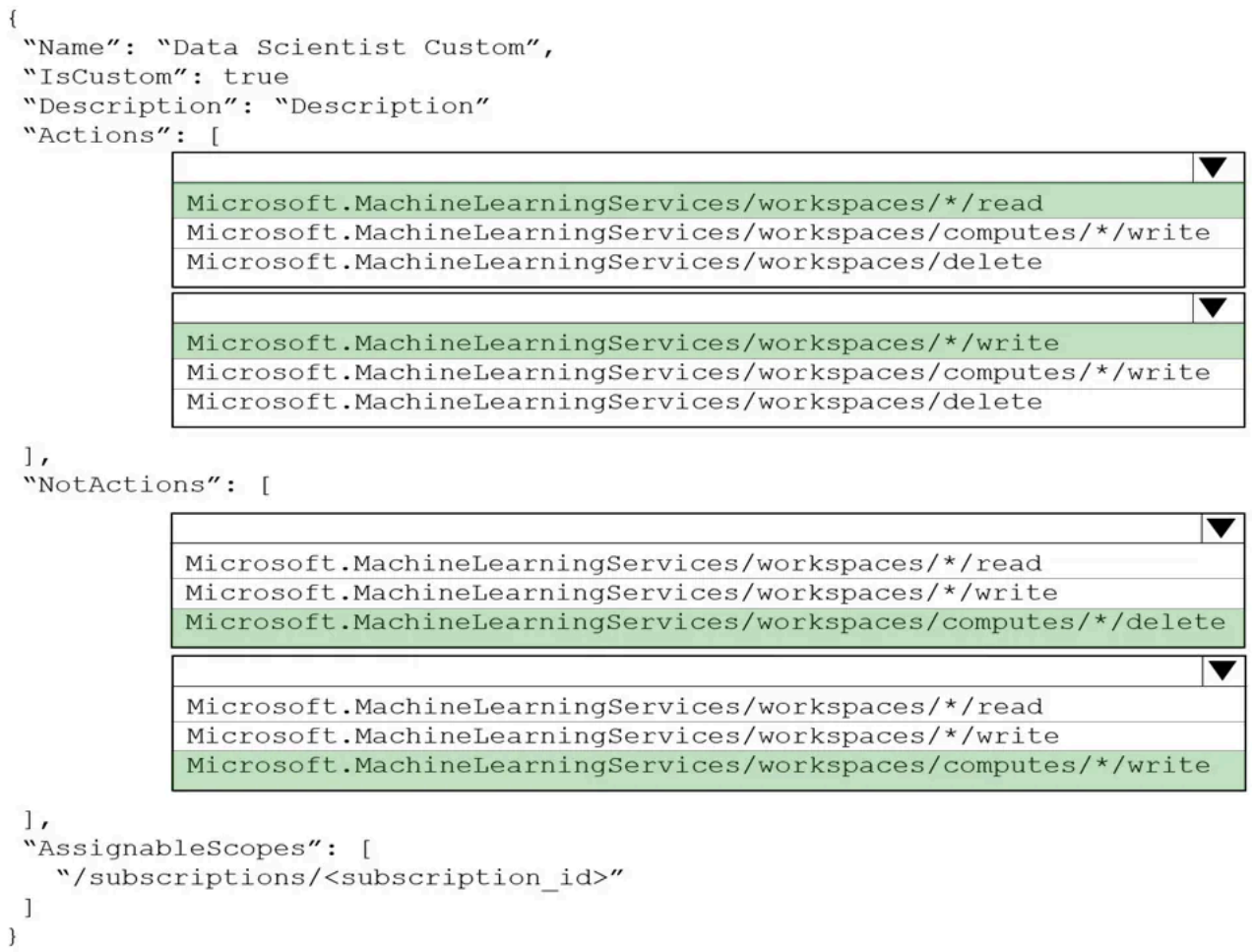
```
 ],
 "NotActions": [
```

- Microsoft.MachineLearningServices/workspaces/*/read
- Microsoft.MachineLearningServices/workspaces/*/write
- Microsoft.MachineLearningServices/workspaces/computes/*/delete

- Microsoft.MachineLearningServices/workspaces/*/read
- Microsoft.MachineLearningServices/workspaces/*/write
- Microsoft.MachineLearningServices/workspaces/computes/*/write

```
 ],
 "AssignableScopes": [
   "/subscriptions/<subscription_id>"
 ]
}
```

Answer Area



Box 1: Microsoft.MachineLearningServices/workspaces/*/read Reader role: Read-only actions in the workspace. Readers can list and view assets, including datastore credentials, in a workspace. Readers can't create or update these assets. Box 2: Microsoft.MachineLearningServices/workspaces/*/write If the roles include Actions that have a wildcard (*), the effective permissions are computed by subtracting the NotActions from the allowed Actions. Box 3: Box 2: Microsoft.MachineLearningServices/workspaces/computes/*/delete Box 4: Microsoft.MachineLearningServices/workspaces/computes/*/write Reference: https://docs.microsoft.com/en-us/azure/role-based-access-control/overview#how-azure-rbac-determines-if-a-user-has-access-to-a-resource

danishanis Highly Voted 1 year, 11 months ago

correct af

upvoted 6 times

klowqw 2 years, 5 months ago

correct

upvoted 4 times

HOTSPOT -

You create an Azure Machine Learning workspace named workspace1. You assign a custom role to a user of workspace1.

The custom role has the following JSON definition:

Instructions: For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

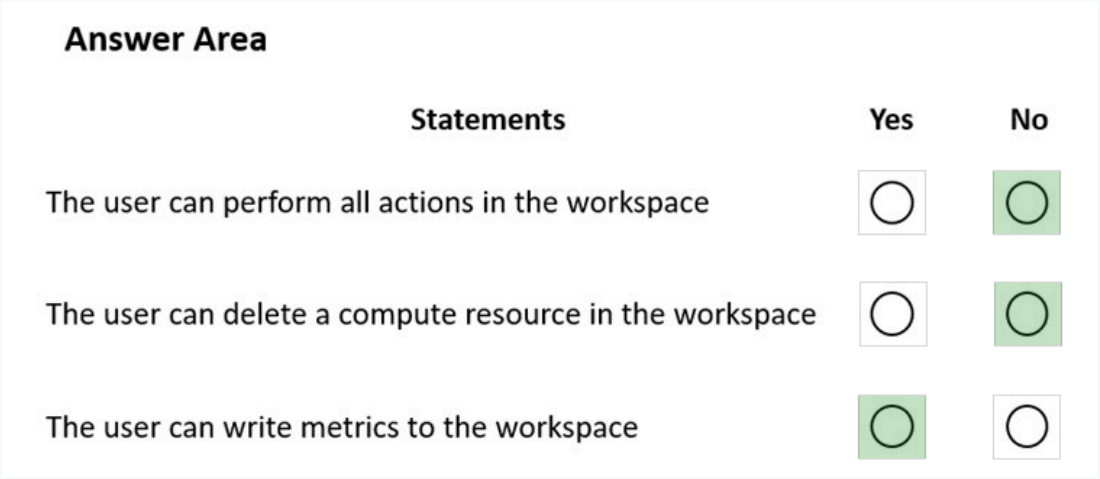
Hot Area:

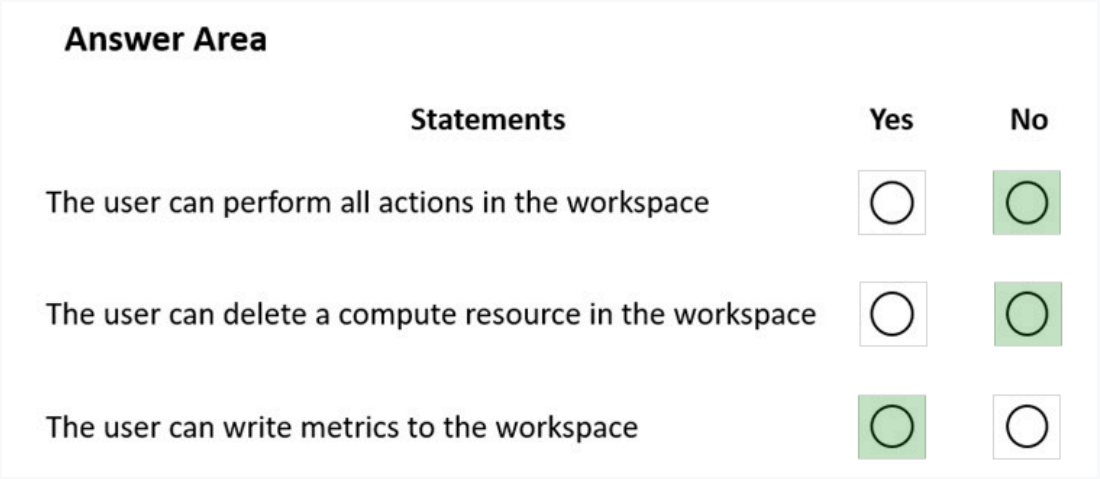
```
{
 "Name": "MyRole",
 "IsCustom": true,
 "Description": "New custom role description.",
 "Actions": ["*"],
 "NotActions": [
   "Microsoft.MachineLearningServices/workspaces/write",
   "Microsoft.MachineLearningServices/workspaces/computes/*/write",
   "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
   "Microsoft. Authorization/*/write"
 ],
 "AssignableScopes": [
   "/subscriptions/<subscription_id>/resourceGroups/resourcegroup1/providers/
   Microsoft.MachineLearningServices/workspaces/workspace1"
 ]
}
```

Answer Area

Statements	Yes	No
The user can perform all actions in the workspace	◯	◯
The user can delete a compute resource in the workspace	◯	◯
The user can write metrics to the workspace	◯	◯

Suggested Answer:





Box 1: No - The actions listed in NotActions are prohibited. If the roles include Actions that have a wildcard (*), the effective permissions are computed by subtracting the NotActions from the allowed Actions.

Box 2: No - Deleting compute resources in the workspace is in the NotActions list.

Box 3: Yes - Writing metrics is not listed in NotActions. Reference: https://docs.microsoft.com/en-us/azure/role-based-access-control/overview#how-azure-rbac-determines-if-a-user-has-access-to-a-resource

PremPatrick Highly Voted 3 years, 3 months ago

Correct. No, No, Yes

upvoted 5 times

ABosco Highly Voted 2 years, 6 months ago

No, No, Yes is correct. The last one is concerning the Roles not related to WS.

upvoted 5 times

smaha_18 Most Recent 8 months ago

In Azure Machine Learning, the relationship between Workspace permissions and logging metrics depends on whether you are talking about the Control Plane (managing the resource) or the Data Plane (running a job and recording results). When a script runs, not writing to the workspace configuration, just appending data to an Asset (the Run) within that workspace. so logging metrics is still possible

upvoted 1 times

PI_Team 2 years, 7 months ago

No,No,NO

Writing metrics specifically is not listed in the NotActions property of the provided role definition. However, the NotActions property does include "Microsoft.MachineLearningServices/workspaces/write", which would likely prevent the user from writing metrics to the workspace, as writing metrics would likely require write permissions on the workspace

SaM

upvoted 3 times

PI_Team 2 years, 7 months ago

No,No,NO Writing metrics specifically is not listed in the NotActions property of the provided role definition. However, the NotActions property does include "Microsoft.MachineLearningServices/workspaces/write", which would likely prevent the user from writing metrics to the workspace, as writing metrics would likely require write permissions on the workspace SaMYou used "likely" twice. So does it or does it not?"Microsoft.MachineLearningServices/workspaces/write", this line only means that It can't create or update the workspace i don't think that writing metrics is part of the "updating workspace"Yes, you are correct! My bad!

upvoted 3 times

phydev 2 years, 7 months ago

You used "likely" twice. So does it or does it not?

upvoted 1 times

Nadine_nm 2 years, 5 months ago

"Microsoft.MachineLearningServices/workspaces/write", this line only means that It can't create or update the workspace i don't think that writing metrics is part of the "updating workspace"Yes, you are correct! My bad!

upvoted 2 times

Nadine_nm 2 years, 5 months ago

"Microsoft.MachineLearningServices/workspaces/write", this line only means that It can't create or update the workspace i don't think that writing metrics is part of the "updating workspace"

upvoted 2 times

PI_Team 2 years, 1 month ago

Yes, you are correct! My bad!

upvoted 4 times

HOTSPOT -

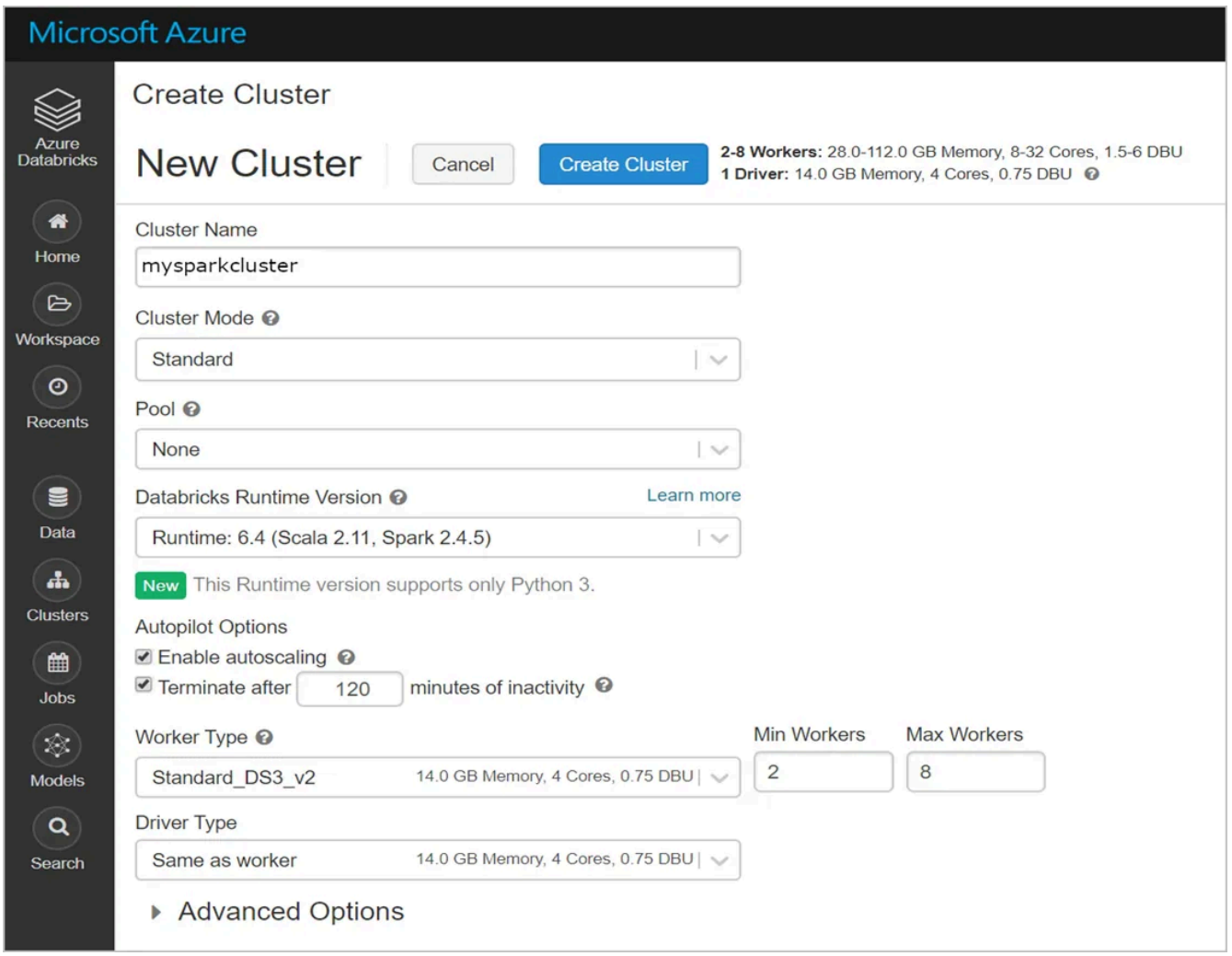
You create a new Azure Databricks workspace.

You configure a new cluster for long-running tasks with mixed loads on the compute cluster as shown in the image below.

Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic.

NOTE: Each correct selection is worth one point.

Hot Area:



Answer Area

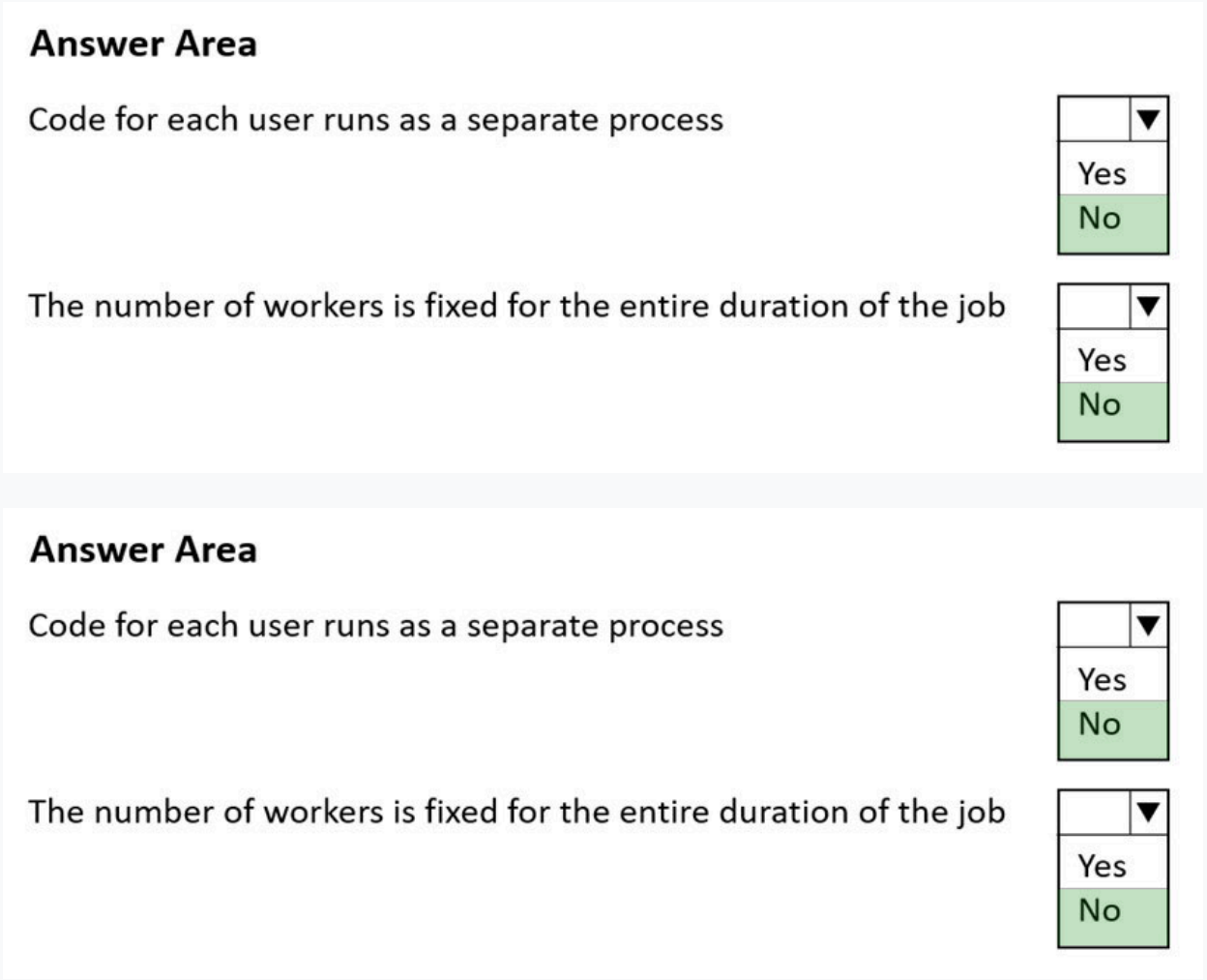
Code for each user runs as a separate process

▼
Yes
No

The number of workers is fixed for the entire duration of the job

▼
Yes
No

Suggested Answer:



Box 1: No - Running user code in separate processes is not possible in Scala.

Box 2: No - Autoscaling is enabled. Minimum 2 workers, Maximum 8 workers. Reference: https://docs.databricks.com/clusters/configure.html

michaelmorar Highly Voted 3 years, 8 months ago

Correct answer: Scala code will be executed inside the Spark JVM (per machine) that is shared between all users, so you can get access to everything that is inside JVM. https://learn.microsoft.com/en-us/answers/questions/924587/azure-databricks-scala-on-high-concurrency-cluster.html High Concurrency clusters can run workloads developed in SQL, Python, and R. The performance and security of High Concurrency clusters is provided by running user code in separate processes, which is not possible in Scala. https://learn.microsoft.com/en-us/azure/databricks/clusters/configure#--high-concurrency-clusters

upvoted 12 times

hiyoww 2 years, 4 months ago

this question is seem related to DP-203 data engineering exam more

upvoted 1 times

cyberfriends 2 years, 10 months ago

why not yes and yes>???

upvoted 1 times

cyberfriends 2 years, 10 months ago

why not yes and yes>???Pool = None - no parallel processes Number of workers vary between 2 and 8 so not fixed

upvoted 1 times

jefimija Most Recent 1 year, 9 months ago

Pool = None - no parallel processes Number of workers vary between 2 and 8 so not fixed

upvoted 1 times

Question #62 *Topic 2*

HOTSPOT

-

You use an Azure Machine Learning workspace. The default datastore contains comma-separated values (CSV) files.

The CSV files must be made available for use in experiments and data processing pipelines. The files must be loaded directly into pandas dataframes.

How should you complete the code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

```
from azureml.core import Workspace
from azureml.core import Dataset

ws = Workspace.from_config()
blob_ds = ws.get_default_datastore()

target_data = [(blob ds, 'data/files/archive/*.csv')]
data1 = [          ▼] . [                              ▼] (path=target_data)
          Dataset                File.from_files
          Datastore              Tabular.from_delimited_files
          Workspace              Tabular.from_pandas_dataframe
registered_data1 = data1.register(workspace=ws, name= 'data1')
```

Suggested Answer:

Answer Area

```
from azureml.core import Workspace
from azureml.core import Dataset

ws = Workspace.from_config()
blob_ds = ws.get_default_datastore()

target_data = [(blob ds, 'data/files/archive/*.csv')]
data1 = [          ▼] . [                              ▼] (path=target_data)
          Dataset                File.from_files
          Datastore              Tabular.from_delimited_files
          Workspace              Tabular.from_pandas_dataframe
registered_data1 = data1.register(workspace=ws, name= 'data1')
```

Answer Area

```
from azureml.core import Workspace
from azureml.core import Dataset

ws = Workspace.from_config()
blob_ds = ws.get_default_datastore()

target_data = [(blob ds, 'data/files/archive/*.csv')]
data1 = [          ▼] . [                              ▼] (path=target_data)
          Dataset                File.from_files
          Datastore              Tabular.from_delimited_files
          Workspace              Tabular.from_pandas_dataframe
registered_data1 = data1.register(workspace=ws, name= 'data1')
```

evangelist Most Recent 1 year, 8 months ago

given answer is correct

upvoted 2 times

Lion007 2 years, 1 month ago

Correct, the answer is: Dataset and Tabular.from_delimited_files To make CSV files available for use in Azure ML experiments and data processing pipelines, and specifically for them to be loaded directly into pandas dataframes, the appropriate method from the Dataset class is Tabular.from_delimited_files. - Tabular.from_delimited_files is designed for structured files like CSVs and will load the data into a format that can be easily converted to pandas dataframes, which is exactly the requirement. Wrong answers: - File.from_files is used when you are working with file datasets that are which a collection of references to the files directly without loading their contents into a structured format, rather than providing a dataset that represents the data in a tabular format like CSV. - Tabular.from_pandas_dataframe is used to create a tabular dataset from an existing pandas dataframe in memory, not from files in a datastore.

upvoted 4 times

NullVoider_0 2 years, 2 months ago

Sorry for the previous response. I've double-checked and updated the code. The answer is correct. from azureml.core import Workspace, Dataset ws = Workspace.from_config() blob_ds = ws.get_default_datastore () target_data = [(blob ds, 'data/files/archive/*.csv')] data1 = Dataset.Tabular.from_delimited_files(path=target_data) registered_datal datal.register(workspaws, name data1')

upvoted 4 times

NullVoider_0 2 years, 2 months ago

The mentioned answer is partially correct. The actual code is given below. from azureml.core import Workspace, Dataset ws = Workspace.from_config() blob_ds = ws.get_default_datastore () target_data = [(blob ds, 'data/files/archive/*.csv')] data1 = Dataset.Files.from_files(path=target_data) registered_datal datal.register(workspaws, name datal')

upvoted 4 times

orionduo 2 years, 5 months ago

It seems correct https://github.com/MicrosoftDocs/azure-docs/blob/main/articles/machine-learning/v1/how-to-create-register-datasets.md#create-a-tabulardataset

upvoted 2 times

ajay0011 2 years, 10 months ago

correct

upvoted 2 times

Question #63 *Topic 2*

HOTSPOT

-

You plan to use a curated environment to run Azure Machine Learning training experiments in a workspace.

You need to display all curated environments and their respective packages in the workspace.

How should you complete the code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

```
from azureml.core import Workspace, Environment
ws = Workspace.from_config()
envs = Environment.list(workspace=ws)
for env in envs:

    if env.startswith(" [ ▼ ] "):
                      AutoML
                      AzureML
                      PyTorch
        print("Name", env)
        print("packages", envs[env].[ ▼ ].conda_dependencies.serialize_to_string())
                                    python
                                    version
                                    docker
```

Suggested Answer:

Answer Area

```
from azureml.core import Workspace, Environment
ws = Workspace.from_config()
envs = Environment.list(workspace=ws)
for env in envs:

    if env.startswith(" [ ▼ ] "):
                      AutoML
                      AzureML
                      PyTorch
        print("Name", env)
        print("packages", envs[env].[ ▼ ].conda_dependencies.serialize_to_string())
                                    python
                                    version
                                    docker
```

Answer Area

```
from azureml.core import Workspace, Environment
ws = Workspace.from_config()
envs = Environment.list(workspace=ws)
for env in envs:

    if env.startswith(" [ ▼ ] "):
                      AutoML
                      AzureML
                      PyTorch
        print("Name", env)
        print("packages", envs[env].[ ▼ ].conda_dependencies.serialize_to_string())
                                    python
                                    version
                                    docker
```

orionduo Most Recent 1 year, 11 months ago

correct https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.conda_dependencies.condadependencies?view=azure-ml-py

upvoted 3 times

vish9 2 years, 2 months ago

The given answer is correct

upvoted 2 times

Question #64 *Topic 2*

You are profiling data by using Azure Machine Learning studio.
You need to detect columns with odd or missing values.
Which statistic should you analyze?

A. Profile

B. Std deviation

C. Error count

D. Type

Suggested Answer: *C*
Official Answer: C

Community vote distribution
B (50%) C (50%)

ajay0011 Highly Voted 3 years, 4 months ago
To detect columns with odd or missing values, you should analyze the Error count statistic. The Error count statistic represents the number of missing or malformed values in each column of your dataset. Analyzing this statistic allows you to detect columns with missing values or with values that don't conform to the expected data type. The Profile statistic provides a summary of the statistical properties of each column, such as the minimum, maximum, mean, median, and quartiles. The Std deviation statistic represents the variation of each column around its mean value. These statistics are useful to understand the distribution and variability of the data, but they don't provide information about missing or malformed values. The Type statistic represents the data type of each column, such as integer, float, or string. While this statistic is useful to understand the structure of the dataset, it doesn't provide information about missing or malformed values. Therefore, the correct answer is option C: Error count.
upvoted 9 times

Abdelsalamhaa Most Recent 2 months, 2 weeks ago
Error Count can't detect odd values stander devetion is
upvoted 1 times

Sadhak 1 year, 9 months ago
C. Error count
upvoted 1 times

Karthikat 2 years, 4 months ago
on exam 3/25/2024
upvoted 2 times

Kanwal001 2 years, 11 months ago
On exam 28 Aug 2023
upvoted 4 times

phydev 3 years, 1 month ago
On exam 20 July 2023.
upvoted 2 times

Jin_22 3 years, 5 months ago
C. Error count To detect columns with odd or missing values, you should analyze the "Error count" statistic. The Error count metric provides information on the number of missing or null values present in each column. By analyzing this metric, you can identify columns that have a high number of missing or null values, which may indicate issues with the data quality or the data collection process. Additionally, you can also identify columns with odd or unexpected values that do not fit the data distribution or that have a high number of outliers.
upvoted 3 times

Question #65 *Topic 2*

You are authoring a notebook in Azure Machine Learning studio.
You must install packages from the notebook into the currently running kernel. The installation must be limited to the currently running kernel only.
You need to install the packages.
Which magic function should you use?

A. !pip

B. %pip

C. !conda

D. %load

Suggested Answer: *B*
Official Answer: B

Community vote distribution
B (100%)

ajay0011 Highly Voted 2 years, 10 months ago
The %pip magic function is used to install Python packages directly from the notebook environment, and it installs packages only in the currently running kernel. This means that any packages you install with %pip will be available only in the notebook session where you install them. The !pip command is used to install Python packages in a system-wide installation of Python. It does not restrict the installation to the currently running kernel, and packages installed with !pip will be available to all Python processes on the system. The !conda command is used to install packages using the Conda package manager. While Conda is a popular package manager, it does not integrate well with the notebook environment and can cause conflicts with the Anaconda distribution of Python that is used in Azure Machine Learning. The %load magic function is used to load the contents of a file into a cell. It is not used for installing packages. Therefore, the correct answer is option B: %pip.
upvoted 12 times

TA_ Most Recent 1 year, 10 months ago
On exam 15-03-2024. Plus, there were a few other questions related to kernels
upvoted 1 times

NullVoider_0 2 years ago
On exam 12-02-2024.
upvoted 2 times

phydev 2 years, 7 months ago
On exam 20 July 2023.
upvoted 4 times

Jin_22 2 years, 11 months ago
The !pip command installs packages from the notebook into the currently running kernel but does not limit the installation to the currently running kernel only. This means that if you have multiple kernels running in your notebook, !pip will install the package in all of them.
upvoted 3 times

Retko 2 years, 11 months ago
Why not !pip?? Can it also be used? Thx
upvoted 1 times

Retko 2 years, 11 months ago
Why not !pip?? Can it also be used? ThxThe %pip magic function is used to install Python packages directly from the notebook environment, and it installs packages only in the currently running kernel. This means that any packages you install with %pip will be available only in the notebook session where you install them. The !pip command is used to install Python packages in a system-wide installation of Python. It does not restrict the installation to the currently running kernel, and packages installed with !pip will be available to all Python processes on the system.
upvoted 1 times

ajay0011 2 years, 10 months ago
The %pip magic function is used to install Python packages directly from the notebook environment, and it installs packages only in the currently running kernel. This means that any packages you install with %pip will be available only in the notebook session where you install them. The !pip

command is used to install Python packages in a system-wide installation of Python. It does not restrict the installation to the currently running kernel, and packages installed with !pip will be available to all Python processes on the system.

upvoted 1 times

Question #66 *Topic 2*

DRAG DROP

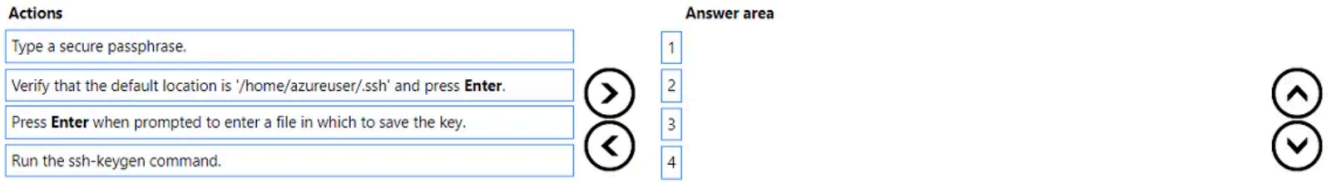
-

You need to implement source control for scripts in an Azure Machine Learning workspace. You use a terminal window in the Azure Machine Learning Notebook tab.

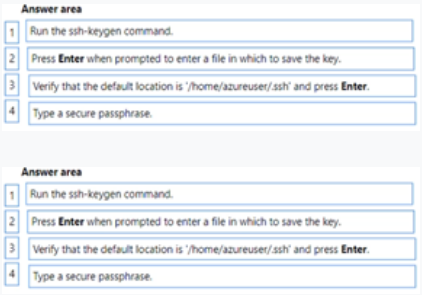
You must authenticate your Git account with SSH.

You need to generate a new SSH key.

Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Suggested Answer:



Kanwal001 Highly Voted 2 years, 5 months ago

On exam 28 Aug 2023

upvoted 7 times

PI_Team Highly Voted 2 years, 7 months ago

Answer is correct: Here is a breakdown of each step: 1. The ssh-keygen command will generate a new SSH key pair. The key pair will consist of a public key and a private key. The public key will be used to authenticate your Git account with SSH, and the private key will be used to access your Git repositories. 2. When you run the ssh-keygen command, you will be prompted to enter a file in which to save the key pair. You can press Enter to accept the default location, which is /home/azureuser/.ssh. 3. The default location for SSH keys is /home/azureuser/.ssh. You should verify that this is the correct location before you press Enter. 4. You will be prompted to type a secure passphrase. The passphrase will be used to protect your private key. You should choose a strong passphrase that you will not forget.

upvoted 5 times

TA_ Most Recent 1 year, 10 months ago

On exam 15-03-2024

upvoted 1 times

NullVoider_0 2 years ago

On exam 12-02-2024.

upvoted 1 times

Jin_22 2 years, 11 months ago

The answer seems correct.

upvoted 1 times

Question #67 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You use Azure Machine Learning designer to load the following datasets into an experiment:
Dataset1 -
Dataset2 -
You need to create a dataset that has the same columns and header row as the input datasets and contains all rows from both input datasets.
Solution: Use the Add Rows module.
Does the solution meet the goal?

Age	Length	Width
3	22	13
7	11	96
18	32	85

Age	Length	Width
11	101	65
6	98	23
33	22	54
17	52	12

A. Yes

B. No

Suggested Answer: *A*
Official Answer: A

sap_dg Highly Voted 2 years, 10 months ago
True. Add Rows module is used to concatenate two datasets. https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/add-rows
upvoted 8 times

Karthikat Most Recent 1 year, 10 months ago
on exam 3/25/2024
upvoted 2 times

NullVoider_0 2 years ago
On exam 12-02-2024.
upvoted 2 times

phydev 2 years, 7 months ago
On exam 20 July 2023.
upvoted 4 times

Question #68 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You use Azure Machine Learning designer to load the following datasets into an experiment:
Dataset1 -
Dataset2 -
You need to create a dataset that has the same columns and header row as the input datasets and contains all rows from both input datasets.
Solution: Use the Apply Transformation module.
Does the solution meet the goal?

Age	Length	Width
3	22	13
7	11	96
18	32	85

Age	Length	Width
11	101	65
6	98	23
33	22	54
17	52	12

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Community vote distribution
A (100%)

Karthikat 2 years, 4 months ago
on exam 3/25/2024
upvoted 2 times

TA_ 2 years, 4 months ago
This set-up of questions on exam 15-03-2024
upvoted 1 times

Plb2 2 years, 5 months ago
Assuming we're speaking of the 'Apply SQL Transformation' component here and reading the documentation I was thinking this could work: "Using the Apply SQL Transformation component, you can: - Create tables for results and save the datasets in a portable database" and "The component can take up to three datasets as inputs". However further down it's stated that with SQLLite "LEFT OUTER JOIN is implemented, but not RIGHT OUTER JOIN or FULL OUTER JOIN". https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/apply-sql-transformation?view=azureml-api-2Correcting the selected anwer; should be B (no)Why would you assume that if there is a component named Apply Transformation exactly? And this component does not fit the problem. The answer should be B.
upvoted 1 times

Plb2 2 years, 5 months ago
Selected Answer: A
Assuming we're speaking of the 'Apply SQL Transformation' component here and reading the documentation I was thinking this could work: "Using the Apply SQL Transformation component, you can: - Create tables for results and save the datasets in a portable database" and "The component can take up to three datasets as inputs".

However further down it's stated that with SQLLite "LEFT OUTER JOIN is implemented, but not RIGHT OUTER JOIN or FULL OUTER JOIN".

https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/apply-sql-transformation?view=azureml-api-2

upvoted 1 times

Plb2 2 years, 5 months ago

Correcting the selected anwer; should be B (no)

upvoted 1 times

DaniloMagone Most Recent 2 years, 3 months ago

Why would you assume that if there is a component named Apply Transformation exactly? And this component does not fit the problem. The answer should be B.

upvoted 1 times

phydev 3 years, 1 month ago

On exam 20 July 2023.

upvoted 4 times

Dp_100_11 3 years, 2 months ago

No, the solution provided using the Apply Transformation module does not meet the goal of creating a dataset that contains all rows from both input datasets with the same columns and header row. The Apply Transformation module is used to apply custom transformations or operations on the dataset, but it does not perform the task of merging or combining two datasets into a single dataset with all rows and columns.

upvoted 4 times

Question #69 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You use Azure Machine Learning designer to load the following datasets into an experiment:
Dataset1 -
Dataset2 -
You need to create a dataset that has the same columns and header row as the input datasets and contains all rows from both input datasets.
Solution: Use the Execute Python Script module.
Does the solution meet the goal?

Age	Length	Width
3	22	13
7	11	96
18	32	85

Age	Length	Width
11	101	65
6	98	23
33	22	54
17	52	12

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Community vote distribution
A (85%) B (15%)

Alex310andra Highly Voted 3 years, 5 months ago
Why not?
upvoted 6 times

ajay0011 Highly Voted 3 years, 4 months ago
Selected Answer: A
Yes, it is possible. with the script we can perform any operation
upvoted 5 times

ajay0011 3 years, 4 months ago
Yes, it is possible. with the script we can perform any operationsame here, with Python script, we can perform any operation
upvoted 5 times

chaymat 3 years, 4 months ago
same here, with Python script, we can perform any operation
upvoted 3 times

25f71e3 Most Recent 1 year, 11 months ago
It is possible using the Execute Python Script module in Azure Machine Learning designer, but this approach is unnecessarily complex for this task. Use Add Rows instead. Poorly written question.
upvoted 1 times

bbe8966 2 years, 2 months ago
No, The statement specifies that a dataset should be created with the same columns and contain all rows from both datasets. When performing a join, regardless of the type, we will not obtain a single dataframe with all the rows; this can only be achieved through concatenation.
upvoted 1 times

Karthikat 2 years, 4 months ago

on exam 3/25/2024

upvoted 1 times

NullVoider_0 2 years, 6 months ago

On exam 12-02-2024.

upvoted 1 times

Mikku123 3 years ago

B seems the correct answer as the question explicitly states that you are using Azure Machine Learning designer to load the datasets into an experiment, and the Execute Python Script module is not used for loading datasets directly.

upvoted 4 times

Ahmed_Gehad 3 years ago

The Execute Python Script module can be used to execute Python code in Azure Machine Learning Studio. However, it cannot be used to create a dataset that has the same columns and header row as the input datasets and contains all rows from both input datasets.

upvoted 1 times

phydev 3 years, 1 month ago

On exam 20 July 2023.

upvoted 3 times

SovanMistry 3 years, 1 month ago

With Script it's possible

upvoted 2 times

Dp_100_11 3 years, 2 months ago

Selected Answer: A

Within the Python script, you can read the input datasets, concatenate or merge the rows using pandas functions (e.g., pd.concat() or pd.merge()), and generate a new dataset that contains all rows from both inputs. By specifying the appropriate logic in the Python script, you can ensure that the resulting dataset has the same columns and header row as the input datasets.

upvoted 4 times

Dp_100_11 3 years, 2 months ago

Within the Python script, you can read the input datasets, concatenate or merge the rows using pandas functions (e.g., pd.concat() or pd.merge()), and generate a new dataset that contains all rows from both inputs. By specifying the appropriate logic in the Python script, you can ensure that the resulting dataset has the same columns and header row as the input datasets.No, You use Azure Machine Learning designer

upvoted 4 times

snegnik 3 years, 2 months ago

No, You use Azure Machine Learning designer

upvoted 2 times

HOTSPOT

-

You must use an Azure Data Science Virtual Machine (DSVM) as a compute target.

You need to attach an existing DSVM to the workspace by using the Azure Machine Learning SDK for Python.

How should you complete the following code segment? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

```
from azureml.core.compute import RemoteCompute, ComputeTarget
compute_target_name = "dsvm"
config = RemoteCompute.                (resource_id='<resource_id>',
```

attach_configuration
get_credentials
detach

```
 ssh_port=22, username='<username>', private_key_file='./.ssh/id_rsa')
compute = ComputeTarget.               (ws, compute_target_name, config)
```

detach
create
attach

```
compute.wait_for_completion(show_output=True)
```

Suggested Answer:





Tommo565 Highly Voted 2 years, 4 months ago

Correct (I think) https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.remote.remotecompute?view=azure-ml-py

upvoted 9 times

Tommo565 Highly Voted 2 years, 4 months ago

Correct (I think) https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.remote.remotecompute?view=azure-ml-pycorrect

upvoted 9 times

ajay0011 2 years, 4 months ago

correct

upvoted 2 times

SovanMistry Most Recent 2 years, 1 month ago

correct You can also use subscription_id, resource_group and vm_name without constructing resourceId. with Attach Configuaration and attach functions

upvoted 2 times

Question #71 *Topic 2*

HOTSPOT

-

You have an Azure Machine Learning workspace.

You run the following code in a Python environment in which the configuration file for your workspace has been downloaded.

Instructions: For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

```
from azureml.core import Workspace
from azureml.core import Experiment
import pandas as pd
import datetime as dt
ws = Workspace.from_config()
experiment = Experiment(workspace=ws, name= 'my_experiment')
run = experiment.start_logging()
print('run_time', dt.datetime.now())

row_count = (len(data))
run.log('observations', row_count)
run.complete()
```

Answer Area

Statements	Yes	No
An error will occur if an experiment named my_experiment does not already exist in the workspace.	○	○
If the experiment does not exist, it will be created. If the experiment does exist, the code will create a new run of the existing experiment.	○	○
After the code completes, a metric named run_time is recorded in the experiment run. The metric will contain the date and time for the run.	○	○
After the code completes, the data.csv file will be available in the run's output.	○	○

Suggested Answer:

Answer Area

Statements	Yes	No
An error will occur if an experiment named my_experiment does not already exist in the workspace.	○	○
If the experiment does not exist, it will be created. If the experiment does exist, the code will create a new run of the existing experiment.	○	○
After the code completes, a metric named run_time is recorded in the experiment run. The metric will contain the date and time for the run.	○	○
After the code completes, the data.csv file will be available in the run's output.	○	○

Answer Area

Statements	Yes	No
An error will occur if an experiment named my_experiment does not already exist in the workspace.	○	○
If the experiment does not exist, it will be created. If the experiment does exist, the code will create a new run of the existing experiment.	○	○
After the code completes, a metric named run_time is recorded in the experiment run. The metric will contain the date and time for the run.	○	○
After the code completes, the data.csv file will be available in the run's output.	○	○

vmcompra Highly Voted 2 years, 7 months ago

NO YES NO (Nothing is recorded, but printed/showed) NO

upvoted 5 times

Plb2 Most Recent 1 year, 11 months ago

run-time will be in the output_logs std_log.txt, but not in the metrics

upvoted 4 times

valgz 2 years, 5 months ago

No, Yes, No, No. I hink the metric that is log is named "observations" and contains the number of rows, not the run_time.

upvoted 3 times

rishi_ram 2 years, 8 months ago

No yes yes no With this code, the current date and time will be printed during the execution of the script, and it will also be captured in the experiment run as part of the output logs.

upvoted 2 times

vish9 2 years, 9 months ago

Given Answer is correct. The code only prints run_time, it does not log it.

upvoted 1 times

labriji 2 years, 9 months ago

Q1 - No Q2 - Yes Q3 - Yes Q4 - No

upvoted 2 times

ajay0011 2 years, 10 months ago

No, Yes, Yes, No C)Yes. The code logs a metric named "run_time" with the current date and time using the dt.datetime.now() function. The metric will contain the timestamp of the run.

upvoted 3 times

sap_dg 2 years, 10 months ago

I will go for No, Yes, Yes, No*No, Yes, No, No

upvoted 2 times

sap_dg 2 years, 10 months ago

I will go for No, Yes, Yes, No

upvoted 2 times

sap_dg Highly Voted 2 years, 10 months ago

*No, Yes, No, No

upvoted 5 times

Question #72 *Topic 2*

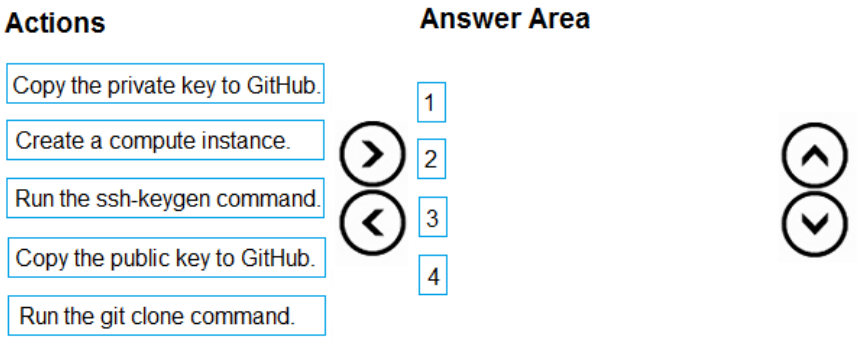
DRAG DROP

-

You create an Azure Machine Learning workspace.

You need to use the shared file system of the workspace to store a clone of a private Git repository.

Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Suggested Answer:

Answer Area

1. Create a compute instance.
2. Run the ssh-keygen command.
3. Copy the public key to GitHub.
4. Run the git clone command.

Answer Area

1. Create a compute instance.
2. Run the ssh-keygen command.
3. Copy the public key to GitHub.
4. Run the git clone command.

PI_Team Highly Voted 2 years, 7 months ago

correct: 1. You need to create a compute instance to access the shared file system of the workspace. 2. You need to run the ssh-keygen command to generate a new SSH key pair. The public key will be used to authenticate your Git account with SSH, and the private key will be used to access your Git repositories. 3. You need to copy the public key to GitHub. The public key will be used by GitHub to authenticate your SSH connection. 4. You can then run the git clone command to clone the private Git repository to the shared file system of the workspace

upvoted 6 times

NullVoider_0 Most Recent 2 years ago

On exam 12-02-2024.

upvoted 2 times

Kanwal001 2 years, 5 months ago

On exam 28 Aug 2023

upvoted 4 times

labriji 2 years, 9 months ago

Given answer is correct 😄 here is the order : * Create a compute instance. * Run the ssh-keygen command. * Copy the public key to GitHub. * Run the git clone command.

upvoted 2 times

Question #73 *Topic 2*

DRAG DROP

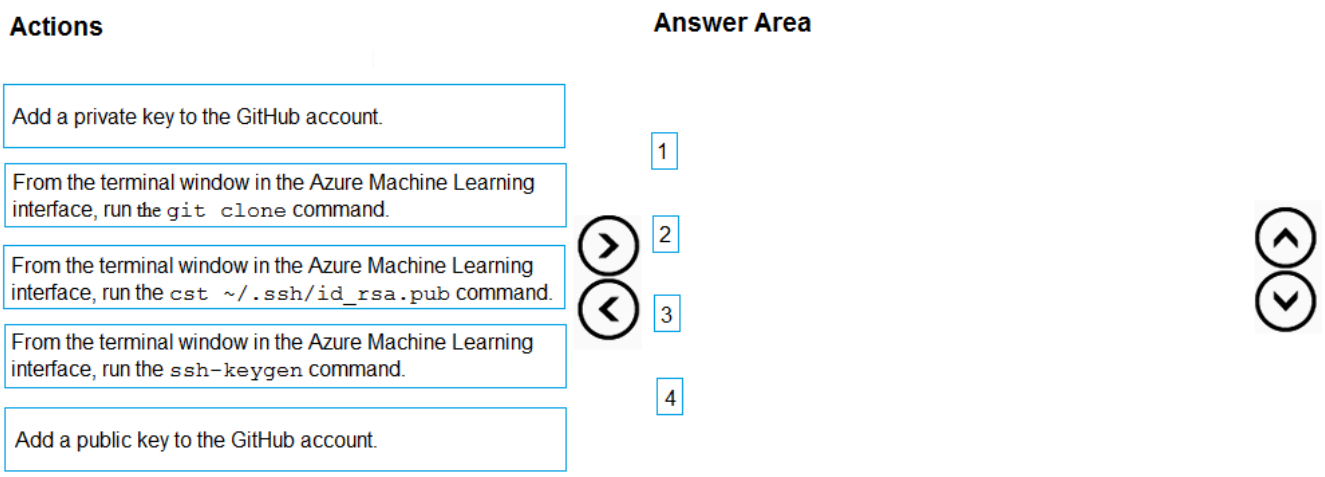
-

You have an existing GitHub repository containing Azure Machine Learning project files.

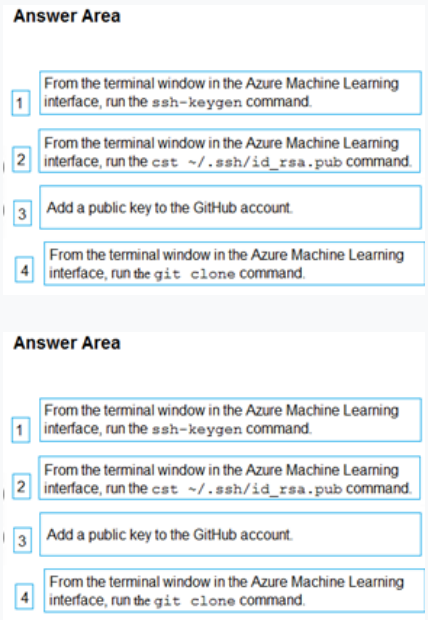
You need to clone the repository to your Azure Machine Learning shared workspace file system.

Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.



Suggested Answer:



Aryabhatta Most Recent 1 year, 11 months ago

The command cat ~/.ssh/id_rsa.pub displays your SSH public key on Linux.

upvoted 3 times

evangelist 2 years, 2 months ago

given answer is correct

upvoted 2 times

og44 2 years, 5 months ago

the second one has a type, need to be cat ~./...

upvoted 1 times

NullVoider_0 2 years, 6 months ago

On exam 12-02-2024.

upvoted 1 times

Awooga 2 years, 6 months ago

On exam 2024-02-06

upvoted 1 times

Kanwal001 2 years, 11 months ago

On exam 28 Aug 2023

upvoted 4 times

labriji 3 years, 3 months ago

Given answer is correct 😄

upvoted 2 times

Question #74 *Topic 2*

You are creating a compute target to train a machine learning experiment.
The compute target must support automated machine learning, machine learning pipelines, and Azure Machine Learning designer training.
You need to configure the compute target.
Which option should you use?

A. Azure HDInsight

B. Azure Machine Learning compute cluster

C. Azure Batch

D. Remote VM

Suggested Answer: *B*
Official Answer: B

Community vote distribution
B (100%)

evangelist Most Recent 1 year, 8 months ago
Azure Machine Learning compute cluster is specifically designed to support various machine learning tasks, including automated machine learning, machine learning pipelines, and Azure Machine Learning designer. It provides scalable compute resources that are optimized for these types of workloads.
upvoted 2 times

Kanwal001 2 years, 5 months ago
On exam 28 Aug 2023
upvoted 3 times

phydev 2 years, 7 months ago
On exam 20 July 2023.
upvoted 3 times

krishna1818 2 years, 8 months ago
Yeah it's Azure Machine Learning Compute that can fulfill all the given criteria's
upvoted 1 times

labriji 2 years, 9 months ago
Given answer is correct 😄
upvoted 1 times

Kvcper333 2 years, 10 months ago
Correct
upvoted 3 times

Question #75 *Topic 2*

You manage an Azure Machine Learning workspace by using the Azure CLI ml extension v2.
You need to define a YAML schema to create a compute cluster.
Which schema should you use?

A. https://azuremlschemas.azureedge.net/latest/computeInstance.schema.json

B. https://azuremlschemas.azureedge.net/latest/amlCompute.schema.json

C. https://azuremlschemas.azureedge.net/latest/vmCompute.schema.json

D. https://azuremlschemas.azureedge.net/latest/kubernetesCompute.schema.json

Suggested Answer: *B*
Official Answer: B

Community vote distribution
B (100%)

evangelist Most Recent 1 year, 8 months ago
amlCompute.schema.json: This schema is specifically designed for Azure Machine Learning compute clusters, which are used for training machine learning models. It supports configurations for scalable and managed compute clusters that can handle automated machine learning, machine learning pipelines, and Azure Machine Learning designer training.
upvoted 2 times

orionduo 2 years, 5 months ago
Correct compute cluster: https://azuremlschemas.azureedge.net/latest/amlCompute.schema.json compute instance: https://azuremlschemas.azureedge.net/latest/computeInstance.schema.json Attached Virtual Machine: https://azuremlschemas.azureedge.net/latest/vmCompute.schema.json Attached Azure Arc-enabled Kubernetes (KubernetesCompute): https://azuremlschemas.azureedge.net/latest/kubernetesCompute.schema.json
upvoted 2 times

labriji 2 years, 9 months ago
Correct Answer is B, here is the link to the docs: https://learn.microsoft.com/en-us/azure/machine-learning/reference-yaml-overview?view=azureml-api-2#:~:text=model.schema.json-,Compute,-Reference
upvoted 3 times

Question #76 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have the following Azure subscriptions and Azure Machine Learning service workspaces:
You need to obtain a reference to the ml-project workspace.
Solution: Run the following Python code:
Does the solution meet the goal?

Subscription	Workspace	Comment
385bdfe5-4cef-4ad4-b977-3f86d92727c9	ml-default	This is default subscription.
5a5891d1-557a-4234-9b83-2e90412b1068	ml-project	The information required to uniquely identify this workspace is stored in the file config.json in the same folder as the Python script.

```
from azureml.core import Workspace
ws = Workspace.from_config()
```

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Community vote distribution
A (50%) B (50%)

PI_Team Highly Voted 3 years ago
Selected Answer: B
The code snippet will only work if the config.json file is present in the current working directory. If the file is not present, the code will throw an exception.

To obtain a reference to the ml-project workspace, we can use the following code:

from azureml.core import Workspace

subscription_id = "5a5891d1-557a-4234-9b83-2e90412b1068" resource_group = "my-resource-group" workspace_name = "ml-project"

ws = Workspace(subscription_id, resource_group, workspace_name)

SaM
upvoted 10 times

PI_Team Highly Voted 3 years ago
The code snippet will only work if the config.json file is present in the current working directory. If the file is not present, the code will throw an exception. To obtain a reference to the ml-project workspace, we can use the following code: from azureml.core import Workspace subscription_id = "5a5891d1-557a-4234-9b83-2e90412b1068" resource_group = "my-resource-group" workspace_name = "ml-project" ws = Workspace(subscription_id, resource_group, workspace_name) SaMit should be Workspace.get()
upvoted 10 times

deyoz 2 years, 5 months ago
it should be Workspace.get()
upvoted 1 times

LadyCasilda 3 years ago
On exam 18 August 2023
upvoted 6 times

ManuelHenriques Most Recent 1 year, 8 months ago

A as the file is in the root folder

upvoted 1 times

jl420 1 year, 9 months ago

yes. this is the purpose of from_config()

upvoted 2 times

FactCheckr4 2 years ago

The Workspace.from_config() method is used to load the workspace configuration from the config.json file, which contains all necessary information to connect to the Azure Machine Learning workspace. This file is typically located in the same directory as your Python script. According to the details in the table, the ml-project workspace information is stored in the config.json file. Using Workspace.from_config() will correctly load the ml-project workspace based on this configuration file. Therefore, the correct answer is: A. Yes

upvoted 2 times

welcomeYM 2 years, 3 months ago

A is the answer.

upvoted 2 times

kzmain 2 years, 3 months ago

A To load the workspace from the configuration file, use the from_config method. https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace?view=azure-ml-py#:~:text=ws%20%3D%20Workspace.from_config()%0A%20%20%20ws.get_details()

upvoted 2 times

Matt2000 2 years, 6 months ago

Yes. Reference: https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace?view=azure-ml-py#azureml-core-workspace-workspace-from-config

upvoted 1 times

dporwal 2 years, 8 months ago

correct answer is b

upvoted 1 times

NullVoider_0 2 years, 8 months ago

With the given code you can only get the workspace configuration to get the reference to the workspace a lot more context is required so the actual given answer is correct not the most voted one.

upvoted 1 times

InversaRadice 2 years, 8 months ago

As ml-project is stored in config.json the script will succeed, so the answer is YES

upvoted 2 times

phydev 3 years, 1 month ago

On exam 20 July 2023.How? This is V1 or the previous version., not the latest one

upvoted 4 times

phydev 3 years, 1 month ago

On exam 20 July 2023.

upvoted 4 times

Ahmed_Gehad 3 years ago

How? This is V1 or the previous version., not the latest one

upvoted 2 times

Norasit 3 years, 1 month ago

If the 'config.json' file is in the same directory as your python script, and you don't pass any input parameters to the 'from_config()' method, it will automatically pick up the configuration from the 'config.json' file in the current directory.

upvoted 3 times

BrahderLau 3 years, 2 months ago

To load the workspace from the configuration file, use the from_config method. ws = Workspace.from_config() ws.get_details() The method provides a simple way to reuse the same workspace across multiple Python notebooks or projects. Users can save the workspace Azure Resource Manager (ARM) properties using the write_config method, and use this method to load the same workspace in different Python notebooks or projects without retyping the workspace ARM properties. Reference: https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace?view=azure-ml-py#remarks

upvoted 4 times

rishi_ram 3 years, 2 months ago

The code ws = Workspace.from_config() will attempt to load the workspace configuration from the config.json file in the current directory. However, it does not specify the specific workspace name "ml-project" mentioned in the question. from azureml.core import Workspace ws = Workspace.from_config(path='./config.json', name='ml-project')

upvoted 2 times

chaymat 3 years, 4 months ago

I think the answer is A

upvoted 2 times

bibinmj 3 years, 4 months ago

Ans: B The config file should be inside `.azureml` folder, not the root of python script. https://azure.github.io/azureml-cheatsheets/docs/cheatsheets/python/v1/workspace/#helpful-methods

upvoted 1 times

Question #77 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have the following Azure subscriptions and Azure Machine Learning service workspaces:
You need to obtain a reference to the ml-project workspace.
Solution: Run the following Python code:
Does the solution meet the goal?

Subscription	Workspace	Comment
385bdfe5-4cef-4ad4-b977-3f86d92727c9	ml-default	This is default subscription.
5a5891d1-557a-4234-9b83-2e90412b1068	ml-project	The information required to uniquely identify this workspace is stored in the file config.json in the same folder as the Python script.

```
from azureml.core import Workspace
ws = Workspace.get(name="ml-project")
```

A. Yes

B. No

Suggested Answer: *A*
Official Answer: A

Community vote distribution
B (87%) A (13%)

Skip73 Highly Voted 2 years, 9 months ago
Correct Answer is B. You need to reference to the workspace using ws = Workspace.from_config() The config.json contains the subscription_id, resource-id and the workspace name which are required to get the reference to the workspace.
upvoted 6 times

evangelist Most Recent 1 year, 8 months ago
The answer is NO, This code attempts to obtain a reference to the Azure Machine Learning workspace named "ml-project" using the Workspace.get method with only the name parameter. However, the Workspace.get method requires additional parameters to uniquely identify the workspace, such as subscription_id and resource_group. Without these parameters, the method will not be able to locate and reference the workspace correctly.
upvoted 4 times

Plb2 1 year, 11 months ago
besides the wsname, subid and rgname should be provided ws = Workspace.get(name="myworkspace", subscription_id=' ', resource_group='myresourcegroup')adding to this; subid is required if user has access to multiple subscriptions, so in this case it is.
upvoted 3 times

Plb2 1 year, 11 months ago
Selected Answer: B
besides the wsname, subid and rgname should be provided

ws = Workspace.get(name="myworkspace", subscription_id=' ', resource_group='myresourcegroup')
upvoted 3 times

Plb2 1 year, 11 months ago
adding to this; subid is required if user has access to multiple subscriptions, so in this case it is.
upvoted 1 times

zishankamal 2 years ago
ws = Workspace.get(name="myworkspace", subscription_id=' ', resource_group='myresourcegroup') subscription_id and resource_group are option parameters.
upvoted 1 times

Matt2000 2 years ago

Yes. Reference: https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace?view=azure-ml-py#azureml-core-workspace-workspace-get

upvoted 1 times

LadyCasilda 2 years, 6 months ago

On exam 18 August 2023

upvoted 3 times

skumarliaison 2 years, 6 months ago

get Return a workspace object for an existing Azure Machine Learning Workspace. Parameters name - str - Required https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace?view=azure-ml-py#azureml-core-workspace-workspace-get

upvoted 2 times

PI_Team 2 years, 6 months ago

The code snippet will not work because the Workspace.get() method requires the subscription ID and resource group of the workspace as well as the workspace name. The config.json file only contains the workspace name, so the code will not be able to find the workspace. from azureml.core import Workspace subscription_id = "5a5891d1-557a-4234-9b83-2e90412b1068" resource_group = "my-resource-group" workspace_name = "ml-project" ws = Workspace(subscription_id, resource_group, workspace_name) SaM

upvoted 3 times

phydev 2 years, 7 months ago

On exam 20 July 2023.

upvoted 4 times

BrahderLau 2 years, 8 months ago

Answer: Yes ws = Workspace.get(name="myworkspace", subscription_id=' ', resource_group='myresourcegroup') Reference: https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace?view=azure-ml-py#remarks* Answer: No. Because subscription_id and resoruce group is not provided

upvoted 2 times

BrahderLau 2 years, 8 months ago

Answer: Yes

ws = Workspace.get(name="myworkspace", subscription_id=' ', resource_group='myresourcegroup')

Reference: https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace?view=azure-ml-py#remarks

upvoted 2 times

BrahderLau 2 years, 8 months ago

* Answer: No. Because subscription_id and resoruce group is not provided

upvoted 4 times

rishi_ram 2 years, 8 months ago

workspace_name is correct parameter not 'name' https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace?view=azure-ml-py Answer is B for sure

upvoted 4 times

Question #78 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have the following Azure subscriptions and Azure Machine Learning service workspaces:
You need to obtain a reference to the ml-project workspace.
Solution: Run the following Python code:
Does the solution meet the goal?

Subscription	Workspace	Comment
385bdfe5-4cef-4ad4-b977-3f86d92727c9	ml-default	This is default subscription.
5a5891d1-557a-4234-9b83-2e90412b1068	ml-project	The information required to uniquely identify this workspace is stored in the file config.json in the same folder as the Python script.

```
from azureml.core import Workspace
ws = Workspace(workspace_name="ml-project")
```

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Community vote distribution
B (80%) A (20%)

LadyCasilda Most Recent 1 year, 12 months ago
On exam 18 August 2023
upvoted 2 times

phydev 2 years ago
On exam 20 July 2023.
upvoted 2 times

vmcompra 2 years, 1 month ago
Answer is B. NO Using the constructor, you need to include at least the "subscription_id", "resource_group" and "workspace_name" as they are REQUIRED. https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace?view=azure-ml-py
upvoted 4 times

Norasit 2 years, 1 month ago
There is no method in the 'Workspace' class that accepts the 'workspace_name' parameter.
upvoted 1 times

BrahderLau 2 years, 2 months ago
Since, subscription_id and resource_group are required, hence the answer is No Workspace(subscription_id, resource_group, workspace_name, auth=None, _location=None, _disable_service_check=False, _workspace_id=None, sku='basic', tags=None, _cloud='AzureCloud')
https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace?view=azure-ml-py#constructor
upvoted 3 times

krishna1818 2 years, 2 months ago
ws = workspace.get(name='...')
upvoted 1 times

rishi_ram 2 years, 2 months ago
Workspace(subscription_id, resource_group, workspace_name, auth=None, _location=None, _disable_service_check=False, _workspace_id=None, sku='basic', tags=None, _cloud='AzureCloud') parameter name is workspace_name not name hence all other options are wrong This is correct ans
upvoted 1 times

Alex310andra 2 years, 4 months ago

What is the difference between this question and the one before?

upvoted 1 times

Alex310andra 2 years, 4 months ago

What is the difference between this question and the one before?Ah sorry "get" is missing

upvoted 1 times

Alex310andra Highly Voted 2 years, 4 months ago

Ah sorry "get" is missing

upvoted 5 times

HOTSPOT

-

You create an Azure Machine Learning workspace. You use the Azure Machine Learning SDK for Python.

You must create a dataset from remote paths. The dataset must be reusable within the workspace.

You need to create the dataset.

How should you complete the following code segment? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

```
from azureml.core import Dataset
from azureml.data.dataset_factory import DataType
web_paths = ['https://domain.blob.core.windows.net/demo/dataset1.tsv',
  'https://domain.blob.core.windows.net/demo/dataset2.tsv']

ds = Dataset [                          ] (path=web_paths)
               Tabular.from_delimited_files
               Tabular.from_parquet_files

ds = ds.     [                          ] (workspace=workspace,
               update
               register
               unregister_all_versions
    name= 'ds',
    description= 'training data')
```

Suggested Answer:

Answer Area

```
from azureml.core import Dataset
from azureml.data.dataset_factory import DataType
web_paths = ['https://domain.blob.core.windows.net/demo/dataset1.tsv',
  'https://domain.blob.core.windows.net/demo/dataset2.tsv']

ds = Dataset [                          ] (path=web_paths)
               Tabular.from_delimited_files
               Tabular.from_parquet_files

ds = ds.     [                          ] (workspace=workspace,
               update
               register
               unregister_all_versions
    name= 'ds',
    description= 'training data')
```

Answer Area

```
from azureml.core import Dataset
from azureml.data.dataset_factory import DataType
web_paths = ['https://domain.blob.core.windows.net/demo/dataset1.tsv',
  'https://domain.blob.core.windows.net/demo/dataset2.tsv']

ds = Dataset [                          ] (path=web_paths)
               Tabular.from_delimited_files
               Tabular.from_parquet_files

ds = ds.     [                          ] (workspace=workspace,
               update
               register
               unregister_all_versions
    name= 'ds',
    description= 'training data')
```

labriji Highly Voted 1 year, 9 months ago

given answer is correct 😄

upvoted 6 times

hammamse 1 year, 10 months ago

given answer is true

upvoted 2 times

Question #80 *Topic 2*

HOTSPOT

-

You train classification and regression models by using automated machine learning.

You must evaluate automated machine learning experiment results. The results include how a classification model is making systematic errors in its predictions and the relationship between the target feature and the regression model's predictions. You must use charts generated by automated machine learning.

You need to choose a chart type for each model type.

Which chart types should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Model type	Chart type
Classification	Confusion matrix / Calibration curve / Predicted vs. true
Regression	Confusion matrix / Calibration curve / Predicted vs. true

Suggested Answer:

Answer Area

Model type	Chart type
Classification	**Confusion matrix** / Calibration curve / Predicted vs. true
Regression	Confusion matrix / Calibration curve / **Predicted vs. true**

Answer Area

Model type	Chart type
Classification	**Confusion matrix** / Calibration curve / Predicted vs. true
Regression	Confusion matrix / Calibration curve / **Predicted vs. true**

Plb2 Most Recent 1 year, 11 months ago

It's 1. Confusion Matrix 2. Prediction vs True The calibration curve plots a model's confidence in its predictions against the proportion of positive samples at each confidence level (https://learn.microsoft.com/en-us/azure/machine-learning/how-to-understand-automated-ml?view=azureml-api-2#calibration-curve) The predicted vs. true chart plots the relationship between the target feature (true/actual values) and the model's predictions (https://learn.microsoft.com/en-us/azure/machine-learning/how-to-understand-automated-ml?view=azureml-api-2#predicted-vs-true)

upvoted 2 times

Tin_Tin 2 years, 1 month ago

1. Confusion matrix 2. Predicted vs. true https://learn.microsoft.com/en-us/azure/machine-learning/how-to-understand-automated-ml?view=azureml-api-2

upvoted 1 times

labriji 2 years, 9 months ago

* for sure we use the "confusion matrix" for classification problem 😄 * for regression, we dont use "confusion matrix", we use >> "Calibration curve": shows the relationship between the predicted values and the actual values. >> "Predicted vs. true": visual representation of the model's performance. i think both are correct 😄

upvoted 4 times

Question #81 *Topic 2*

HOTSPOT

-

You create an Azure Data Lake Storage Gen2 storage account named storage1containing a file system named fs1 and a folder named folder1.

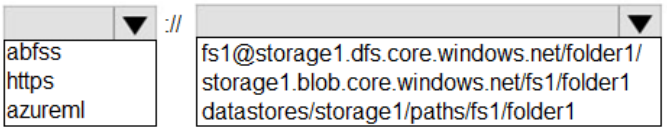
The contents of folder1 must be accessible from jobs on compute targets in the Azure Machine Learning workspace.

You need to construct a URI to reference folder1.

How should you construct the URI? To answer, select the appropriate options in the answer area.

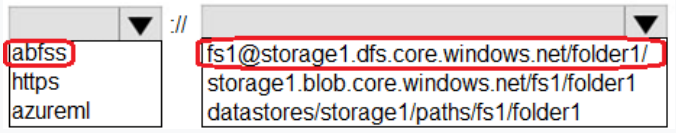
NOTE: Each correct selection is worth one point.

Answer Area

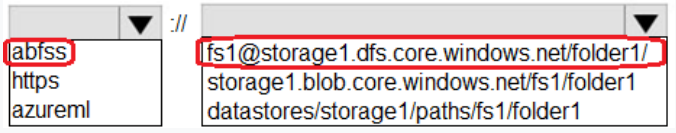


Suggested Answer:

Answer Area



Answer Area



ajay0011 Highly Voted 2 years, 10 months ago

This is a URI (Uniform Resource Identifier) that points to a folder named "folder1" in a file system named "fs1" inside an Azure Data Lake Storage Gen2 storage account named "storage1".

The URI is using the ABFSS (Azure Blob File System Service) protocol to access the storage account. abfss:// @ .dfs.core.windows.net/

upvoted 9 times

ajay0011 Highly Voted 2 years, 10 months ago

This is a URI (Uniform Resource Identifier) that points to a folder named "folder1" in a file system named "fs1" inside an Azure Data Lake Storage Gen2 storage account named "storage1". The URI is using the ABFSS (Azure Blob File System Service) protocol to access the storage account. abfss:// @ .dfs.core.windows.net/ Correct. The URI abfss:// @ .dfs.core.windows.net/ points to a folder named folder1 in a file system named fs1 inside an Azure Data Lake Storage Gen2 storage account named storage1. The URI is using the ABFSS (Azure Blob File System Service) protocol to access the storage account. Here is a breakdown of the URI: abfss: The scheme identifier for the ABFSS protocol. : The name of the file system. @: The delimiter between the file system name and the storage account name. : The name of the storage account. .dfs.core.windows.net: The domain name for the Azure Data Lake Storage Gen2 service. : The path to the folder.

upvoted 9 times

PI_Team 2 years, 6 months ago

Correct. The URI abfss:// @ .dfs.core.windows.net/ points to a folder named folder1 in a file system named fs1 inside an Azure Data Lake Storage Gen2 storage account named storage1. The URI is using the ABFSS (Azure Blob File System Service) protocol to access the storage account. Here is a breakdown of the URI: abfss: The scheme identifier for the ABFSS protocol. : The name of the file system. @: The delimiter between the file system name and the storage account name. : The name of the storage account. .dfs.core.windows.net: The domain name for the Azure Data Lake Storage Gen2 service. : The path to the folder.

upvoted 2 times

Kanwal001 2 years, 5 months ago

On exam 28 Aug 2023

upvoted 5 times

evangelist Most Recent 1 year, 8 months ago

Correct Selections: abfss storage1.dfs.core.windows.net/fs1/folder1 Explanation: abfss: This is the Azure Blob File System Secure (ABFSS) protocol, used for accessing data in Azure Data Lake Storage Gen2. storage1.dfs.core.windows.net: This is the domain for accessing the Data Lake Storage Gen2 account named storage1. fs1/folder1: This is the path within the file system fs1 to the folder folder1.

upvoted 1 times

sai384957324 1 year, 10 months ago

abfss is correct

upvoted 1 times

damaldon 2 years, 7 months ago

correct, for ADLS Gen2: abfss:// @ .dfs.core.windows.net/

upvoted 1 times

pranav33 2 years, 7 months ago

Correct. To construct the URI for referencing the contents of `folder1` in the Azure Data Lake Storage Gen2 storage account, you would need to combine the storage account name, file system name, and folder name in a specific format. Here's how you should construct the URI: ``` abfss:// .dfs.core.windows.net/ / ``` In this case, the correct options for constructing the URI would be: - `storage-account-name`: storage1 - `file-system-name`: fs1 - `folder-name`: folder1 So the correct URI would be: ``` abfss://storage1.dfs.core.windows.net/fs1/folder1 ``` Make sure to replace ` `, ` `, and ` ` with the actual names of your storage account, file system, and folder respectively.

upvoted 1 times

Question #82 *Topic 2*

HOTSPOT

-

You train a model by using Azure Machine Learning. You use Azure Blob Storage to store production data.

The model must be re-trained when new data is uploaded to Azure Blob Storage. You need to minimize development and coding.

You need to configure Azure services to develop a re-training solution.

Which Azure services should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Requirement	Azure service
Identify when new data is uploaded.	Event Grid Event Hubs Functions
Trigger re-training.	Event Grid Functions Logic Apps

Suggested Answer:

Answer Area

Requirement	Azure service
Identify when new data is uploaded.	Event Grid Event Hubs **Functions**
Trigger re-training.	**Event Grid** Functions Logic Apps

Answer Area

Requirement	Azure service
Identify when new data is uploaded.	Event Grid Event Hubs **Functions**
Trigger re-training.	**Event Grid** Functions Logic Apps

SoftAI Highly Voted 2 years, 10 months ago

First one is event grid https://learn.microsoft.com/en-us/azure/data-explorer/create-event-grid-connection?tabs=portal-adx%2Cportal-2 the last one is logic apps: https://learn.microsoft.com/en-us/azure/machine-learning/v1/how-to-trigger-published-pipeline?view=azureml-api-1

upvoted 10 times

chaymat Highly Voted 2 years, 10 months ago

Event Grid, Logic Apps

upvoted 7 times

sanctafrax Most Recent 1 year, 6 months ago

Given answer seems correct. Event grid needs to be filled explicitly (by someone), it does not happen by itself. you need some kind of function. Therefore detecting whether or not there is new data can be done by a timed based triggered function which pushes an event to an event grid. Then trigger the retraining based on the event grid entry.

upvoted 1 times

evangelist 1 year, 8 months ago

The other way around, 1: event grid to monitor event of new data 2: triggering functions to re-train the model

upvoted 1 times

Kanwal001 2 years, 5 months ago

On exam 28 Aug 2023

upvoted 6 times

MarinaMijailovic 2 years, 6 months ago

Event Hubs: Used for high-volume, streaming data ingestion. Event Grid: Used for reactive programming based on event-driven architectures, distributing event notifications.

upvoted 1 times

PI_Team 2 years, 6 months ago

Identify when new data is uploaded: You can use Event Grid to subscribe to events that are emitted when new data is uploaded to Azure Blob Storage. Event Grid will send an event to your subscription whenever a new blob is created or modified. Trigger re-training: Logic Apps can be a good choice for linking data to blob storage and then to ML Studio. Logic Apps are a more complex solution than Functions, but they offer more flexibility and integration capabilities. I think Logic Apps would be a better choice over Functions for triggering as the question asks for linking different services together. As far as I know, Logic Apps can be used to integrate with other Azure services, such as Azure Notification Hubs and Azure Service Bus and if we need to integrate with other Azure services, then Logic Apps may be a better choice. SaM

upvoted 5 times

Norasit 2 years, 7 months ago

Trigger re-training can be use Functions and Logic Apps. In this problem it sound like simple process so I think Functions is better than Logic Apps.

upvoted 1 times

Norasit 2 years, 7 months ago

Trigger re-training can be use Functions and Logic Apps. In this problem it sound like simple process so I think Functions is better than Logic Apps."You need to minimize development and coding." So it has to be Logic Apps over Functions

upvoted 1 times

DaniloMagone 1 year, 9 months ago

"You need to minimize development and coding." So it has to be Logic Apps over Functions

upvoted 1 times

rishi_ram 2 years, 8 months ago

Azure Event Grid: Azure Event Grid can be used to monitor events in Azure Blob Storage. You can configure Event Grid to trigger an event when new data is uploaded to Azure Blob Storage. Azure Functions: Azure Functions is a serverless compute service that can be used to run your re-training code in response to the event triggered by Azure Event Grid. You can write your re-training logic as a function and configure it to be executed when the event is received. By combining Azure Event Grid and Azure Functions, you can create an automated re-training pipeline that triggers the re-training process whenever new data is uploaded to Azure Blob Storage, without the need for extensive coding or manual intervention.it is better to mention that you use ChatGPT

upvoted 2 times

rishi_ram 2 years, 8 months ago

Azure Event Grid: Azure Event Grid can be used to monitor events in Azure Blob Storage. You can configure Event Grid to trigger an event when new data is uploaded to Azure Blob Storage.

Azure Functions: Azure Functions is a serverless compute service that can be used to run your re-training code in response to the event triggered by Azure Event Grid. You can write your re-training logic as a function and configure it to be executed when the event is received.

By combining Azure Event Grid and Azure Functions, you can create an automated re-training pipeline that triggers the re-training process whenever new data is uploaded to Azure Blob Storage, without the need for extensive coding or manual intervention.

upvoted 2 times

snegnik 2 years, 8 months ago

it is better to mention that you use ChatGPT

upvoted 4 times

labriji 2 years, 9 months ago

Identify when new data is uploaded >> Event Grid Trigger re-training >> Azure Logic Apps basta 😅

upvoted 4 times

ajay0011 2 years, 10 months ago

Answer is reverse. you can configure Azure Blob Storage to trigger an event in Azure Event Grid when new data is uploaded. Then, you can create an Azure Function that listens for the event and initiates the re-training process

upvoted 2 times

hammamse 2 years, 10 months ago

I think this is Event Grid + Event Grid?

upvoted 2 times

Tommo565 2 years, 10 months ago

Could also be Event Grid + Event Grid

upvoted 3 times

Tommo565 2 years, 11 months ago

I think this is Event Grid + Logic Apps?

upvoted 5 times

Question #83 *Topic 2*

You use the Azure Machine Learning SDK for Python v1 and notebooks to train a model. You create a compute target, an environment, and a training script by using Python code.
You need to prepare information to submit a training run.
Which class should you use?

A. ScriptRun

B. ScriptRunConfig

C. RunConfiguration

D. Run

Suggested Answer: *B*
Official Answer: B

Community vote distribution
B (100%)

evangelist Most Recent 1 year, 8 months ago
Define the script run configuration src = ScriptRunConfig(source_directory='.', script='train.py', compute_target=compute_target, environment=env)
upvoted 1 times

PI_Team 2 years, 6 months ago
from azureml.core import ScriptRunConfig # Create a ScriptRunConfig object run_config = ScriptRunConfig(name="my-training-run", script="train.py", compute_target="my-compute-target", environment="my-environment",) # Submit the run run = run_config.submit() # Wait for the run to complete run.wait_for_completion()
upvoted 2 times

krishna1818 2 years, 8 months ago
ScriptRunConfig()
upvoted 1 times

Dp_100_11 2 years, 8 months ago
The correct class to use for preparing information to submit a training run in the given scenario is B. ScriptRunConfig. from azureml.core import ScriptRunConfig, Environment, Workspace # Create a ScriptRunConfig script_run_config = ScriptRunConfig(source_directory='path/to/scripts', script='train.py', compute_target='my_compute_target', environment='my_environment') # Submit the training run run = Experiment(workspace=Workspace.get('my_workspace'), name='my_experiment').submit(script_run_config)
upvoted 2 times

ajay0011 2 years, 10 months ago
To submit a training run in Azure Machine Learning using Python code, you should create a ScriptRunConfig object. This class defines the configuration information needed for running a training script on a compute target. The ScriptRunConfig class allows you to specify the compute target, the environment, the training script, and any script arguments or data references needed by the script.
upvoted 3 times

HOTSPOT

-

You have an Azure Machine Learning workspace.

You need to use the Azure Machine Learning SDK for Python to create and register the Azure Data Lake Storage Generation 2 datastore for the workspace.

How should you complete the following code segment? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

```
import azureml.core
from azureml.core import Workspace, Datastore
ws = Workspace.from_config()

datastore = Datastore.[ register_azure_data_lake_gen2 | register_dbfs | register_azure_blob_container ](
    workspace=ws,
    datastore_name='datastore_name',
    account_name='account_name',
    [ database_name=database_name | filesystem='test' | container_name=container_name ],
    tenant_id=tenant_id,
    client_id=client_id,
    client_secret=client_secret)
```

Suggested Answer:

Answer Area

```
import azureml.core
from azureml.core import Workspace, Datastore
ws = Workspace.from_config()

datastore = Datastore.[register_azure_data_lake_gen2 | register_dbfs | register_azure_blob_container](

    workspace=ws,
    datastore_name='datastore_name',
    account_name='account_name',
    [database_name=database_name | filesystem='test' | container_name=container_name],
    tenant_id=tenant_id,
    client_id=client_id,
    client_secret=client_secret)
```

Answer Area

```
import azureml.core
from azureml.core import Workspace, Datastore
ws = Workspace.from_config()

datastore = Datastore.[register_azure_data_lake_gen2 | register_dbfs | register_azure_blob_container](

    workspace=ws,
    datastore_name='datastore_name',
    account_name='account_name',
    [database_name=database_name | filesystem='test' | container_name=container_name],
    tenant_id=tenant_id,
    client_id=client_id,
    client_secret=client_secret)
```

A_PL300 Highly Voted 1 year, 11 months ago

On Exam 9-4-23

upvoted 5 times

PI_Team 2 years ago

Answer is correct. Here is the link to the Microsoft document for it: https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.datastore(class)?view=azure-ml-py#azureml-core-datastore-register-azure-data-lake-gen2 SaM

upvoted 2 times

Ahmed_Gehad 2 years ago

I think this question will be invalid in the new version of the exam. currently it's azure.ai.ml & we use abfs(s) now for data stores in an Azure Data Lake Storage Gen 2

upvoted 3 times

damaldon 2 years, 1 month ago

Cprrect. static register_azure_data_lake_gen2(workspace, datastore_name, filesystem, account_name, tenant_id=None, client_id=None, client_secret=None, resource_url=None, authority_url=None, protocol=None, endpoint=None, overwrite=False, subscription_id=None, resource_group=None, grant_workspace_access=False)

upvoted 1 times

labriji 2 years, 3 months ago

Given answer is correct 😄 snippet code : datastore = Datastore.```register_azure_data_lake_gen2```(workspace=ws, datastore_name='datastore_name', account_name='account_name', ```filesystem```='test', tenant_id='tenant_id', client_id='client_id', client_secret='client_secret')

upvoted 1 times

bibabrian 2 years, 4 months ago

https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.datastore(class)?view=azure-ml-py#azureml-core-datastore-register-azure-data-lake-gen2

upvoted 3 times

HOTSPOT

-

You manage an Azure Machine Learning workspace. You create a training script named sample_training_script.py. The script is used to train a predictive model in the conda environment defined by a file named environment.yml.

You need to run the script as an experiment.

How should you complete the following code segment? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

```
import azureml.core
from azureml.core import Workspace
from azureml.core import Experiment, ScriptRunConfig, Environment
from azureml.widgets import RunDetails
ws = Workspace.from_config()
env = Environment.from_conda_specification("experiment_env", "environment.yml")
object1 = [ScriptRun / ScriptRunConfig / RunConfiguration](source_directory='sample_folder',
                                                           script='sample_training_script.py',
                                                           environment=env)
object2 = Experiment(workspace=ws, name='sample_object2')
run = object2.[get_runs / reactivate / submit](object1)
run.wait_for_completion()
```

Suggested Answer:

Answer Area



Answer Area



orionduo Most Recent 1 year, 11 months ago

from azureml.core import ScriptRunConfig, Experiment # create or load an experiment experiment = Experiment(workspace, 'MyExperiment') # create or retrieve a compute target cluster = workspace.compute_targets['MyCluster'] # create or retrieve an environment env = Environment.get(ws, name='MyEnvironment') # configure and submit your training run config = ScriptRunConfig(source_directory='.', script='train.py', arguments=['--arg1', arg1_val, '--arg2', arg2_val], compute_target=cluster, environment=env) script_run = experiment.submit(config)

upvoted 2 times

Ahmed_Gehad 2 years ago

I think we shall use mlflow now for experiments.

upvoted 2 times

chaymat 2 years, 4 months ago

correct

upvoted 3 times

Question #86 *Topic 2*

You have an Azure Machine Learning workspace. You are connecting an Azure Data Lake Storage Gen2 account to the workspace as a data store. You need to authorize access from the workspace to the Azure Data Lake Storage Gen2 account.
What should you use?

A. Service principal

B. SAS token

C. Managed identity

D. Account key

Suggested Answer: *C*
Official Answer: C

Community vote distribution
A (60%) C (40%)

jmh106 Most Recent 7 months, 3 weeks ago
Cuando conectas Azure Data Lake Storage Gen2 a Azure Machine Learning, la opción recomendada y moderna para autorizar el acceso es: Identidad administrada (Managed Identity) ¿Por qué? Azure ML ya dispone de una identidad administrada (system-assigned o user-assigned). Puedes asignarle permisos RBAC en el ADLS Gen2, por ejemplo: Storage Blob Data Reader Storage Blob Data Contributor No necesitas gestionar secretos, claves ni tokens. Es la opción más segura y alineada con buenas prácticas de Azure.
upvoted 1 times

victorlie 1 year, 5 months ago
Managed Identity is a feature of Azure Active Directory that allows Azure services to authenticate and access other Azure resources securely without needing to manage credentials
upvoted 1 times

Secure_Defense 1 year, 6 months ago
Managed Identity, here's the answer. https://learn.microsoft.com/en-us/azure/machine-learning/how-to-identity-based-service-authentication?view=azureml-api-2&tabs=cli
upvoted 1 times

cayenne06 1 year, 8 months ago
Managed Identity: https://learn.microsoft.com/en-us/azure/machine-learning/how-to-identity-based-service-authentication?view=azureml-api-2&tabs=cli
upvoted 1 times

gunn_m 1 year, 8 months ago
C. Managed identity
upvoted 2 times

sai384957324 2 years, 4 months ago
The answer is Service Principle.
upvoted 1 times

Matt2000 2 years, 6 months ago
Managed Identity. If azure services talk to each other, managed identity is recommended (was also on the DP-203). Note that managed identity is a specific type of service principal. Reference: https://stackoverflow.com/questions/61322079/difference-between-service-principal-and-managed-identities-in-azure
upvoted 2 times

Kanwal001 2 years, 11 months ago
On exam 28 Aug 2023
upvoted 4 times

PI_Team 3 years ago
A. Service principal Explanation: A service principal is an identity that can be used by applications, services, or automation tools to access specific resources within Azure. By creating a service principal and granting it the necessary permissions, you can authorize the Azure Machine Learning workspace to access the Azure Data Lake Storage Gen2 account securely. SaM

upvoted 4 times

heidousl 3 years, 3 months ago

Answer is A

upvoted 1 times

ajay0011 3 years, 3 months ago

Service Principle is correct.

upvoted 1 times

chaymat 3 years, 4 months ago

I think we can use Service principle as well as Managed Identity

upvoted 2 times

avotofu 3 years, 4 months ago

A. Service principle Adding Azure Data Lake Gen. 2 as a datastore can only be authenticated via service principle.Sorry, question is about 'access' to datastore. Still answer should be A. service principle. https://learn.microsoft.com/en-us/azure/machine-learning/v1/how-to-access-data?view=azureml-api-1

upvoted 1 times

avotofu 3 years, 4 months ago

A. Service principle Adding Azure Data Lake Gen. 2 as a datastore can only be authenticated via service principle.

upvoted 1 times

avotofu 3 years, 4 months ago

Sorry, question is about 'access' to datastore. Still answer should be A. service principle. https://learn.microsoft.com/en-us/azure/machine-learning/v1/how-to-access-data?view=azureml-api-1

upvoted 3 times

Question #87 *Topic 2*

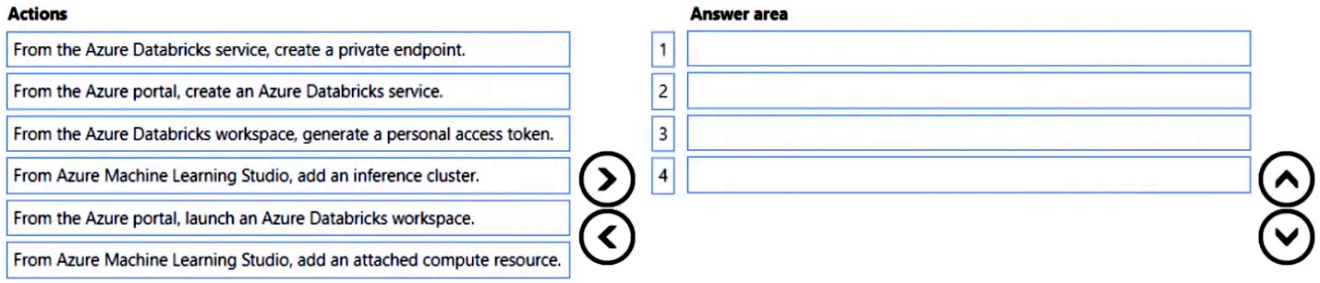
DRAG DROP

-

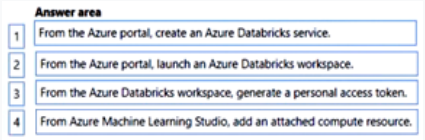
You provision an Azure Machine Learning workspace in a new Azure subscription.

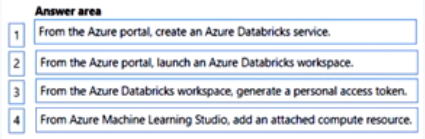
You need to attach Azure Databricks as a compute resource from the Azure Machine Learning workspace.

Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Suggested Answer:





evangelist Most Recent 1 year, 8 months ago

given answer is correct

upvoted 2 times

james2033 2 years, 4 months ago

1. From the Azure portal, create an Azure Databricks service. 2. From the Azure portal, launch an Azure Databricks workspace. 3. From the Azure Databricks workspace, generate a personal access token. 4. From Azure Machine Learning Studio, add an attached compute resource.

upvoted 3 times

SoftAI 2 years, 10 months ago

Seems correct, according to https://www.element61.be/en/resource/how-integrate-azure-databricks-azure-machine-learning-running-big-data-machine-learning

upvoted 1 times

Question #88 *Topic 2*

HOTSPOT

-

You are designing a machine learning solution.

You have the following requirements:

• Use a training script to train a machine learning model.

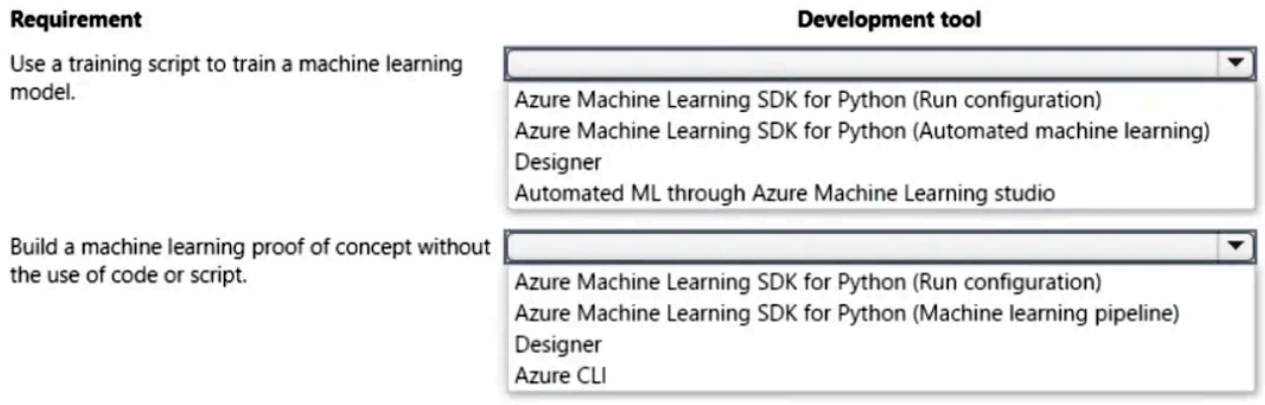
• Build a machine learning proof of concept without the use of code or script.

You need to select a development tool for each requirement.

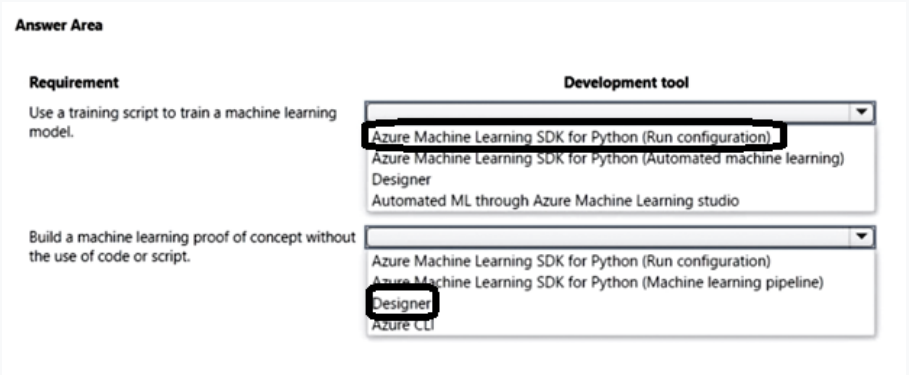
Which development tool should you use? To answer, select the appropriate options in the answer area.

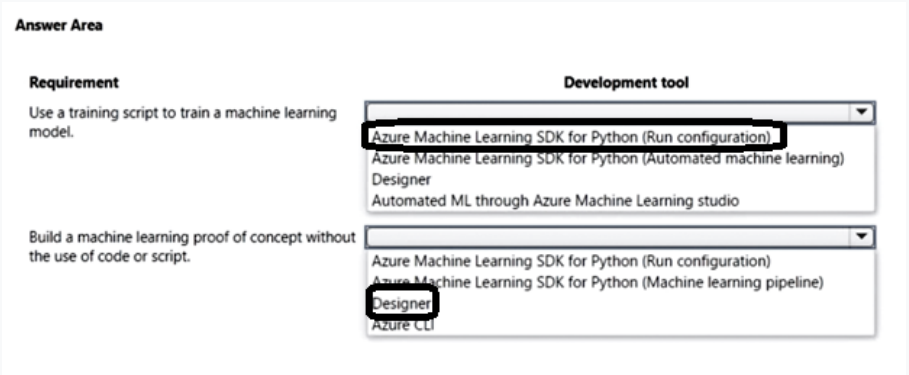
NOTE: Each correct selection is worth one point.

Answer Area



Suggested Answer:





Karthikat Most Recent 1 year, 10 months ago

on exam 3/25/2024

upvoted 1 times

james2033 2 years, 4 months ago

Use a training script to train a machine learning model --> Azure Machine Learning SDK for Python (Run configuration). Build a machine learning proof of concept without the use of code or script. --> Designer.

upvoted 2 times

abuluxury 2 years, 7 months ago

Don't see how the top answer is correct if the say 'without the use of code' SDK uses code right?Only bottom answer says ‘without use of code’

upvoted 1 times

abuluxury 2 years, 7 months ago

Don't see how the top answer is correct if the say 'without the use of code' SDK uses code right?

upvoted 1 times

oso0348 2 years, 6 months ago

Only bottom answer says ‘without use of code’

upvoted 2 times

labriji 2 years, 9 months ago

Given answer looks correct 😄

upvoted 2 times

Question #89 *Topic 2*

You manage an Azure Machine Learning workspace named workspace1.
You must develop Python SDK v2 code to attach an Azure Synapse Spark pool as a compute target in workspace1. The code must invoke the constructor of the SynapseSparkCompute class.
You need to invoke the constructor.
What should you use?

A. Synapse workspace web URL and Spark pool name

B. resource ID of the Synapse Spark pool and a user-defined name

C. pool URL of the Synapse Spark pool and a system-assigned name

D. Synapse workspace name and workspace web URL

Suggested Answer: *B*
Official Answer: B

damaldon Highly Voted 2 years, 7 months ago
Correct. To attach a Synapse Compute using Python SDK, first create an instance of azure.ai.ml.MLClient class. This provides convenient functions for interaction with Azure Machine Learning services. The following code sample uses azure.identity.DefaultAzureCredential for connecting to a workspace in resource group of a specified Azure subscription. In the following code sample, define the SynapseSparkCompute with the parameters: name - user-defined name of the new attached Synapse Spark pool. resource_id - resource ID of the Synapse Spark pool created earlier in the Azure Synapse Analytics workspace.
upvoted 5 times

evangelist Most Recent 1 year, 8 months ago
given answer is correct
upvoted 1 times

Karthikat 1 year, 10 months ago
on exam 3/25/2024
upvoted 1 times

ymj_000 2 years, 3 months ago
For reference: https://learn.microsoft.com/en-us/azure/machine-learning/how-to-manage-synapse-spark-pool?view=azureml-api-2&tabs=sdk
upvoted 1 times

Mal42 2 years, 5 months ago
On exam 18 Aug 2023
upvoted 3 times

Batman160591 2 years, 8 months ago
Answer is B. To invoke the constructor of the SynapseSparkCompute class in Azure Machine Learning SDK v2, you need to provide the resource ID of the Synapse Spark pool and a user-defined name.
upvoted 1 times

Question #90 *Topic 2*

HOTSPOT

-

You manage an Azure Machine Learning workspace by using the Python SDK v2.

You must create a compute cluster in the workspace. The compute cluster must run workloads and property handle interruptions. You start by calculating the maximum amount of compute resources required by the workloads and size the cluster to match the calculations.

The cluster definition includes the following properties and values:

• names="mlcluster"

• size="STANDARD_DS3_v2"

• min_instances=1

• max_instances=4

• tier="dedicated"

The cost of the compute resources must be minimized when a workload is active or idle. Cluster property changes must not affect the maximum amount of compute resources available to the workloads run on the cluster.

You need to modify the cluster properties to minimize the cost of compute resources.

Which properties should you modify? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Workload status	Property
active	size / tier / max_instances
load	size / min_instances / max_instances

Suggested Answer:





damaldon Highly Voted 2 years, 1 month ago

Correct. Lower your compute cluster cost with low priority VMs. from azure.ai.ml.entities import AmlCompute cluster_low_pri = AmlCompute(name="low-pri-example", size="STANDARD_DS3_v2", min_instances=0, max_instances=2, idle_time_before_scale_down=120, tier="low_priority",) ml_client.begin_create_or_update(cluster_low_pri).result()

upvoted 7 times

Mal42 Highly Voted 1 year, 12 months ago

On exam 18 Aug 2023

upvoted 6 times

Question #91 *Topic 2*

HOTSPOT

-

You manage an Azure Machine Learning workspace. You create an experiment named experiment by using the Azure Machine Learning Python SDK v2 and MLflow.

You are reviewing the results of experiment by using the following code segment:

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

```
runs = mlflow.search_runs(
    experiment_names=["experiment1"],
    max_results=5,
    order_by=["start_time ASC"])

runs[runs.status == "FAILED"]
```

Answer Area

Statements	Yes	No
Aborted runs are returned.	○	○
The latest five experiment runs are returned.	○	○
The jobs that are returned have been canceled or killed by the user or system.	○	○
All metrics and their values are returned for the returned experiment runs.	○	○

Suggested Answer:

Answer Area

Statements	Yes	No
Aborted runs are returned.	○	○
The latest five experiment runs are returned.	○	○
The jobs that are returned have been canceled or killed by the user or system.	○	○
All metrics and their values are returned for the returned experiment runs.	○	○

Answer Area

Statements	Yes	No
Aborted runs are returned.	○	○
The latest five experiment runs are returned.	○	○
The jobs that are returned have been canceled or killed by the user or system.	○	○
All metrics and their values are returned for the returned experiment runs.	○	○

gunn_m Most Recent 1 year, 8 months ago

Is Killed by system'different from 'Failed'?

upvoted 1 times

og44 2 years, 5 months ago

Shouldn't be the first one YES. As runs[runs.status=="Failed"] will return aborted runs. is it about the atributes.status? https://learn.microsoft.com/en-us/azure/machine-learning/how-to-track-experiments-mlflow?view=azureml-api-2#getting-params-and-metrics-from-a-runAborted means cancelled by the user explicitly, not to be confused with "failed" jobs.

upvoted 1 times

og44 2 years, 5 months ago

Shouldn't be the first one YES. As runs[runs.status=="Failed"] will return aborted runs. is it about the atributes.status?

https://learn.microsoft.com/en-us/azure/machine-learning/how-to-track-experiments-mlflow?view=azureml-api-2#getting-params-and-metrics-from-a-run

upvoted 1 times

zainkhazi 2 years, 5 months ago

Aborted means cancelled by the user explicitly, not to be confused with "failed" jobs.

upvoted 1 times

Mal42 2 years, 12 months ago

On exam 18 Aug 2023

upvoted 3 times

ABosco 3 years ago

All NOs is correct.

upvoted 4 times

Batman160591 3 years, 1 month ago

NO,YES,NO,NO

upvoted 1 times

Batman160591 3 years, 1 month ago

NO,YES,NO,NOorder by is ASC based on time so the 5 returned will be oldest not latest. So all will be NO

upvoted 1 times

mkk888 Highly Voted 3 years, 1 month ago

order by is ASC based on time so the 5 returned will be oldest not latest. So all will be NO

upvoted 9 times

Question #92 *Topic 2*

You manage an Azure Machine Learning workspace. You have an environment for training jobs which uses an existing Docker image.
A new version of the Docker image is available.
You need to use the latest version of the Docker image for the environment configuration by using the Azure Machine Learning SDK v2.
What should you do?

A. Modify the conda_file to specify the new version of the Docker image.

B. Use the Environment class to create a new version of the environment.

C. Use the create_or_update method to change the tag of the image.

D. Change the description parameter of the environment configuration.

Suggested Answer: *A*
Official Answer: A

Community vote distribution
B (73%) C (27%)

evangelist Most Recent 1 year, 8 months ago
B is the correct answer
upvoted 2 times

sl_mslconsulting 1 year, 9 months ago
Link: https://learn.microsoft.com/en-us/azure/machine-learning/how-to-manage-environments-v2?view=azureml-api-2&tabs=python#create-an-environment-from-a-docker-image Code: env_docker_image = Environment(image="pytorch/pytorch:latest", name="docker-image-example", description="Environment created from a Docker image.",) ml_client.environments.create_or_update(env_docker_image)
upvoted 3 times

deyoz 1 year, 11 months ago
For environments, only description and tags can be updated. All other properties are immutable; if you need to change any of those properties you should create a new version of the environment. https://learn.microsoft.com/en-us/azure/machine-learning/how-to-manage-environments-v2?view=azureml-api-2&tabs=python
upvoted 1 times

dporwal 2 years, 1 month ago
B is the correct answer
upvoted 1 times

NullVoider_0 2 years, 2 months ago
Selected Answer: C
Since the environment already uses a Docker image, the proper way to update to the latest image version is to call the create_or_update() method on that environment configuration object. This allows changing the image tag/version to point to the new Docker image.

Please change the answer to C it's misleading.
upvoted 1 times

NullVoider_0 2 years, 2 months ago
Since the environment already uses a Docker image, the proper way to update to the latest image version is to call the create_or_update() method on that environment configuration object. This allows changing the image tag/version to point to the new Docker image. Please change the answer to C it's misleading.For environments, only description and tags can be updated. All other properties are immutable; if you need to change any of those properties you should create a new version of the environment => https://learn.microsoft.com/en-us/azure/machine-learning/how-to-manage-environments-v2?view=azureml-api-2&tabs=python
upvoted 1 times

Sara_E 2 years ago
For environments, only description and tags can be updated. All other properties are immutable; if you need to change any of those properties you should create a new version of the environment => https://learn.microsoft.com/en-us/azure/machine-learning/how-to-manage-environments-v2?view=azureml-api-2&tabs=python
upvoted 2 times

Mal42 2 years, 5 months ago

On exam 18 Aug 2023

upvoted 3 times

Ahmed_Gehad 2 years, 6 months ago

We set the image in the Environment class as follows from azure.ai.ml.entities import Environment env_docker_image = Environment(image="mcr.microsoft.com/azureml/openmpi3.1.2-ubuntu18.04", name="aml-docker-image-example", description="Environment created from a Azure ML Docker image.",) ml_client.environments.create_or_update(env_docker_image)seem ans should be both B and C

upvoted 4 times

Ahmed_Gehad 2 years, 6 months ago

Selected Answer: B

We set the image in the Environment class as follows

from azure.ai.ml.entities import Environment

env_docker_image = Environment(image="mcr.microsoft.com/azureml/openmpi3.1.2-ubuntu18.04", name="aml-docker-image-example", description="Environment created from a Azure ML Docker image.",) ml_client.environments.create_or_update(env_docker_image)

upvoted 4 times

hiyoww 2 years, 1 month ago

seem ans should be both B and C

upvoted 1 times

damaldon 2 years, 7 months ago

Correct. You need to create a new version using the Environment class. env_docker_image = Environment(image="pytorch/pytorch:latest", name="docker-image-example", description="Environment created from a Docker image.",) ml_client.environments.create_or_update(env_docker_image)

upvoted 3 times

pranav33 2 years, 7 months ago

B looks correct

upvoted 3 times

Batman160591 2 years, 8 months ago

B. Use the Environment class to create a new version of the environment. To update the Docker image, you can create a new version of the environment by using the Environment class provided by the Azure Machine Learning SDK. You can update the Docker image reference in the environment's Docker section to use the latest version.

upvoted 4 times

Question #93 *Topic 2*

HOTSPOT

-

You manage an Azure Machine Learning workspace by using the Python SDK v2.

You must create an automated machine learning job to generate a classification model by using data files stored in Parquet format.

You must configure an autoscaling compute target and a data asset for the job.

You need to configure the resources for the job.

Which resource configuration should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Resource type	Resource value
Compute target	Azure Data Lake Analytics Azure HDInsight Azure Databricks
Data asset	mltable uri_file uri folder

Suggested Answer:

Answer Area

Resource type	Resource value
Compute target	Azure Data Lake Analytics Azure HDInsight **Azure Databricks**
Data asset	mltable uri_file **uri folder**

Answer Area

Resource type	Resource value
Compute target	Azure Data Lake Analytics Azure HDInsight **Azure Databricks**
Data asset	mltable uri_file **uri folder**

f11c733 Highly Voted 2 years, 2 months ago

The correct answers are Azure Databricks and mltable.

upvoted 6 times

LadyCasilda Highly Voted 3 years ago

On exam 18 August 2023

upvoted 5 times

Plb2 2 years, 5 months ago

https://learn.microsoft.com/en-us/azure/machine-learning/how-to-mltable?view=azureml-api-2&tabs=cli Azure Machine Learning doesn't require use of Azure Machine Learning Tables (mltable) for tabular data. You can use Azure Machine Learning File (uri_file) and Folder (uri_folder) types, and your own parsing logic loads the data into a Pandas or Spark data frame. If you have a simple CSV file or Parquet folder, it's easier to use Azure Machine Learning Files/Folders instead of Tables.

upvoted 2 times

Lion007 2 years, 7 months ago

Correct: Azure Databricks and uri_folder Compute target: Azure Databricks This is because Azure Databricks supports autoscaling of workers required to run your job, which dynamically reallocates workers to match the computational demands of your job, thereby achieving high cluster utilization without the need for provisioning the cluster to match a specific workload. Data asset: uri_folder This option allows the machine learning job to access all the Parquet files stored in the specified directory. If you have multiple Parquet data files, you would use a URI that points to a folder containing all these files.

upvoted 4 times

damaldon 3 years, 1 month ago

Correct. uri_folder Read a folder of parquet/CSV files into Pandas/Spark.

upvoted 4 times

Batman160591 3 years, 1 month ago

Seems correct:)

upvoted 1 times

Question #94 *Topic 2*

HOTSPOT

-

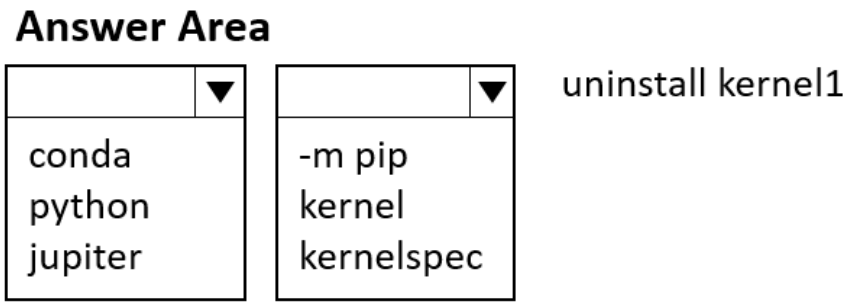
You manage an Azure Machine Learning workspace named workspace1 with a compute instance named compute1.

You must remove a kernel named kernel1 from compute1. You connect to compute1 by using a terminal window from workspace1.

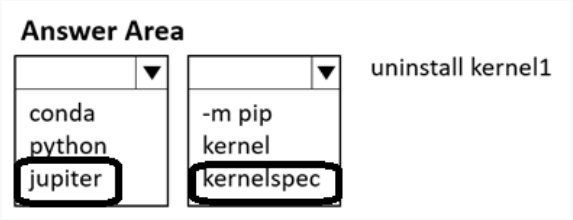
You need to enter a command in the terminal window to remove kernel.

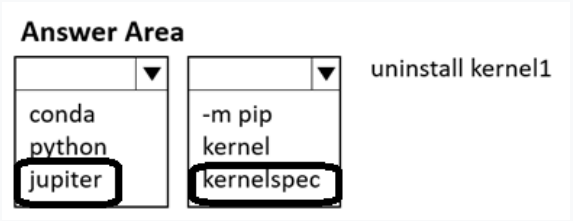
Which command should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



Suggested Answer:





Batman160591 Highly Voted 2 years, 1 month ago

Answer: jupyter kernelspec uninstall kernel1 This command uninstalls the specified kernel from Jupyter, which is the underlying technology used for compute instances in Azure Machine Learning. By running this command, you will remove the "kernel1" from the list of available kernels on the compute instance. Please make sure you have the necessary permissions to execute the command on the compute instance.

upvoted 6 times

ymj_000 Most Recent 1 year, 9 months ago

For reference: https://learn.microsoft.com/en-us/azure/machine-learning/how-to-access-terminal?view=azureml-api-2

upvoted 3 times

A_PL300 1 year, 11 months ago

on exam 4 September 2023

upvoted 4 times

LadyCasilda 1 year, 12 months ago

On exam 18 August 2023

upvoted 4 times

Question #95 *Topic 2*

You manage an Azure Machine Learning workspace. The workspace includes an Azure Machine Learning Kubernetes compute target configured as an Azure Kubernetes Service (AKS) cluster named AKS1. AKS1 is configured to enable the targeting of different nodes to train workloads.
You must run a command job on AKS1 by using the Azure ML Python SDK v2. The command job must select different types of compute nodes.
The compute node types must be specified by using a command parameter.
You need to configure the command parameter.
Which parameter should you use?

A. environment

B. compute

C. limits

D. instance_type

Suggested Answer: *D*
Official Answer: D

Community vote distribution
D (64%) B (36%)

Norasit Highly Voted 2 years, 7 months ago
D. instance_type
upvoted 9 times

LadyCasilda Highly Voted 2 years, 6 months ago
On exam 18 August 2023
upvoted 6 times

evangelist Most Recent 1 year, 8 months ago
compute node type= instance type
upvoted 1 times

sl_mslconsulting 1 year, 9 months ago
AKS is the compute target you want to use. instance_type is about the size of the VM you are using. command_node.set_resources(instance_count=1, instance_type="STANDARD_D2_v2", properties={"key": "new_val"}, shm_size="3g",)
upvoted 1 times

Plb2 1 year, 11 months ago
With the parameter 'instance_type' on the method command.set_resources the type of compute instance to run the job on can be configured. If not specified, the job will run on the default compute target. https://learn.microsoft.com/en-us/python/api/azure-ai-ml/azure.ai.ml.entities.command?view=azure-python#azure-ai-ml-entities-command-set-resources
upvoted 1 times

NullVoider_0 2 years, 1 month ago
B. compute: The 'compute' parameter is used to specify the compute target on which the job will run. In the context of Azure ML and AKS, you can use this parameter to specify different node types in your AKS cluster. This is done by referencing the specific AKS compute targets that are configured for different node types.
upvoted 1 times

Ahmed_Gehad 2 years, 6 months ago
The instance_type is optional, which is the type of VM used as supported by the compute target. The compute is required and is the name of the compute where the command job is executed.
upvoted 6 times

cloud6190 2 years, 6 months ago
B. Compute is the right answer.
upvoted 2 times

damaldon 2 years, 7 months ago
Correct. from azure.ai.ml import command # define the command command_job = command(command="python -c "print('Hello world!')"", environment="AzureML-lightgbm-3.2-ubuntu18.04-py37-cpu@latest", compute=" ", instance_type=" ")

upvoted 4 times

Batman160591 2 years, 8 months ago

To configure the command parameter to select different types of compute nodes when running a command job on an Azure Machine Learning Kubernetes compute target (AKS) using the Azure ML Python SDK v2, you should use: B. compute The "compute" parameter allows you to specify the target compute resource for the command job. In this case, you want to target different types of compute nodes. By setting the "compute" parameter to the desired node type, you can instruct the command job to use the specific compute nodes for the job execution.

upvoted 3 times

HOTSPOT

-

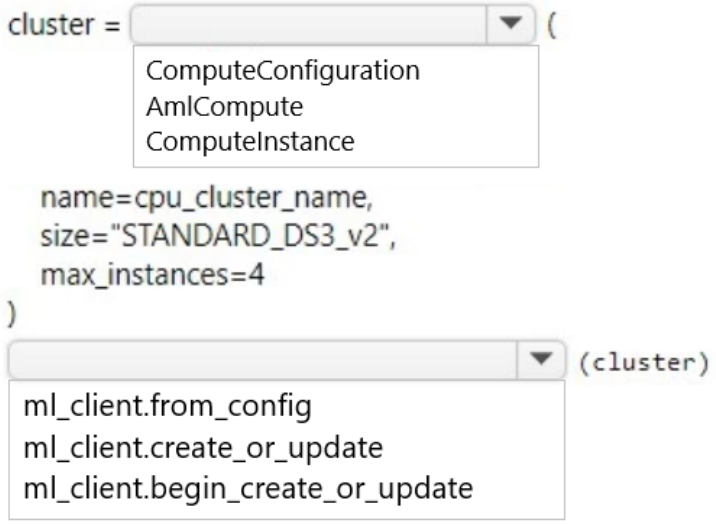
You create a new Azure Machine Learning workspace with a compute cluster.

You need to create the compute cluster asynchronously by using the Azure Machine Learning Python SDK v2.

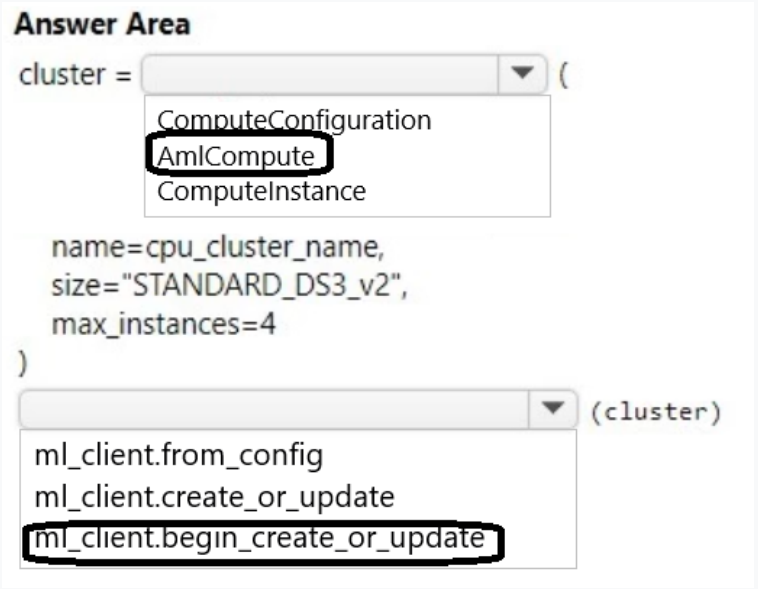
How should you complete the code segment? To answer, select the appropriate options in the answer area.

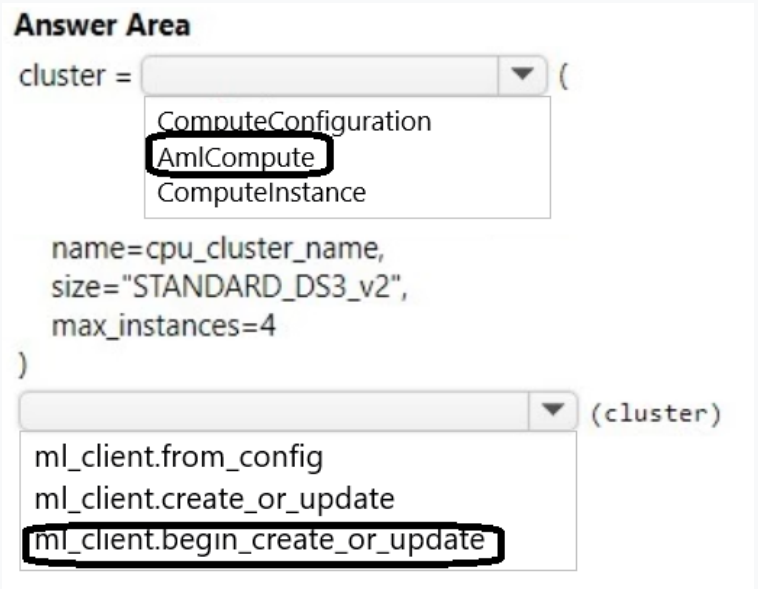
NOTE: Each correct selection is worth one point.

Answer Area



Suggested Answer:





damaldon Most Recent 2 years, 1 month ago

Correct. begin_create_or_update(resource_group_name: str, domain_name: str, domain_info: '_models.Domain', **kwargs: Any) → LROPoller['_models.Domain'][source] Create or update a domain. Asynchronously creates or updates a new domain with the specified parameters.

upvoted 3 times

Question #97 *Topic 2*

DRAG DROP

-

You manage an Azure Machine Learning workspace named workspace1 by using the Python SDK v2.

You must register datastores in workspace1 for Azure Blob storage and Azure Files storage to meet the following requirements:

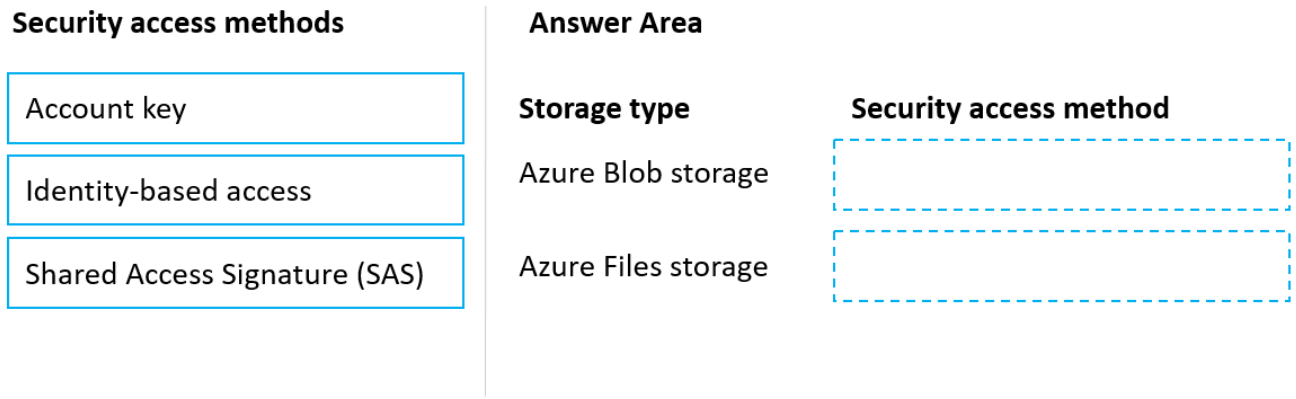
• Azure Active Directory (Azure AD) authentication must be used for access to storage when possible.

• Credentials and secrets stored in workspace1 must be valid for a specified time period when accessing Azure Files storage.

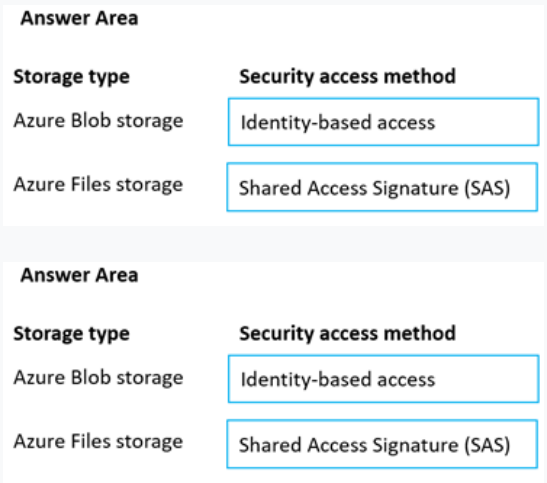
You need to configure a security access method used to register the Azure Blob and Azure Files storage in workspace1.

Which security access method should you configure? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



Suggested Answer:



LadyCasilda 3 years ago

On exam 18 August 2023

upvoted 7 times

A_PL300 2 years, 11 months ago

on exam 4 September 2023

upvoted 5 times

jl420 Most Recent 1 year, 9 months ago

Identity / SAS

upvoted 1 times

phdykd 3 years ago

Account key and SAS token is correct for both. agreed with abuluxury

upvoted 1 times

PI_Team 3 years ago

correct1. Azure Blob storage: Identity-based access (Azure AD authentication) allows you to use Azure AD credentials to authenticate and access Azure Blob storage. This is the recommended and more secure method for accessing Blob storage, as it avoids exposing account keys directly. 2. Azure File Storage: Shared access signature (SAS) is used to provide temporary access to resources in Azure File Storage. When using SAS, you can generate a token with specific permissions and expiration time, which can be used to access the Azure File share. This approach allows you to control the access to the file share for a specified time period without exposing the storage account key. SaMagreed with explanation. Link: https://learn.microsoft.com/en-us/azure/machine-learning/how-to-administrate-data-authentication?view=azureml-api-2#identity-based-data-authentication

upvoted 1 times

PI_Team 3 years ago

correct

upvoted 1 times

PI_Team Highly Voted 3 years ago

1. Azure Blob storage: Identity-based access (Azure AD authentication) allows you to use Azure AD credentials to authenticate and access Azure Blob storage. This is the recommended and more secure method for accessing Blob storage, as it avoids exposing account keys directly.

2. Azure File Storage: Shared access signature (SAS) is used to provide temporary access to resources in Azure File Storage. When using SAS, you can generate a token with specific permissions and expiration time, which can be used to access the Azure File share. This approach allows you to control the access to the file share for a specified time period without exposing the storage account key.

SaM

upvoted 8 times

PI_Team Highly Voted 3 years ago

1. Azure Blob storage: Identity-based access (Azure AD authentication) allows you to use Azure AD credentials to authenticate and access Azure Blob storage. This is the recommended and more secure method for accessing Blob storage, as it avoids exposing account keys directly. 2. Azure File Storage: Shared access signature (SAS) is used to provide temporary access to resources in Azure File Storage. When using SAS, you can generate a token with specific permissions and expiration time, which can be used to access the Azure File share. This approach allows you to control the access to the file share for a specified time period without exposing the storage account key. SaMagreed with explanation. Link: https://learn.microsoft.com/en-us/azure/machine-learning/how-to-administrate-data-authentication?view=azureml-api-2#identity-based-data-authentication

upvoted 8 times

sl_mslconsulting 2 years, 3 months ago

agreed with explanation. Link: https://learn.microsoft.com/en-us/azure/machine-learning/how-to-administrate-data-authentication?view=azureml-api-2#identity-based-data-authentication

upvoted 1 times

abuluxury 3 years ago

Don't see why you can't use sas for both or account key

upvoted 2 times

abuluxury 3 years ago

Don't see why you can't use sas for both or account keybecause it said "Azure Active Directory (Azure AD) authentication must be used for access to storage when possible"

upvoted 2 times

jl420 1 year, 9 months ago

because it said "Azure Active Directory (Azure AD) authentication must be used for access to storage when possible"

upvoted 1 times

Batman160591 3 years, 1 month ago

seems correct

upvoted 2 times

Question #98 *Topic 2*

You manage an Azure Machine Learning workspace named workspace1.
You must develop Python SDK v2 code to add a compute instance to workspace1. The code must import all required modules and call the constructor of the ComputeInstance class.
You need to add the instantiated compute instance to workspace1.
What should you use?

A. constructor of the azure.ai.ml.ComputeSchedule class

B. constructor of the azure.ai.ml.ComputePowerAction enum

C. begin_create_or_update method of an instance of the azure.ai.ml.MLClient class

D. set_resources method of an instance of the azure.ai.ml.Command class

Suggested Answer: *C*
Official Answer: C

Community vote distribution
C (100%)

evangelist Most Recent 1 year, 8 months ago
begin_create_or_update create instance asynchrounously, given answer is correct
upvoted 1 times

mackcsc 2 years, 4 months ago
Correct
upvoted 2 times

Batman160591 2 years, 8 months ago
Seems correct. To add a compute instance to an Azure Machine Learning workspace using the Python SDK v2, you should use: C. begin_create_or_update method of an instance of the azure.ai.ml.MLClient class
upvoted 2 times

Question #99 *Topic 2*

HOTSPOT

-

You manage an Azure Machine Learning workspace.

You must define the execution environments for your jobs and encapsulate the dependencies for your code.

You need to configure the environment from a Docker build context.

How should you complete the code segment? To answer, select the appropriate option in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

```
docker_context = [ ▼ ](
                   Environment
                   Job
                   Command
                   Component
    [ ▼ ]=BuildContext(path="docker-contexts/python-and-pip"),
    build
    image
    datastore
    properties
    name="docker-context "
)
ml_client.environments.create_or_update(docker_context)
```

Suggested Answer:

Answer Area

```
docker_context = [ ▼ ](
                   Environment
                   Job
                   Command
                   Component
    [ ▼ ]=BuildContext(path="docker-contexts/python-and-pip"),
    build
    image
    datastore
    properties
    name="docker-context "
)
ml_client.environments.create_or_update(docker_context)
```

Answer Area

```
docker_context = [ ▼ ](
                   Environment
                   Job
                   Command
                   Component
    [ ▼ ]=BuildContext(path="docker-contexts/python-and-pip"),
    build
    image
    datastore
    properties
    name="docker-context "
)
ml_client.environments.create_or_update(docker_context)
```

Kanwal001 Highly Voted 2 years, 5 months ago

On exam 28 Aug 2023

upvoted 6 times

PI_Team Highly Voted 2 years, 6 months ago

Correct. from azureml.core import Environment from azureml.core.environment import BuildContext env_docker_context = Environment(build=BuildContext(path="docker-contexts/python-and-pip"), name="docker-context-example", description="Environment created from a Docker context.",) SaM

upvoted 6 times

TA_ Most Recent 1 year, 10 months ago

On exam 15-03-2024

upvoted 1 times

Batman160591 2 years, 8 months ago

seems correct

upvoted 1 times

Question #100 *Topic 2*

HOTSPOT

-

You create an Azure Machine Learning workspace. You use the Azure Machine Learning Python SDK v2 to create a compute cluster.

The compute cluster must run a training script. Costs associated with running the training script must be minimized.

You need to complete the Python script to create the compute cluster.

How should you complete the script? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

```
from azure.ai.ml.entities import AmlCompute
try:
   cpu_cluster = ml_client.compute.get("cpu-cluster")
except Exception:
   cpu_cluster = [AmlCompute / ComputeInstance / KubernetesCompute] (
      name="cpu-cluster",
      size="STANDARD_DS3_V2",
      max_instances=4,
      [tier="LowPriority" / min_instances=0 / min_instances=1] ,
      )
      cpu_cluster =
ml_client.begin_create_or_update(cpu_cluster)
)
```

Suggested Answer:

Answer Area

```
from azure.ai.ml.entities import AmlCompute
try:
   cpu_cluster = ml_client.compute.get("cpu-cluster")
except Exception:
   cpu_cluster = [AmlCompute | ComputeInstance | KubernetesCompute] (
      name="cpu-cluster",
      size="STANDARD_DS3_V2",
      max_instances=4,
      [tier="LowPriority" | min_instances=0 | min_instances=1],
      )
      cpu_cluster =
ml_client.begin_create_or_update(cpu_cluster)
)
```



haby Highly Voted 2 years, 2 months ago

I think should be tier="LowPriority" since min_instances has default value 0. If dont set min_instances value, it will be 0 by default.

upvoted 8 times

sl_mslconsulting 1 year, 9 months ago

should be tier="LowPriority". Setting min_instances to 0 won't help you cut down the cost if you are actively training a model.Besides it's asking "Costs associated with running the training script must be minimized." so the nodes won't be idle.

upvoted 2 times

sl_mslconsulting 1 year, 9 months ago

should be tier="LowPriority". Setting min_instances to 0 won't help you cut down the cost if you are actively training a model.

upvoted 2 times

sl_mslconsulting Most Recent 1 year, 9 months ago

Besides it's asking "Costs associated with running the training script must be minimized." so the nodes won't be idle.

upvoted 1 times

vv_bb 2 years, 3 months ago

The correct answer is: - AmlCompute - tier="LowPriority"While there is no doubts what the answer should be for the first part, let me explain the answer for the second. 1) According to the question, we should minimize the cost "...associated with running the training script...". It means the goal is to minimize the costs exactly for training, not for the rest of the time. Why? Well, because for example the cluster may even be deleted immediately after the training completed, etc. We can't be sure why we've being asked not to consider epy сщыеы other than necessary for training, but that's what we've being asked for. And during the training, cluster eventually use all the possible nodes, so defining the [min_instances = 0] will not affect the costs for training at all.2) According to the page below, if you'll not define the min_instances, then the default value = 0 will be used. So actually there is even no need to explicitly define [min_instances = 0] While, if you're not defining tier, then the default value = 'dedicated' will be used. https://learn.microsoft.com/en-us/azure/machine-learning/reference-yaml-compute-aml?view=azureml-api-2

upvoted 4 times

vv_bb 2 years, 3 months ago

The correct answer is: - AmlCompute - tier="LowPriority"

upvoted 4 times

vv_bb 2 years, 3 months ago

While there is no doubts what the answer should be for the first part, let me explain the answer for the second. 1) According to the question, we should minimize the cost "...associated with running the training script...". It means the goal is to minimize the costs exactly for training, not for the rest of the time. Why? Well, because for example the cluster may even be deleted immediately after the training completed, etc. We can't be sure why we've being asked not to consider epy сщыеы other than necessary for training, but that's what we've being asked for. And during the training, cluster eventually use all the possible nodes, so defining the [min_instances = 0] will not affect the costs for training at all.2) According to the page below, if you'll not define the min_instances, then the default value = 0 will be used. So actually there is even no need to explicitly define [min_instances = 0] While, if you're not defining tier, then the default value = 'dedicated' will be used. https://learn.microsoft.com/en-us/azure/machine-learning/reference-yaml-compute-aml?view=azureml-api-2

upvoted 1 times

vv_bb 2 years, 3 months ago

While there is no doubts what the answer should be for the first part, let me explain the answer for the second.

1) According to the question, we should minimize the cost "...associated with running the training script...". It means the goal is to minimize the costs exactly for training, not for the rest of the time. Why? Well, because for example the cluster may even be deleted immediately after the training completed, etc. We can't be sure why we've being asked not to consider epy сщыеы other than necessary for training, but that's what we've being asked for.

And during the training, cluster eventually use all the possible nodes, so defining the [min_instances = 0] will not affect the costs for training at all.

upvoted 1 times

vv_bb 2 years, 3 months ago

2) According to the page below, if you'll not define the min_instances, then the default value = 0 will be used. So actually there is even no need to explicitly define [min_instances = 0] While, if you're not defining tier, then the default value = 'dedicated' will be used. https://learn.microsoft.com/en-us/azure/machine-learning/reference-yaml-compute-aml?view=azureml-api-2

upvoted 3 times

damaldon 2 years, 7 months ago

Correct.

upvoted 2 times

Batman160591 2 years, 8 months ago

seems correct

upvoted 1 times

Question #101 *Topic 2*

You create a workspace to include a compute instance by using Azure Machine Learning Studio. You are developing a Python SDK v2 notebook in the workspace.
You need to use Intellisense in the notebook.
What should you do?

A. Stop the compute instance.

B. Start the compute instance.

C. Run a %pip magic function on the compute instance.

D. Run a !pip magic function on the compute instance.

Suggested Answer: *C*
Official Answer: C

Community vote distribution
B (100%)

Ahmed_Gehad Highly Voted 2 years, 6 months ago
- Option A (Stop the compute instance) will not enable Intellisense. Stopping the compute instance will shut it down, and you won't be able to access its resources until you start it again. - Option C (Run a %pip magic function on the compute instance) and Option D (Run a !pip magic function on the compute instance) are not related to enabling Intellisense. These magic functions are used to install Python packages in the notebook's environment. While installing packages may be necessary for code execution, it won't impact the availability of Intellisense.
upvoted 5 times

smaha_18 Most Recent 8 months ago
https://learn.microsoft.com/en-us/azure/machine-learning/how-to-run-jupyter-notebooks?view=azureml-api-2. Document clearly says that when a compute instance is running, you can also use code completion, powered by Intellisense, in any Python notebook.
upvoted 1 times

evangelist 1 year, 8 months ago
Correct is B, intellisense needs a running compute instance to support this feature
upvoted 1 times

jojashi 1 year, 8 months ago
Correct. C Chat GPT says it is C. To enable Intellisense in a Python SDK v2 notebook within an Azure Machine Learning Studio workspace, you don't need to perform any actions related to starting or stopping the compute instance. Instead, you need to ensure that the required Python packages are installed in your environment to support Intellisense.
upvoted 1 times

Matt2000 2 years ago
B seems to be correct. See https://learn.microsoft.com/en-us/azure/machine-learning/how-to-run-jupyter-notebooks?view=azureml-api-2
upvoted 2 times

ferren 2 years, 3 months ago
claude says it is B
upvoted 1 times

cyberfriends 2 years, 4 months ago
You can use %pip install to install packages and enable Intellisense. For example, if you want to install a package like numpy, you can run the following command in your Jupyter notebook cell: %p ip install numpy
upvoted 1 times

PradhanManva 2 years, 4 months ago
B is correct
upvoted 2 times

ABosco 2 years, 6 months ago
B is correct When a compute instance is running, you can also use code completion, powered by Intellisense, in any Python notebook.
upvoted 2 times

PI_Team 2 years, 6 months ago

To use Intellisense in a Python SDK v2 notebook in Azure Machine Learning Studio, you should start the compute instance. This will enable you to use advanced features like full IntelliSense and inline error highlighting directly in your Jupyter notebooks1. So the correct answer is B. Start the compute instance. SaM

upvoted 2 times

phdykd 2 years, 7 months ago

To use it, you must ensure that your compute instance is running and connected to your notebook. the correct answer is: B. Start the compute instance.

upvoted 1 times

fhlos 2 years, 7 months ago

C - is correct answer

upvoted 1 times

Batman160591 2 years, 8 months ago

B seems the correct

upvoted 2 times

Question #102 *Topic 2*

DRAG DROP

-

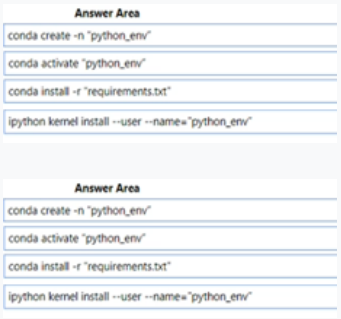
You manage an Azure Machine Learning workspace named workspace1 with a compute instance named compute1. You connect to compute1 by using a terminal window from workspace1. You create a file named "requirements.txt" containing Python dependencies to include Jupyter.

You need to add a new Jupyter kernel to compute1.

Which four commands should you use? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Suggested Answer:



evangelist Most Recent 1 year, 8 months ago

given answer is correct

upvoted 1 times

jojashi 1 year, 8 months ago

Correct. create new environment --> activate new environment --> install packages to the new environment --> create a kernel for that environment https://learn.microsoft.com/en-us/azure/machine-learning/how-to-access-terminal?view=azureml-api-2

upvoted 2 times

damaldon 2 years, 7 months ago

Correct

upvoted 3 times

phdykd 2 years, 7 months ago

no, given answer is correct!

upvoted 4 times

Batman160591 2 years, 8 months ago

1. ipykernel install --user --name myenv --display-name "Python (myenv)" 2. conda activate myenv 3. pip install -r requirements.txt 4. python -m ipykernel install --user --name myenv --display-name "Python (myenv)"

upvoted 1 times

Batman160591 2 years, 8 months ago

1. ipykernel install --user --name myenv --display-name "Python (myenv)" 2. conda activate myenv 3. pip install -r requirements.txt 4. python -m ipykernel install --user --name myenv --display-name "Python (myenv)"yes https://learn.microsoft.com/en-us/azure/machine-learning/how-to-access-terminal?view=azureml-api-2#add-new-kernels

upvoted 1 times

Plb2 1 year, 11 months ago

yes https://learn.microsoft.com/en-us/azure/machine-learning/how-to-access-terminal?view=azureml-api-2#add-new-kernels

upvoted 1 times

Question #103 *Topic 2*

HOTSPOT

-

You are creating data wrangling and model training solutions in an Azure Machine Learning workspace.

You must use the same Python notebook to perform both data wrangling and model training.

You need to use the Azure Machine Learning Python SDK v2 to define and configure the Synapse Spark pool asynchronously in the workspace as dedicated compute.

How should you complete the code segment? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

```
synapse_compute = [ ▼ ] (name=synapse_name,
```

- SparkComponent
- SynapseSparkCompute
- Spark
- ResourceConfiguration

```
resource_id=synapse_resource)
[ ▼ ] (synapse_compute)
```

- ml_client.from_config
- ml_client.create_or_update
- ml_client.begin_create_or_update

Suggested Answer:

Answer Area

```
synapse_compute = [ ▼ ] (name=synapse_name,
```

- SparkComponent
- **SynapseSparkCompute**
- Spark
- ResourceConfiguration

```
resource_id=synapse_resource)
[ ▼ ] (synapse_compute)
```

- ml_client.from_config
- ml_client.create_or_update
- **ml_client.begin_create_or_update**

Answer Area

```
synapse_compute = [ ▼ ] (name=synapse_name,
```

- SparkComponent
- **SynapseSparkCompute**
- Spark
- ResourceConfiguration

```
resource_id=synapse_resource)
[ ▼ ] (synapse_compute)
```

- ml_client.from_config
- ml_client.create_or_update
- **ml_client.begin_create_or_update**

sai384957324 Most Recent 1 year, 10 months ago

correct

upvoted 1 times

TA_ 1 year, 10 months ago

This or very similar on exam 15-03-2024. Several questions related to Synapse Spark

upvoted 1 times

damaldon 2 years, 7 months ago

Correct.

upvoted 2 times

Batman160591 2 years, 8 months ago

Seems correct

upvoted 1 times

HOTSPOT

-

You manage an Azure Machine Learning workspace named workspace1 by using the Python SDK v2.

You must register datastores in workspace1 for Azure Blob and Azure Data Lake Gen2 storage to meet the following requirements:

• Data scientists accessing the datastore must have the same level of access.

• Access must be restricted to specified containers or folders.

You need to configure a security access method used to register the Azure Blob and Azure Data Lake Gen2 storage in workspace1.

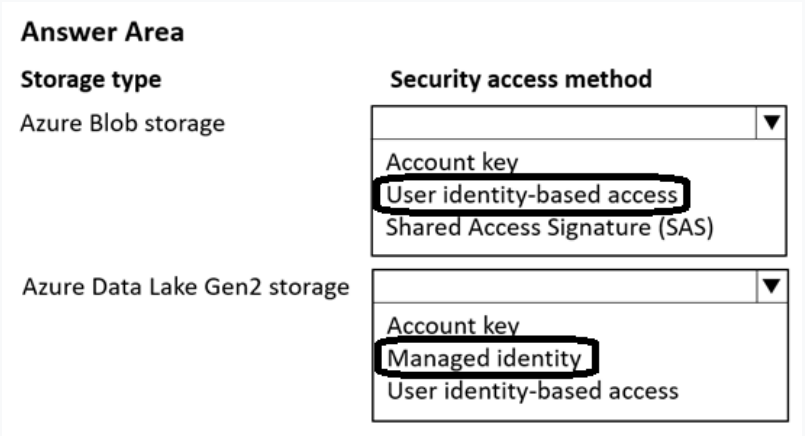
Which security access method should you configure? To answer, select the appropriate options in the answer area.

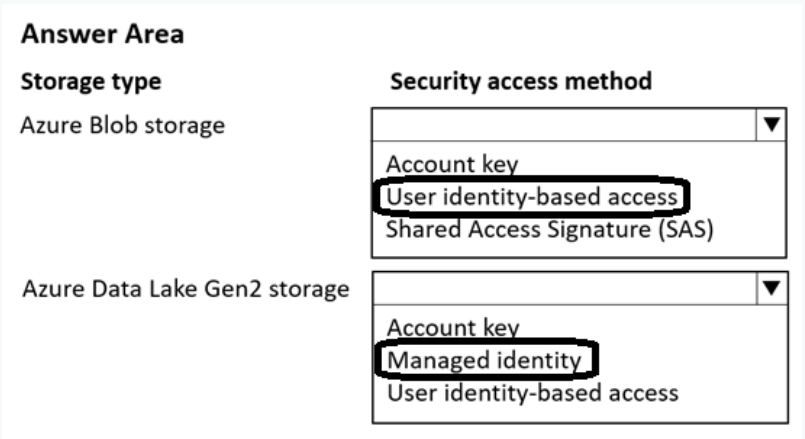
NOTE: Each correct selection is worth one point.

Answer Area

Storage type	Security access method
Azure Blob storage	Account key User identity-based access Shared Access Signature (SAS)
Azure Data Lake Gen2 storage	Account key Managed identity User identity-based access

Suggested Answer:





PI_Team Highly Voted 3 years ago

Both should be User identity-based access User identity-based access allows you to restrict access to the datastore to specific users or groups. This is the best option for your scenario because it ensures that all data scientists accessing the datastore have the same level of access, and it allows you to restrict access to specified containers or folders. Account key and SAS tokens are both credential-based access methods. This means that the credentials for accessing the datastore are stored in the Azure Machine Learning workspace. This is not ideal for your scenario because it would allow any user with Reader workspace access to access the datastore. SaM

upvoted 7 times

Batman160591 Highly Voted 3 years, 1 month ago

To meet the requirements of providing the same level of access to data scientists and restricting access to specified containers or folders when registering Azure Data Lake Gen2 storage in an Azure Machine Learning workspace using the Python SDK v2, you should configure the security access method as: C. User identity-based access By using user identity-based access, you can leverage Azure Active Directory (Azure AD) to control access to the Azure Data Lake Gen2 storage. This approach allows you to assign appropriate roles and permissions to individual users or groups in Azure AD, providing granular access control at the user level. With user identity-based access, you can ensure that only authorized users have access to the specified containers or folders within the Azure Data Lake Gen2 storage. This level of security is essential when you want to restrict access to sensitive data or ensure compliance with data governance policies. On the other hand, the options "Account key" (A) and "Managed identity" (B) do not meet the requirement of providing the same level of access to data scientists or restricting access to specified containers or folders.

upvoted 5 times

Fefnut Most Recent 1 year, 9 months ago

- Azure Blob storage: SAS since User identity-based access does not give access to the actual data https://learn.microsoft.com/en-us/azure/storage/blobs/sas-service-create-python?tabs=container. SAS keys can be shared - Azure Data Lake Gen2: Managed identity since you can create groups to access specific data https://learn.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-access-control#access-control-lists-on-files-and-directories

upvoted 1 times

deyoz 2 years, 5 months ago

I do agree with below answers in discussions

upvoted 1 times

Question #105 *Topic 2*

DRAG DROP

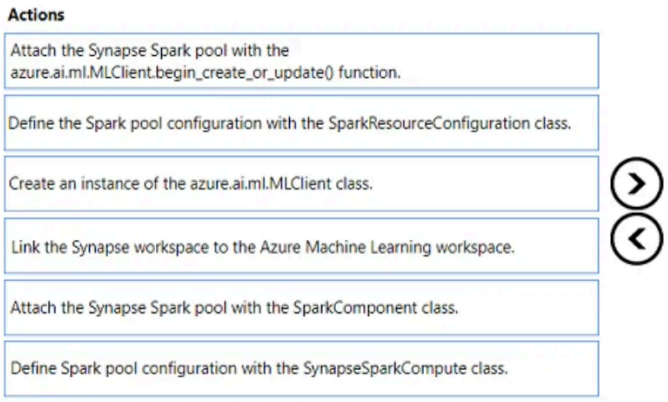
-

You create an Azure Machine Learning workspace and an Azure Synapse Analytics workspace with a Spark pool. The workspaces are contained within the same Azure subscription.

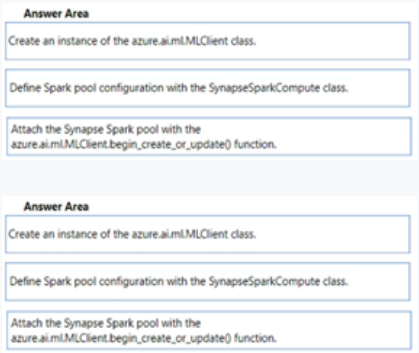
You must manage the Synapse Spark pool from the Azure Machine Learning workspace.

You need to attach the Synapse Spark pool in Azure Machine Learning by using the Python SDK v2.

Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Suggested Answer:



TA_ 2 years, 4 months ago

On exam 15-03-2024

upvoted 1 times

Borjavvvvv 2 years, 8 months ago

To attach a Synapse Compute using Python SDK, first create an instance of azure.ai.ml.MLClient class. This provides convenient functions for interaction with Azure Machine Learning services. The following code sample uses azure.identity.DefaultAzureCredential for connecting to a workspace in resource group of a specified Azure subscription. In the following code sample, define the SynapseSparkCompute with the parameters: name - user-defined name of the new attached Synapse Spark pool. resource_id - resource ID of the Synapse Spark pool created earlier in the Azure Synapse Analytics workspace. An azure.ai.ml.MLClient.begin_create_or_update() function call attaches the defined Synapse Spark pool to the Azure Machine Learning workspace.

upvoted 4 times

Mikku123 3 years ago

My Bad - The correct sequence is: Create an instance of the azure.ai.MLClient class Define the spark pool configuration with the SparkResourceConfiguration class Attach the Synapse Spark pool with the azure.ai.ml.MLClient.begin_create_or_update() functionNot SparkResourceConfiguration calss but SynapseSparkCompute class.

upvoted 3 times

Mikku123 3 years ago

My Bad - The correct sequence is:

Create an instance of the azure.ai.MLClient class Define the spark pool configuration with the SparkResourceConfiguration class Attach the Synapse Spark pool with the azure.ai.ml.MLClient.begin_create_or_update() function

upvoted 3 times

Sadhak Most Recent 1 year, 9 months ago

Not SparkResourceConfiguration calss but SynapseSparkCompute class.

upvoted 1 times

Mikku123 3 years ago

correct answer!

upvoted 3 times

Question #106 *Topic 2*

You manage an Azure Machine Learning workspace.
You need to define an environment from a Docker image by using the Azure Machine Learning Python SDK v2.
Which parameter should you use?

A. properties

B. image

C. build

D. conda_file

Suggested Answer: *B*
Official Answer: B

Community vote distribution
B (67%) C (33%)

PI_Team Highly Voted 3 years ago
Correct. The main difference between using the build and image parameters when defining an environment using the Azure Machine Learning Python SDK v2 is whether you want to define the environment from a Dockerfile or a Docker image. If you want to define an environment from a Dockerfile, you should use the build parameter and set it to a dictionary that specifies the contents of the Dockerfile using the dockerfile key. If you want to define an environment from a Docker image, you should use the image parameter and set it to the name of the Docker image you want to use. SaM
upvoted 17 times

dreamlimit666 Most Recent 1 year, 9 months ago
The correct answer is: B. image When defining an environment from a Docker image using the Azure Machine Learning Python SDK v2, the image parameter should be used. This parameter specifies the Docker image that will be used as the base environment for running the experiment.
upvoted 2 times

kay1101 2 years, 2 months ago
The answer is correct. define from docker image using image parameter. https://learn.microsoft.com/en-us/azure/machine-learning/how-to-manage-environments-v2?view=azureml-api-2&tabs=python
upvoted 1 times

Matt2000 2 years, 6 months ago
image is correct. Reference: https://learn.microsoft.com/en-us/azure/machine-learning/how-to-manage-environments-v2?view=azureml-api-2&tabs=python
upvoted 1 times

Lion007 2 years, 7 months ago
The Correct answer is: C. build Justification: When defining an environment from a Docker image using the Azure Machine Learning Python SDK v2, the appropriate parameter to use is build. This is because in the context of the Azure Machine Learning environment, the build parameter is used to specify the configuration for building the environment, including specifying the Docker image to be used. C. build: This is the correct parameter. Within the Environment class in SDK v2, the build parameter is used to define the build context, which includes specifying the Docker image. When creating an environment from a Docker image, you typically use the BuildContext class within build to specify the Docker image.
upvoted 1 times

damaldon 3 years, 1 month ago
Correct. env_docker_image = Environment(image="pytorch/pytorch:latest", name="docker-image-example", description="Environment created from a Docker image.",) ml_client.environments.create_or_update(env_docker_image)
upvoted 2 times

Question #107 *Topic 2*

You create an Azure Machine Learning managed compute resource. The compute resource is configured as follows:
• Minimum nodes: 2
• Maximum nodes: 4
You must decrease the minimum number of nodes and increase the maximum number of nodes to the following values:
• Minimum nodes: 0
• Maximum nodes: 8
You need to reconfigure the compute resource.
Which three methods can you use? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. Azure Machine Learning designer

B. MLClient class in Python SDK v2

C. Azure Machine Learning studio

D. Azure CLI ml extension v2

E. BuildContext class in Python SDK v2

Suggested Answer: *A, C, D*
Official Answer: ACD

Community vote distribution
BCD (86%) ABD (14%)

umair_hanu Highly Voted 2 years, 7 months ago
BCD should be the answer.
upvoted 7 times

Techlover74 Most Recent 2 years ago
BCD is the answer.
upvoted 1 times

NullVoider_0 2 years, 2 months ago
You cannot scale the instances from the AML Designer.
upvoted 2 times

PI_Team 2 years, 2 months ago
BCD should be the answer.
upvoted 1 times

Hisayuki 2 years, 3 months ago
The compute autoscales down to zero nodes when it isn't used. Dedicated VMs are created to run your jobs as needed. Use the following examples to create a compute cluster: - Python SDK - Azure CLI - Studio https://learn.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-cluster?view=azureml-api-2&tabs=python#create
upvoted 2 times

cyberfriends 2 years, 3 months ago
ABD is the answer
upvoted 1 times

ABosco 2 years, 6 months ago
BCD is correct
upvoted 2 times

PI_Team 2 years, 6 months ago
I presume Azure ML studio is correct as well, however, since we have only 3 choices, maybe we need to choose the closet ones! :)wrong answer...
BCD is correct sorry... I tested it, you can't change it with designer
upvoted 1 times

PI_Team 2 years, 6 months ago
Selected Answer: ABD

I presume Azure ML studio is correct as well, however, since we have only 3 choices, maybe we need to choose the closet ones! :)

upvoted 1 times

PI_Team 2 years, 2 months ago

wrong answer... BCD is correct sorry... I tested it, you can't change it with designer

upvoted 2 times

damaldon 2 years, 7 months ago

ACD is correct

upvoted 1 times

Question #108 *Topic 2*

You plan to use automated machine learning by using Azure Machine Learning Python SDK v2 to train a regression model. You have data that has features with missing values, and categorical features with few distinct values.
You need to control whether automated machine learning automatically imputes missing values and encode categorical features as part of the training task.
Which enum of the automl package should you use?

A. ForecastHorizonMode

B. RegressionModels

C. FeaturizationMode

D. RegressionPrimaryMetrics

Suggested Answer: *C*
Official Answer: C

Community vote distribution
C (100%)

f82411e Most Recent 1 year, 2 months ago
A. ForecastHorizonMode: relacionado con series temporales, no aplicable aquí. B. RegressionModels: especifica qué modelos usar, pero no controla la featurization. D. RegressionPrimaryMetrics: define la métrica de evaluación, no el preprocesamiento.
upvoted 1 times

Lion007 2 years, 1 month ago
Correct: C. FeaturizationMode Constructor: FeaturizationMode(value, names=None, *, module=None, qualname=None, type=None, start=1, boundary=None) Fields: AUTO, CUSTOM, OFF See https://learn.microsoft.com/en-us/python/api/azure-ai-ml/azure.ai.ml.automl.featurizationmode?view=azure-python
upvoted 2 times

Mikku123 2 years, 6 months ago
correct!
upvoted 3 times

Question #109 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You use Azure Machine Learning designer to load the following datasets into an experiment:
Dataset1 -
Dataset2 -
You need to create a dataset that has the same columns and header row as the input datasets and contains all rows from both input datasets.
Solution: Use the Join Data module.
Does the solution meet the goal?

Age	Length	Width
3	22	13
7	11	96
18	32	85

Age	Length	Width
11	101	65
6	98	23
33	22	54
17	52	12

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Community vote distribution
B (83%) A (17%)

jmh106 Most Recent 7 months, 3 weeks ago
Debes usar el módulo Add Rows
upvoted 1 times

f82411e 1 year, 2 months ago
Debes usar el módulo Add Rows
upvoted 1 times

avinyc 1 year, 7 months ago
"Join Data" is for SQL type joins, "Add Rows" will be suitable in this case. Hence the solution does not meet the guideline.
upvoted 1 times

jojashi 1 year, 8 months ago
Correct. B I checked on the Designer module description.. Module: Join Data Description: Joins two datasets on selected key columns. we don't need to select key columns and there is no matching rows. We only need to append data. So, Add Row module is right
upvoted 1 times

kay1101 1 year, 9 months ago
not sure about this one, thought it would be append more than join.
upvoted 1 times

sl_mslconsulting 1 year, 9 months ago
It won't work using the Join Data even you use Full Outer Join For each of the rows in either table that have no matching rows in the other, the result includes a row containing missing values.
upvoted 1 times

Plb2 1 year, 11 months ago

Yes this should work, FULL OUTER JOIN is supported https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/join-data?view=azureml-api-2#how-to-configure-join-data

upvoted 1 times

Plb2 1 year, 11 months ago

Yes this should work, FULL OUTER JOIN is supported https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/join-data?view=azureml-api-2#how-to-configure-join-datawhereas on Apply SQL transformation (one of the other questions) FULL OUTER JOIN is not supported; https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/apply-sql-transformation?view=azureml-api-2

upvoted 1 times

Plb2 1 year, 11 months ago

whereas on Apply SQL transformation (one of the other questions) FULL OUTER JOIN is not supported; https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/apply-sql-transformation?view=azureml-api-2

upvoted 1 times

zishankamal 2 years ago

We use the Join Data component to merge two datasets using a database-style join operation. Full Outer Join: A full outer join returns all rows from the left table (table1) and from the right table (table2). https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/join-data?view=azureml-api-2 "Add Rows" or "Execute Python Script" are also possible ways to do this.

upvoted 1 times

edogawa 2 years, 4 months ago

It is the 'Add Rows' component which does it.

upvoted 2 times

umair_hanu 2 years, 7 months ago

b is correct

upvoted 2 times

Question #110 *Topic 2*

HOTSPOT

-

You collect data from a nearby weather station. You have a pandas dataframe named weather_df that includes the following data:

The data is collected every 12 hours: noon and midnight.

You plan to use automated machine learning to create a time-series model that predicts temperature over the next seven days. For the initial round of training, you want to train a maximum of 50 different models.

You must use the Azure Machine Learning SDK v2 to run an automated machine learning experiment to train these models.

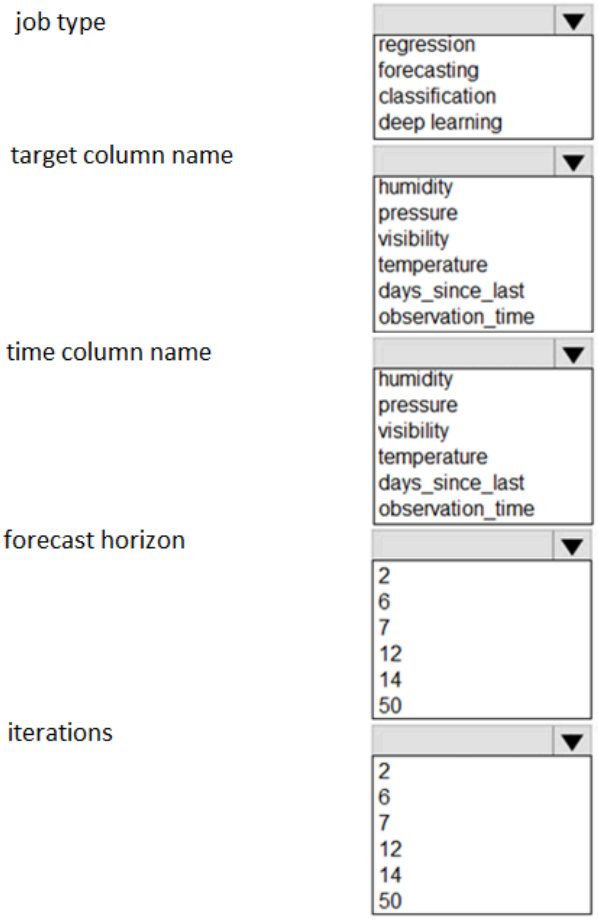
You need to configure the automated machine learning job and its settings.

How should you configure parameters of the classes that implement the job? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Temperature	Observation_time	Humidity	Pressure	Visibility	Days_since_last observation
74	2019/10/2 00:00	0.62	29.87	3	0.5
89	2019/10/2 12:00	0.70	28.88	10	0.5
72	2019/10/3 00:00	0.64	30.00	8	0.5
80	2019/10/3 12:00	0.66	29.75	7	0.5

Answer Area



Suggested Answer:

Answer Area

job type: regression

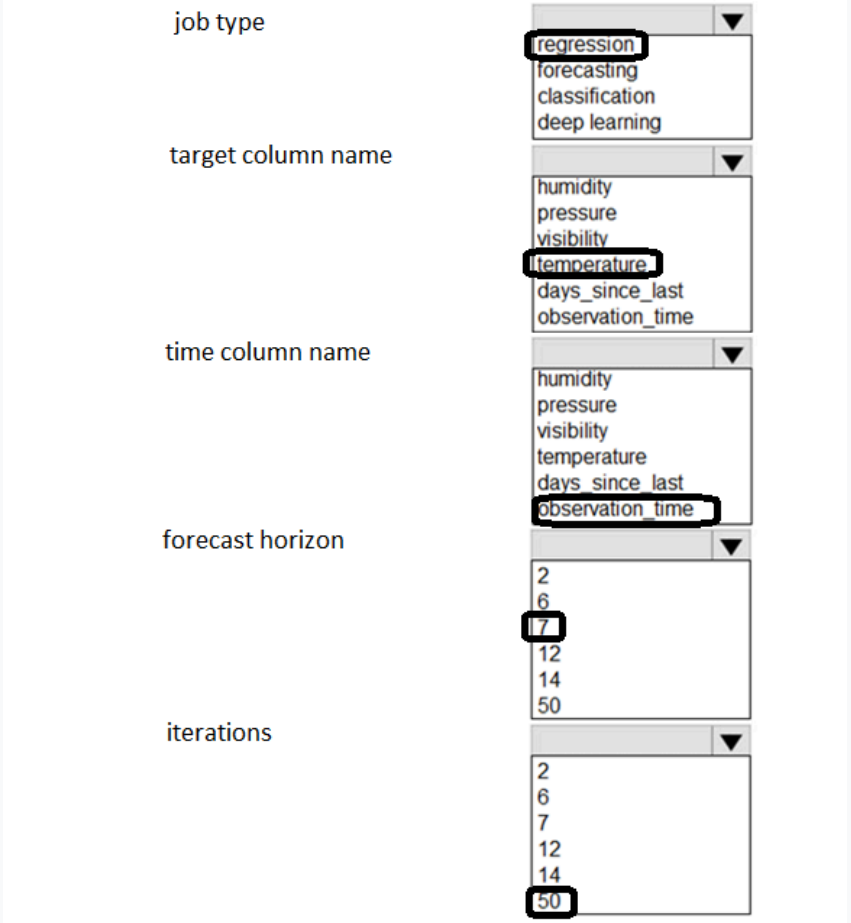
target column name: temperature

time column name: observation_time

forecast horizon: 7

iterations: 50

Answer Area



miumau Highly Voted 2 years, 6 months ago

Shouldn't the forecast horizon be 14, since the data is in 12 hour intervals & we are asked to get data for the next 7 days? https://learn.microsoft.com/en-us/azure/machine-learning/reference-automated-ml-forecasting?view=azureml-api-2#forecasting "The maximum forecast horizon in units of time-series frequency. These units are based on the inferred time interval of your training data, (Ex: monthly, weekly) that the forecaster uses to predict."

upvoted 15 times

umair_hanu Highly Voted 2 years, 7 months ago

IT SHOULD BE FORECASTING

upvoted 9 times

umair_hanu 2 years, 7 months ago

IT SHOULD BE FORECASTINGAgree, forecasting

upvoted 9 times

damaldon 2 years, 7 months ago

Agree, forecasting

upvoted 1 times

a2c6d4a Most Recent 1 year, 8 months ago

the correct configuration for the parameters of the classes that implement the job is as follows: Job type: Forecasting target column name: Temperature time column name: Observation time forecast horizon: 7 days iterations: 50

upvoted 1 times

TA_ 1 year, 10 months ago

On exam 15-03-2024

upvoted 2 times

NullVoider_0 2 years ago

Forcast Horizon is 14 since the data collected is every 12 hours which is twice a day for 7 days.

upvoted 1 times

NullVoider_0 2 years ago

Format Horizon is 14 since the data collected is every 12 hours which is twice a day for 7 days.

upvoted 1 times

PradhanManva 2 years, 4 months ago

CHATGPT - # Define your time-series forecasting task type (e.g., forecasting_future, forecasting_horizon). # Specify the column you want to predict (e.g., 'temperature') and the target lags (e.g., [1, 2, 3]). # This is important for time-series forecasting. task = 'forecasting_future' target_column_name = 'temperature' ///********** # Configure AutoML settings automl_config = AutoMLConfig(task=task, primary_metric='normalized_root_mean_squared_error', training_data=weather_df, label_column_name=target_column_name, max_models=50, # Maximum number of models to train time_column_name='timestamp', # Name of the time column in your DataFrame forecast_horizon=7) ********/

upvoted 1 times

orionduo 2 years, 5 months ago

A regression model provides a function that describes the relationship between one or more independent variables and a response, dependent, or target variable. For example, the relationship between height and weight may be described by a linear regression model. The first anwser should be forecasting

upvoted 3 times

orionduo 2 years, 5 months ago

A regression model provides a function that describes the relationship between one or more independent variables and a response, dependent, or target variable. For example, the relationship between height and weight may be described by a linear regression model. The first anwser should be forecastingmy mistake regression model can be used as prediction as well

upvoted 3 times

orionduo 2 years, 5 months ago

my mistake regression model can be used as prediction as well

upvoted 1 times

PI_Team 2 years, 6 months ago

Task is forecasting

upvoted 1 times

Question #111 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have the following Azure subscriptions and Azure Machine Learning service workspaces:
You need to obtain a reference to the ml-project workspace.
Solution: Run the following Python code:
Does the solution meet the goal?

Subscription	Workspace	Comment
385bdfe5-4cef-4ad4-b977-3f86d92727c9	ml-default	This is default subscription.
5a5891d1-557a-4234-9b83-2e90412b1068	ml-project	The information required to uniquely identify this workspace is stored in the file config.json in the same folder as the Python script.

```
from azure.ai ml import MLClient
ws = MLClient.workspaces.get("ml-project")
```

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Community vote distribution
B (100%)

AzureGeek79 Most Recent 1 year, 10 months ago
Correct answer is A, see reference below: https://learn.microsoft.com/en-us/python/api/azure-ai-ml/azure.ai.ml.operations.workspaceoperations?view=azure-python#azure-ai-ml-operations-workspaceoperations-get
upvoted 2 times

sl_mslconsulting 2 years, 3 months ago
you need to specify subscription id when instantiating the MLClient.
upvoted 1 times

Techlover74 2 years, 6 months ago
B is the correct answer because the argument is not passed with the key in get method.
upvoted 1 times

Question #112 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have the following Azure subscriptions and Azure Machine Learning service workspaces:
You need to obtain a reference to the ml-project workspace.
Solution: Run the following Python code:
Does the solution meet the goal?

Subscription	Workspace	Comment
385bdfe5-4cef-4ad4-b977-3f86d92727c9	ml-default	This is default subscription.
5a5891d1-557a-4234-9b83-2e90412b1068	ml-project	The information required to uniquely identify this workspace is stored in the file config.json in the same folder as the Python script.

```
from azure.ai.ml import MLClient
ws = MLClient(workspace_name= "ml-project")
```

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Community vote distribution
B (100%)

sl_mslconsulting Most Recent 1 year, 9 months ago
Need to specify the subscription id when instantiating the MLClient
upvoted 1 times

Techlover74 2 years ago
B is correct.
upvoted 1 times

Matt2000 2 years ago
Same question as the one before (topic 2, 111).My mistake. The questions are different.
upvoted 1 times

Matt2000 2 years ago
Same question as the one before (topic 2, 111).
upvoted 1 times

Matt2000 2 years ago
My mistake. The questions are different.
upvoted 1 times

Question #113 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have the following Azure subscriptions and Azure Machine Learning service workspaces:
You need to obtain a reference to the ml-project workspace.
Solution: Run the following Python code:
Does the solution meet the goal?

Subscription	Workspace	Comment
385bdfe5-4cef-4ad4-b977-3f86d92727c9	ml-default	This is default subscription.
5a5891d1-557a-4234-9b83-2e90412b1068	ml-project	The information required to uniquely identify this workspace is stored in the file config.json in the same folder as the Python script.

```
from azure.ai.ml import MLClient
ws_from_config = MLClient.from_config()
```

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Community vote distribution
A (50%) B (50%)

jmh106 Most Recent 7 months, 3 weeks ago
La credencial NO es obligatoria Si no se pasa, el SDK usa automáticamente: DefaultAzureCredential() Esto está documentado y es el comportamiento estándar del SDK v2.
upvoted 1 times

jmh106 7 months, 3 weeks ago
Cumple exactamente el escenario MLClient.from_config(): Lee subscription_id, resource_group y workspace_name desde config.json Usa las credenciales configuradas Devuelve un MLClient válido, que en SDK v2 es la referencia al workspace El archivo config.json contiene la información de ml-project, tal como indica la tabla del enunciado. Es la forma recomendada y canónica en SDK v2 para este caso.
upvoted 1 times

Murzfam 1 year, 6 months ago
The questions is reference to ml-project workspace. The solutions is only about auth, so the answer is B.
upvoted 1 times

jl420 1 year, 9 months ago
MLClient.from_config() can be used without explicitly providing credentials, as long as your authentication requirements are met in one of the following ways: Default Authentication with Azure CLI Login: If you've already authenticated using the Azure CLI (e.g., az login), MLClient.from_config() will use your active Azure CLI session for authentication. This is often convenient for development environments and local testing. Managed Identity (for Azure Compute Resources): If you're running your code on an Azure resource with a managed identity (like an Azure Virtual Machine, Azure Kubernetes Service, or Azure Machine Learning Compute Instance), the SDK can authenticate using the managed identity associated with that resource. No additional credentials are needed, as long as the managed identity has access to the Azure ML workspace. Environment-Based Authentication: If environment variables for Azure credentials are set (e.g., AZURE_CLIENT_ID, AZURE_TENANT_ID, and AZURE_CLIENT_SECRET for a service principal), MLClient.from_config() will pick up these credentials automatically.
upvoted 2 times

evangelist 2 years, 2 months ago
Answer should be Yes, it appeared previously in the exam questions
upvoted 1 times

sl_mslconsulting 2 years, 3 months ago

credential is a required parameter in the from_config method . Link: https://learn.microsoft.com/en-us/python/api/azure-ai-ml/azure.ai.ml.mlclient?view=azure-python#azure-ai-ml-mlclient-from-config

upvoted 2 times

Plb2 2 years, 5 months ago

credential-parameter is required

upvoted 2 times

zishankamal 2 years, 6 months ago

Unsure. The credential parameter is not specified which is mandatory. Otherwise class and method is correct.

upvoted 2 times

HOTSPOT

-

You manage an Azure Machine Learning workspace named workspace1.

You must register an Azure Blob storage datastore in workspace1 by using an access key. You develop Python SDK v2 code to import all modules required to register the datastore.

You need to complete the Python SDK v2 code to define the datastore.

How should you complete the code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

```
store = AzureBlobDatastore(
        name="blobdatastore",
        description="Sample blob datastore",
        account_name="sampleaccount1",
        [ filesystem / container_name / file_share_name ] ="samplecontainer1"
        protocol=" [ abfss / wasbs / azureml ] ",
        credentials=AccountKeyCredentials{
                "account_key":
            "6L4VGVXLy30FM/RwmfLGppGj+xNetfUwDVBNPYwYDQDT1E+SirZcjP1a+xCHl
        +D9BTwLnYKhrERO+AStap8yKw=="
                },
          )
```

Suggested Answer:

Answer Area

```
store = AzureBlobDatastore(
        name="blobdatastore",
        description="Sample blob datastore",
        account_name="sampleaccount1",
        [ filesystem | container_name | file_share_name ] ="samplecontainer1"
        protocol=" [ abfss | wasbs | azureml ] ",
        credentials=AccountKeyCredentials{
                "account_key":
            "6L4VGVXLy30FM/RwmfLGppGj+xNetfUwDVBNPYwYDQDT1E+SirZcjP1a+xCHI
        +D9BTwLnYKhrERO+AStap8yKw=="
                },
    )
```

Answer Area

```
store = AzureBlobDatastore(
        name="blobdatastore",
        description="Sample blob datastore",
        account_name="sampleaccount1",
        [ filesystem | container_name | file_share_name ] ="samplecontainer1"
        protocol=" [ abfss | wasbs | azureml ] ",
        credentials=AccountKeyCredentials{
                "account_key":
            "6L4VGVXLy30FM/RwmfLGppGj+xNetfUwDVBNPYwYDQDT1E+SirZcjP1a+xCHI
        +D9BTwLnYKhrERO+AStap8yKw=="
                },
    )
```


Matt2000 Highly Voted 2 years, 6 months ago

wasb[s] has been deprecated. Reference: https://learn.microsoft.com/en-us/azure/databricks/connect/storage/azure-storage. abfs[s] seems to be recommended instead.

upvoted 5 times

jl420 Highly Voted 1 year, 9 months ago

container_name; abfss, because wasb is deprecated

upvoted 5 times

Karthikat 2 years, 5 months ago

answer is correct https://github.com/Azure/azureml-examples/blob/main/sdk/python/resources/datastores/datastore.ipynb

upvoted 2 times

Question #115 *Topic 2*

You manage an Azure Machine Learning workspace.
You plan to import data from Azure Data Lake Storage Gen2.
You need to build a URI that represents the storage location.
Which protocol should you use?

A. https

B. adl

C. abfss

D. wasbs

Suggested Answer: *C*
Official Answer: C

Community vote distribution
C (100%)

3a0b61c Most Recent 1 year, 11 months ago
correct https://learn.microsoft.com/en-us/azure/machine-learning/concept-data?view=azureml-api-2#uri
upvoted 2 times

thea_ipynb69 1 year, 11 months ago
correct
upvoted 3 times

Question #116 *Topic 2*

You manage an Azure Machine Learning workspace. You have a folder that contains a CSV file. The folder is registered as a folder data asset.
You plan to use the folder data asset for data wrangling during interactive development.
You need to access and load the folder data asset into a Pandas data frame.
Which method should you use to achieve this goal?

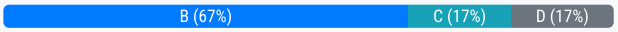
A. mltable.from_parquet_files()

B. mltable.from_delimited_files()

C. mltable.from_data_lake()

D. mltable.load()

Suggested Answer: *B*
Official Answer: B

Community vote distribution


445f1bd Most Recent 1 year, 1 month ago
Is load() guys. Check this ... # load mltable from azureml data asset short uri from mltable import load from azure.ai.ml import MLClient from azure.identity import DefaultAzureCredential credential = DefaultAzureCredential() ml_client = MLClient(credential, , ,) tbl = load('azureml: : ', ml_client=ml_client) in this link https://learn.microsoft.com/en-us/python/api/mltable/mltable?view=azure-ml-py#mltable-from-delimited-files
upvoted 1 times

avinyc 1 year, 7 months ago
mltable.from_delimited_files()
upvoted 1 times

evangelist 1 year, 8 months ago
csv=>delimited file
upvoted 1 times

sai384957324 1 year, 10 months ago
B is correct
upvoted 1 times

401d299_111 1 year, 10 months ago
correct
upvoted 1 times

3a0b61c 1 year, 11 months ago
correct https://learn.microsoft.com/en-us/python/api/mltable/mltable?view=azure-ml-py#mltable-from-delimited-filessorry. B is correct.
upvoted 1 times

3a0b61c 1 year, 11 months ago
Selected Answer: C
correct https://learn.microsoft.com/en-us/python/api/mltable/mltable?view=azure-ml-py#mltable-from-delimited-files
upvoted 1 times

3a0b61c 1 year, 11 months ago
sorry. B is correct.
upvoted 1 times

nicorg5 1 year, 11 months ago
answer is correct
upvoted 1 times

Question #117 *Topic 2*

You manage an Azure Machine Learning workspace named proj1.
You plan to use assets defined in proj1 to create a pipeline in the Machine Learning studio designer.
You need to set the Registry name filter to display only the list of assets defined in proj1.
What should you set the Registry name filter to?

A. proj1

B. azureml-meta

C. azureml

D. workspace

Suggested Answer: *A*
Official Answer: A

Community vote distribution
D (100%)

smaha_18 Most Recent 8 months ago
https://learn.microsoft.com/en-us/azure/machine-learning/concept-designer?view=azureml-api-2 Designer creates pipelines using building blocks from Azure Machine Learning asset libraries. The asset libraries include these pipeline building blocks: Data Models Components The Data, Model, and Component tabs on the left side of Designer show assets in your workspace and in all Azure Machine Learning registries that you can access. To view assets from specific registries, select the Registry name filter. Assets you created in your current workspace are in the Workspace registry. Assets provided by Azure Machine Learning are in the azureml registry.
upvoted 1 times

testgm 1 year, 8 months ago
Workspace
upvoted 1 times

Sadhak 1 year, 9 months ago
workspace is the right answer.
upvoted 1 times

goran74 1 year, 11 months ago
You should set the Registry name filter to D. workspace. In Azure Machine Learning, setting the Registry name filter to "workspace" ensures that only the assets defined within the current workspace (in this case, proj1) are displayed, rather than assets from other registries.
upvoted 3 times

Question #118 *Topic 2*

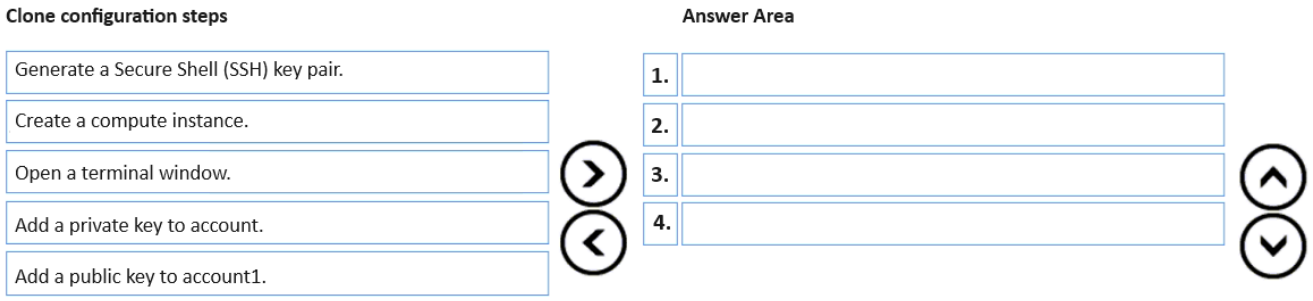
DRAG DROP

-

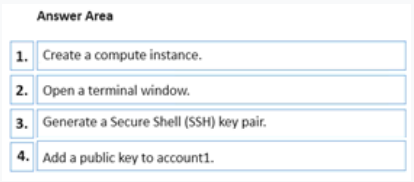
You have an Azure Machine Learning workspace named WS1 and a GitHub account named account1 that hosts a private repository named repo1.

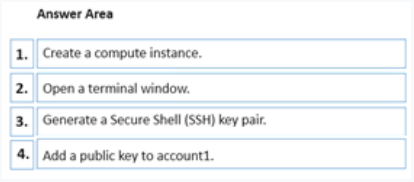
You need to clone repo1 to make it available directly from WS1. The configuration must maximize the performance of the repo1 clone.

Which four actions should you perform in sequence?



Suggested Answer:





f11c733 Most Recent 2 years, 2 months ago

Correct

upvoted 3 times

Question #119 *Topic 2*

You manage an Azure Machine Learning workspace. You design a training job that is configured with a serverless compute.
The serverless compute must have a specific instance type and count.
You need to configure the serverless compute by using Azure Machine Learning Python SDK v2.
What should you do?

A. Specify the compute name by using the compute parameter of the command job.

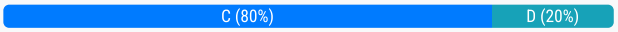
B. Configure the tier parameter to Dedicated VM.

C. Initialize and specify the ResourceConfiguration class.

D. Initialize AmiCompute class with size and type specification.

Suggested Answer: *C*
Official Answer: C

Community vote distribution



f82411e Most Recent 1 year, 2 months ago
from azure.ai.ml.entities import CommandJob, ResourceConfiguration job = CommandJob(... resources=ResourceConfiguration(instance_count=2, instance_type="Standard_DS3_v2"))
upvoted 2 times

ulg 1 year, 6 months ago
To configure serverless compute with a specific instance type and count in Azure Machine Learning using the Python SDK v2, you need to use the ResourceConfiguration class.https://learn.microsoft.com/en-us/azure/machine-learning/how-to-use-serverless-compute?view=azureml-api-2&tabs=python
upvoted 1 times

ulg 1 year, 6 months ago
Selected Answer: C
To configure serverless compute with a specific instance type and count in Azure Machine Learning using the Python SDK v2, you need to use the ResourceConfiguration class.
upvoted 1 times

ulg 1 year, 6 months ago
https://learn.microsoft.com/en-us/azure/machine-learning/how-to-use-serverless-compute?view=azureml-api-2&tabs=python
upvoted 1 times

avinyc 1 year, 7 months ago
Correct answer should be ResourceConfiguration (C). The AmlCompute class is used for creating managed compute clusters, not for configuring serverless compute.
upvoted 1 times

Sadhak 1 year, 8 months ago
https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute(class)?view=azure-ml-py
upvoted 1 times

AzureGeek79 1 year, 10 months ago
Given answer is correct.
upvoted 1 times

HOTSPOT

-

You manage an Azure subscription that contains the following resources:

You plan to implement a solution that will automatically trigger the retraining of the model implemented by MLPipeline1. The trigger must be invoked if data drift is detected in Dataset1.

You need to select the components to invoke and run the solution. The solution must minimize coding implementation and maintenance efforts.

Which components should you select? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

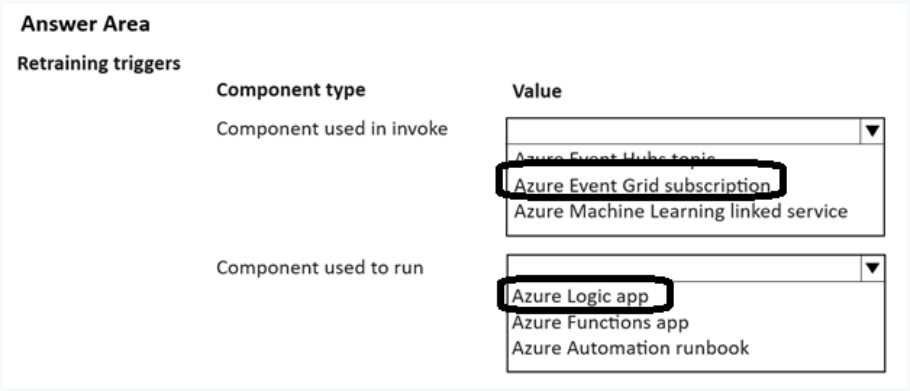
Name	Resource type	Description
Workspace1	Azure Machine Learning workspace	Contains a training pipeline named **MLPipeline1** and a dataset named **Dataset1**
DataFactory1	Azure Data Factory	Contains a pipeline named **DFPipeline1** that runs **MLPipeline1**

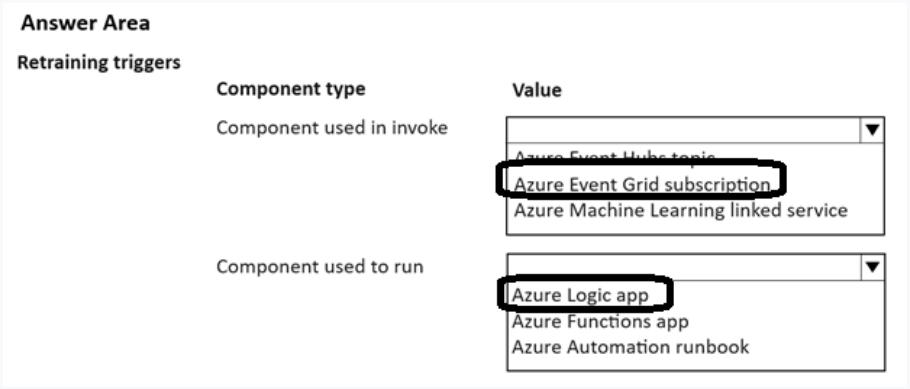
Answer Area

Retraining triggers

Component type	Value
Component used in invoke	Azure Event Hubs topic Azure Event Grid subscription Azure Machine Learning linked service
Component used to run	Azure Logic app Azure Functions app Azure Automation runbook

Suggested Answer:





445f1bd Most Recent 1 year ago

Event grid Logic apps (since it says minimal code)

upvoted 1 times

jefimija 1 year, 10 months ago

Invoke Component: Azure Event Grid Run Component: Azure FunctionsOr it could be Logic Apps Can someone explain the difference between Azure Functins and Logic Apps?I think it's Logic App since is a visual designer with minimal or no code, whereas Functions is more code-centric

upvoted 3 times

jefimija 1 year, 10 months ago

Invoke Component: Azure Event Grid Run Component: Azure Functions

upvoted 3 times

jefimija 1 year, 9 months ago

Or it could be Logic Apps

Can someone explain the difference between Azure Functins and Logic Apps?

upvoted 1 times

jefimija 1 year, 9 months ago

Or it could be Logic Apps Can someone explain the difference between Azure Functins and Logic Apps?I think it's Logic App since is a visual designer with minimal or no code, whereas Functions is more code-centric

upvoted 1 times

nposteraro 1 year, 1 month ago

I think it's Logic App since is a visual designer with minimal or no code, whereas Functions is more code-centric

upvoted 1 times

Question #121 *Topic 2*

You manage an Azure Machine Learning workspace.
You must create and configure a compute cluster for a training job by using Python SDK v2.
You need to create a persistent Azure Machine Learning compute resource, specifying the fewest possible properties.
Which two properties should you define? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. size

B. win_instances

C. type

D. name

E. max_instances

Suggested Answer: *A, E*
Official Answer: AE

Community vote distribution
AE (71%) CD (14%) AD (14%)

f82411e Most Recent 1 year, 2 months ago
B. win_instances: Not a valid property. C. type: Not directly specified; the type is implied by the class AmlCompute. E. max_instances: Optional — useful to auto-scale, but not required. cpu_cluster = AmlCompute(name="cpu-cluster", size="Standard_DS11_v2")
upvoted 1 times

avinyc 1 year, 7 months ago
https://learn.microsoft.com/en-us/AZURE/machine-learning/how-to-create-attach-compute-cluster?view=azureml-api-1&tabs=python
upvoted 2 times

gunn_m 1 year, 8 months ago
Sorry, my last answer was wrong, the correct answer would be A and E
upvoted 1 times

gunn_m 1 year, 8 months ago
from azure.ai.ml.entities import AmlCompute from azure.ai.ml import MLClient ml_client = MLClient(credential=DefaultAzureCredential(), subscription_id="your-subscription-id", resource_group_name="your-resource-group", workspace_name="your-workspace-name") compute_cluster = AmlCompute(name="my-compute-cluster", size="Standard_DS3_v2") ml_client.compute.begin_create_or_update(compute_cluster) A and D
upvoted 1 times

Sadhak 1 year, 9 months ago
To create a persistent Azure Machine Learning Compute resource in Python, specify the size and max_instances properties. Azure Machine Learning then uses smart defaults for the other properties. size: The VM family of the nodes created by Azure Machine Learning Compute. max_instances: The maximum number of nodes to autoscale up to when you run a job on Azure Machine Learning Compute. https://learn.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-cluster?view=azureml-api-2&tabs=python
upvoted 1 times

Sadhak 1 year, 9 months ago
To create a persistent Azure Machine Learning Compute resource in Python, specify the size and max_instances properties. Azure Machine Learning then uses smart defaults for the other properties. size: The VM family of the nodes created by Azure Machine Learning Compute. https://learn.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-cluster?view=azureml-api-2&tabs=python
upvoted 2 times

onurag 1 year, 9 months ago
should be size and name, type is not essential property
upvoted 1 times

PrenCarr 1 year, 10 months ago
To create a persistent Azure Machine Learning compute resource with the fewest possible properties using the Python SDK v2, you should define: type © name (D) These two properties are essential for creating the compute resource. The type specifies the kind of compute resource, and the name gives it a unique identifier within your workspace.

upvoted 1 times

Question #122 *Topic 2*

You manage an Azure Machine Learning workspace named Workspace1.
You plan to create a pipeline in the Azure Machine Learning Studio designer. The pipeline must include a custom component.
You need to ensure the custom component can be used in the pipeline.
What should you do first?

A. Create a pipeline endpoint.

B. Add a linked service to Workspace1.

C. Upload a .json file to Workspace1.

D. Create a datastore.

E. Upload a .yaml file to Workspace1.

Suggested Answer: *E*
Official Answer: E

Community vote distribution
E (100%)

jl420 Most Recent 1 year, 9 months ago
Although E seems correct, it assumes the custom component already exists. However, the question does not say "import existing", it says "must include" which means I can create it on the fly IN DESIGNER GUI without ever using a YAML file, unless I want to export it. What IS required for any pipeline component to work is D. Create a DATASTORE. In fact, this should be the first thing before you can upload a YAML. Going with D. as the answer, although I admit it could still be E and was just poorly written.
upvoted 1 times

PrenCarr 1 year, 10 months ago
correct
upvoted 1 times

Question #123 *Topic 2*

You have an Azure Machine Learning workspace and a serverless Spark compute resource.
You plan to run the same Spark session every 40 minutes. Each session runs for 15 minutes.
During testing, you observe a 15-minute delay at the start of the Spark sessions.
You need to reduce the delay to less than 1 minute.
What should you do?

A. Configure the session timeout to be 25 minutes.

B. Increase the number of nodes to 16.

C. Enable dynamically allocated executors.

D. Enable an isolated compute.

Suggested Answer: *A*
Official Answer: A

Community vote distribution
A (100%)

jl420 Most Recent 1 year, 9 months ago
By configuring the session timeout to be 25 minutes, you allow the Spark session to remain active between runs, preventing it from shutting down between each 40-minute interval. Since each session runs for 15 minutes, setting the timeout to 25 minutes keeps the session warm and reduces the initialization delay.
upvoted 2 times

Question #124 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You manage an Azure Machine Learning workspace. The development environment for managing the workspace is configured to use Python SDK v2 in Azure Machine Learning Notebooks.
A Synapse Spark Compute is currently attached and uses system-assigned identity.
You need to use Python code to update the Synapse Spark Compute to use a user-assigned identity.
Solution: Create an instance of the MLClient class.
Does the solution meet the goal?

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Sadhak Most Recent 1 year, 9 months ago

To update a Synapse Spark Compute to use a user-assigned identity in Python code, you need to access the Azure Synapse Analytics management client and configure the "identity" property of your Synapse Spark pool object with the desired user-assigned managed identity details, specifying the "type" as "UserAssigned" and providing the relevant "principalId" of the user-assigned identity. Key steps: Import necessary libraries. Code from azure.mgmt.synapse import SynapseManagementClient from azure.identity import DefaultAzureCredential

upvoted 1 times

Question #125 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You manage an Azure Machine Learning workspace. The development environment for managing the workspace is configured to use Python SDK v2 in Azure Machine Learning Notebooks.
A Synapse Spark Compute is currently attached and uses system-assigned identity.
You need to use Python code to update the Synapse Spark Compute to use a user-assigned identity.
Solution: Initialize the DefaultAzureCredential class.
Does the solution meet the goal?

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Sadhak Most Recent 1 year, 9 months ago
from azure.identity import DefaultAzureCredential from azure.mgmt.synapse import SynapseManagementClient
upvoted 1 times

Question #126 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You manage an Azure Machine Learning workspace. The development environment for managing the workspace is configured to use Python SDK v2 in Azure Machine Learning Notebooks.
A Synapse Spark Compute is currently attached and uses system-assigned identity.
You need to use Python code to update the Synapse Spark Compute to use a user-assigned identity.
Solution: Configure the IdentityConfiguration class with the appropriate identity type.
Does the solution meet the goal?

A. Yes

B. No

Suggested Answer: *A*
Official Answer: A

Community vote distribution
A (100%)

KeiNek Most Recent 1 year, 6 months ago
To use system-assigned identity, pass IdentityConfiguration, with type set to SystemAssigned, as the identity parameter of the SynapseSparkCompute class. This code snippet updates a Synapse Spark pool to use a system-assigned identity : synapse_identity = IdentityConfiguration(type="SystemAssigned") Ref : https://learn.microsoft.com/en-us/azure/machine-learning/how-to-manage-synapse-spark-pool?view=azureml-api-2&tabs=sdk#update-the-synapse-spark-pool
upvoted 1 times

D0ktor 1 year, 9 months ago
Seems correct to me
upvoted 2 times

Question #127 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have an Azure Machine Learning workspace that includes an AmlCompute cluster and a batch endpoint.
You clone a repository that contains an MLflow model to your local computer.
You need to ensure that you can deploy the model to the batch endpoint.
Solution: Add a compute resource to the workspace.
Does the solution meet the goal?

A. Yes

B. No

Suggested Answer: *A*
Official Answer: A

Community vote distribution
B (100%)

f82411e Most Recent 1 year, 2 months ago
B - no
upvoted 1 times

gunn_m 1 year, 8 months ago
B - No
upvoted 1 times

Arvindu89 1 year, 9 months ago
You also need to register the model and create batch EndPoint
upvoted 1 times

Question #128 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have an Azure Machine Learning workspace that includes an AmlCompute cluster and a batch endpoint.
You clone a repository that contains an MLflow model to your local computer.
You need to ensure that you can deploy the model to the batch endpoint.
Solution: Register the model in the workspace.
Does the solution meet the goal?

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Community vote distribution
A (100%)

f82411e Most Recent 1 year, 2 months ago
A - SI
upvoted 1 times

jl420 1 year, 9 months ago
Registering the model is the required next step. Followed by creating a deployment config to deploy the batch endpoint using existing cluster.
upvoted 1 times

jefimija 1 year, 9 months ago
A. Yes Explanation: Registering the MLflow model in the Azure Machine Learning workspace is a necessary step before deploying the model to a batch endpoint. The model must be registered so that it can be referenced during deployment. Once registered, you can create a batch deployment that uses the model along with the appropriate compute resources. Therefore, registering the model meets the goal of ensuring that the model can be deployed to the batch endpoint.
upvoted 2 times

Question #129 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have an Azure Machine Learning workspace that includes an AmlCompute cluster and a batch endpoint.
You clone a repository that contains an MLflow model to your local computer.
You need to ensure that you can deploy the model to the batch endpoint.
Solution: Create a datastore in the workspace.
Does the solution meet the goal?

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Community vote distribution
B (100%)

f82411e Most Recent 1 year, 2 months ago
B - No
upvoted 1 times

Question #130 *Topic 2*

HOTSPOT

-

You design a data processing strategy for a machine learning project.

The data that must be processed includes unstructured flat files that must be processed in real time.

The data transformation must be executed on a serverless compute and optimized for big data analytical workloads.

You need to select the Azure services for the data science team.

Which storage and data processing service should you use? To answer, select the appropriate option in the answer area.

NOTE: Each correct selection is worth one point.

Data ingestion strategy

Requirement	Service
Data storage for model training workloads	Azure Blob Storage Azure Data Lake Storage Gen 2 Azure SQL Database
Data processing solution	Azure Data Factory Azure Databricks Azure Synapse Analytics

Suggested Answer:

Data ingestion strategy

Requirement	Service
Data storage for model training workloads	Azure Blob Storage Azure Data Lake Storage Gen 2 **Azure SQL Database**
Data processing solution	Azure Data Factory Azure Databricks **Azure Synapse Analytics**

Data ingestion strategy

Requirement	Service
Data storage for model training workloads	Azure Blob Storage Azure Data Lake Storage Gen 2 **Azure SQL Database**
Data processing solution	Azure Data Factory Azure Databricks **Azure Synapse Analytics**

jefimija Highly Voted 1 year, 10 months ago

Data storage for model training workloads: Azure Data Lake Storage Gen 2 Data processing solution: Azure Databricks

upvoted 13 times

smaha_18 Most Recent 7 months, 3 weeks ago

Azure Data Lake Storage Gen 2 is designed for big data analytics and unstructured data (like flat files). Azure Databricks is a serverless and scalable data processing service designed for big data processing. It integrates closely with Apache Spark, which is highly efficient for big data analytical workloads, and real-time data processing. Databricks is ideal for processing unstructured data and applying machine learning models on large datasets.

upvoted 1 times

f82411e 1 year, 2 months ago

Almacenamiento: Azure Data Lake Storage Gen2 Procesamiento: Azure Databricks

upvoted 1 times

D0ktor 1 year, 9 months ago

Here SERVERLESS is the key word so it should be Azure Data Factory

upvoted 2 times

Question #131 *Topic 2*

DRAG DROP

-

You manage an Azure Machine Learning workspace named workspace1.

You plan to create a registry named registry01 with the help of the following registry.yml (line numbers are used for reference only):

You need to use Azure Machine Learning Python SDK v2 with Python 3.10 in a notebook to interact with workspace1.

Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

```
01 name: DemoRegistry1
02 tags:
03   description: A registry with one primary region and an additional region
04   foo: bar
05 location: eastus
06 replication_locations:
07   - location: eastus
08   - location: eastus2
```

Actions

- Create and load registry01.yml.
- Use begin_create() to create a registry.
- Install Azure Machine Learning Python SDK v2 .
- Load registry01.yml with the get('registry01') method.
- Connect to the workspace.

Sharing machine learning assets with registries



Suggested Answer:

Sharing machine learning assets with registries

- Install Azure Machine Learning Python SDK v2 .
- Use begin_create() to create a registry.
- Connect to the workspace.

Sharing machine learning assets with registries

- Install Azure Machine Learning Python SDK v2 .
- Use begin_create() to create a registry.
- Connect to the workspace.

gunn_m Most Recent 1 year, 8 months ago

The answer is wrong Install Azure Machine Learning Python SDK v2. Connect to the workspace. Use begin_create() to create a registry.

upvoted 4 times

Arvindu89 1 year, 9 months ago

Actions 1. Install Azure machine Learning python ADK v2 2. Connect to workspace 3. User begin_create() to create a registry

upvoted 2 times

jefimija 1 year, 10 months ago

1. install 2. connect 3. use begin create

upvoted 3 times

Question #132 *Topic 2*

DRAG DROP

-

You have an Azure Machine Learning workspace.

You plan to use the terminal to configure a compute instance to run a notebook.

You need to add a new R kernel to the compute instance.

In which order should you perform the actions? To answer, move all actions from the list of actions to the answer area and arrange them in the correct order.

Actions

- IRkernel::installspec(name = 'irenv', displayname = 'New R Env')
- q()
- R
- conda activate r_env
- conda create -n r_env r-essentials r-base

Add a new R kernel to the compute instance



Suggested Answer:

Add a new R kernel to the compute instance

1. conda create -n r_env r-essentials r-base
2. conda activate r_env
3. R
4. IRkernel::installspec(name = 'irenv', displayname = 'New R Env')
5. q()

Add a new R kernel to the compute instance

1. conda create -n r_env r-essentials r-base
2. conda activate r_env
3. R
4. IRkernel::installspec(name = 'irenv', displayname = 'New R Env')
5. q()

Sadhak Most Recent 1 year, 9 months ago

what is R and q()

upvoted 2 times

jefimija 1 year, 10 months ago

correct

upvoted 4 times

Question #133 *Topic 2*

HOTSPOT

-

You manage an Azure Machine Learning workspace named Workspace1 and an Azure Blob Storage accessed by using the URL https://storage1.blob.core.windows.net/data1.

You plan to create an Azure Blob datastore in Workspace1. The datastore must target the Blob Storage by using Azure Machine Learning Python SDK v2. Access authorization to the datastore must be limited to a specific amount of time.

You need to select the parameters of the AzureBlobDatastore class that will point to the target datastore and authorize access to it.

Which parameters should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Datastore parameters

Parameter purpose	Value
Point to the target datastore.	container_name="data1" filesystem="data1" file_share_name="data1"
Authorize access.	credentials=AccountKeyConfiguration credentials=SaSTokenConfiguration credentials=ServicePrincipalCredentials

Suggested Answer:

Datastore parameters

Parameter purpose	Value
Point to the target datastore.	container_name="data1" filesystem="data1" file_share_name="data1"
Authorize access.	credentials=AccountKeyConfiguration credentials=SaSTokenConfiguration credentials=ServicePrincipalCredentials

Datastore parameters

Parameter purpose	Value
Point to the target datastore.	container_name="data1" filesystem="data1" file_share_name="data1"
Authorize access.	credentials=AccountKeyConfiguration credentials=SaSTokenConfiguration credentials=ServicePrincipalCredentials

Sadhak Most Recent 1 year, 9 months ago

https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.data.azure_storage_datastore.azureblobdatastore?view=azure-ml-py

upvoted 1 times

Sadhak 1 year, 9 months ago

Answer is correct.

upvoted 3 times

Question #134 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You manage an Azure Machine Learning workspace. The development environment for managing the workspace is configured to use Python SDK v2 in Azure Machine Learning Notebooks.
A Synapse Spark Compute is currently attached and uses system-assigned identity.
You need to use Python code to update the Synapse Spark Compute to use a user-assigned identity.
Solution: Pass the UserAssignedIdentity class object to the SynapseSparkCompute class.
Does the solution meet the goal?

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Community vote distribution


astone42 Most Recent 1 year, 7 months ago
A Synapse Spark pool can also use a user-assigned identity. For a user-assigned identity, you can pass a managed identity definition, using the IdentityConfiguration class, as the identity parameter of the SynapseSparkCompute class. For the managed identity definition used in this way, set the type to UserAssigned. In addition, pass a user_assigned_identities parameter. The parameter user_assigned_identities is a list of objects of the UserAssignedIdentity class. The resource_idof the user-assigned identity populates each UserAssignedIdentity class object
https://learn.microsoft.com/en-us/azure/machine-learning/how-to-manage-synapse-spark-pool?view=azureml-api-2&tabs=sdk#update-the-synapse-spark-pool
upvoted 1 times

colin1919 1 year, 8 months ago
No. It is tricky, but no. You don't pass the UserAssignedIdentity() to SynapseSparkCompute() but to the user_assigned_identities INSIDE the ManagedIdentityConfiguration(), which in turn is passed to the SynapseSparkCompute() class.
upvoted 1 times

Sadhak 1 year, 8 months ago
spark_compute = ml_client.compute.get(" ") spark_compute.identity = ManagedIdentityConfiguration(type="UserAssigned", user_assigned_identities=["/subscriptions/ /resourcegroups/ /providers/Microsoft.ManagedIdentity/userAssignedIdentities/ "])
upvoted 2 times

Sadhak 1 year, 9 months ago
The answer is "Yes"
upvoted 2 times

Arvindu89 1 year, 9 months ago
The answer is "Yes" spark_compute = ml_client.compute.get(" ") spark_compute.identity = ManagedIdentityConfiguration(type="UserAssigned", user_assigned_identities=["/subscriptions/ /resourcegroups/ /providers/Microsoft.ManagedIdentity/userAssignedIdentities/ "])
upvoted 3 times

Question #135 *Topic 2*

DRAG DROP

-

You have an Azure Machine Learning workspace named WS1.

You plan to use WS1 to train two models named model1 and model2. For model1, you plan to use automated machine learning. For model2, you plan to use Azure Machine Learning designer.

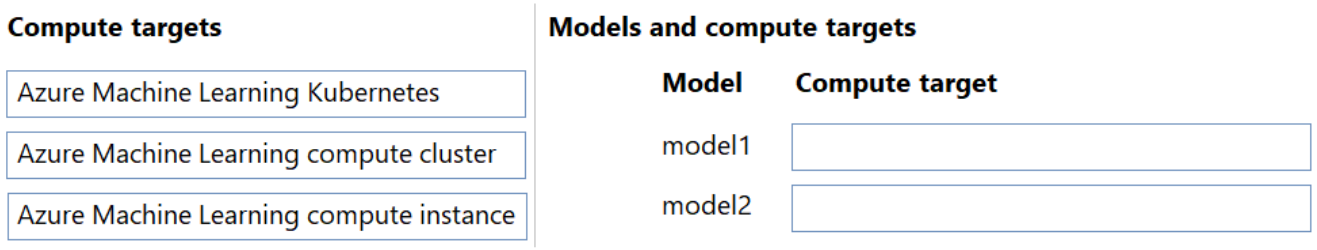
You need to determine the compute targets you should use to train each model. Your solution must ensure the following:

• The compute target for model1 supports auto-shutdown/auto-start based on a schedule.

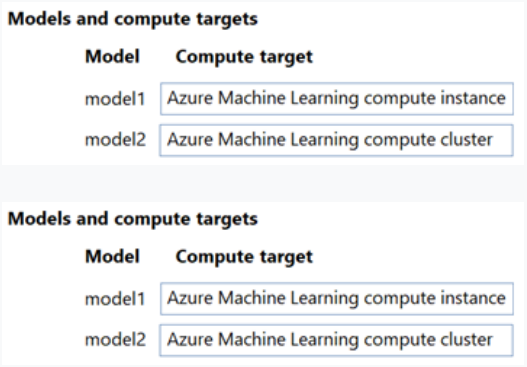
• The compute target for model2 supports the use of low-priority Azure Virtual Machines.

Which compute targets should you use? To answer, move the appropriate compute targets to the correct model. You may use each compute target once, more than once, or not at all. You may need to move the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.



Suggested Answer:



smaha_18 Most Recent 7 months, 3 weeks ago

Answer is correct. ComputeInstance has a parameter named idle_time_before_shutdown_minutes and scheduling can be done. Tier = 'low-priority' can be configured via compute cluster. Model 1- Compute Instance, Model 2 - Compute Cluster

upvoted 1 times

f82411e 1 year, 2 months ago

Cluster compute 1 and 2

upvoted 1 times

gunn_m 1 year, 8 months ago

1 Azure Machine Learning compute cluster 2 Azure Machine Learning compute cluster

upvoted 3 times

gunn_m 1 year, 8 months ago

1 Azure Machine Learning compute cluster 2 Azure Machine Learning compute clusterI did a mistake... 1 Azure Machine Learning compute cluster 2 Azure Machine Learning compute instace (azure virtual machines)

upvoted 3 times

gunn_m 1 year, 8 months ago

I did a mistake... 1 Azure Machine Learning compute cluster 2 Azure Machine Learning compute instace (azure virtual machines)

upvoted 1 times

Arvindu89 1 year, 9 months ago

The answer is correct

upvoted 3 times

Question #136 *Topic 2*

You manage an Azure Machine Learning Workspace named Workspase1 and an Azure Files share named Share1.
You plan to create an Azure Files datastore in Workspace1 to target Share1.
You need to configure permanent access to Share1 from the Azure Files datastore.
Which authorization method should you use?

A. Secondary access key

B. Anonymous access

C. Account SAS key

D. Service SAS key

Suggested Answer: *D*
Official Answer: D

Community vote distribution
A (55%) C (27%) D (18%)

jmh106 Most Recent 7 months, 3 weeks ago
A. Clave de acceso secundaria ✔ Es una clave de cuenta (igual que la primaria). Proporciona acceso permanente. Es una práctica habitual usar la secundaria para permitir rotación sin interrupciones. Cumple el requisito.
upvoted 1 times

f82411e 1 year, 2 months ago
C. Account SAS key Provides delegated access to the entire storage resource (such as Azure Files). Can be configured with far expiration (or renewed periodically) for persistent access. Supports Azure Machine Learning data stores.
upvoted 1 times

jl420 1 year, 9 months ago
The correct answer is: A. Secondary access key Explanation: To configure permanent access to an Azure Files share from an Azure Files datastore in an Azure Machine Learning workspace, you generally use the primary or secondary access key of the Azure Storage account that contains the file share. This provides long-term access without the need to renew credentials, as the access keys remain valid until they are manually regenerated. Why the Other Options Don't Meet the Goal: B. Anonymous access: Azure Files does not support anonymous access. Access requires authentication through either account keys, SAS tokens, or identity-based methods. C. Account SAS key: While an Account SAS key provides scoped access to storage resources, it is typically short-lived and must be regenerated periodically, making it unsuitable for permanent access. D. Service SAS key: Similar to the Account SAS, a Service SAS key is also short-lived and used for temporary access to specific resources. It must be renewed, so it does not meet the requirement for permanent access.
upvoted 3 times

Arvindu89 1 year, 9 months ago
Account SAS and Service SAS are both types of Shared Access Signatures in Azure, but they differ in terms of scope and control. Account SAS: Grants access to resources within a storage account, such as blobs, files, queues, or tables. Offers broader permissions and can cover multiple services within the storage account. Allows you to specify permissions, including read, write, delete, list, and more for the entire account. Service SAS: Grants access to specific resources within a single service, like a specific blob or file share. Offers more fine-grained control compared to Account SAS. Allows you to specify permissions and set constraints, like IP ranges, protocols, and expiration times, but only for the specified resource.
upvoted 2 times

kfgg 1 year, 9 months ago
A user delegation SAS has a maximum expiry interval of 7 days, regardless of the SAS expiration policy. I think answer should be A?
https://learn.microsoft.com/en-us/azure/storage/common/sas-expiration-policy?tabs=azure-portal#configure-a-sas-expiration-policy
https://learn.microsoft.com/en-us/azure/storage/common/storage-account-keys-manage?tabs=azure-portal
upvoted 2 times

jefimija 1 year, 10 months ago
I never came across service sas key, only account sas key
upvoted 2 times

Question #137 *Topic 2*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have an Azure Machine Learning workspace that includes an AmlCompute cluster and a batch endpoint.
You clone a repository that contains an MLflow model to your local computer.
You need to ensure that you can deploy the model to the batch endpoint.
Solution: Create a data asset in the workspace.
Does the solution meet the goal?

A. Yes

B. No

Suggested Answer: *B*
Official Answer: B

Currently there are no comments in this discussion, be the first to comment!

Question #138 *Topic 2*

DRAG DROP

-

You manage an Azure Machine Learning workspace that has an Azure Machine Learning datastore.

Data must be loaded from the following sources:

• a credential-less Azure Blob Storage

• an Azure Data Lake Storage (ADLS) Gen 2 which is not a credential-less datastore

You need to define the authentication mechanisms to access data in the Azure Machine Learning datastore.

Which data access mechanism should you use? To answer, move the appropriate data access mechanisms to the correct storage types. You may use each data access mechanism once, more than once, or not at all. You may need to move the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Answer Area

Data access mechanisms

- Access key
- Shared access signature token
- Service principal
- User identity passthrough

Access data from the Azure Machine Learning datastore

Storage account type	Data access mechanism
Azure Blob Storage	
Azure Data Lake Storage (ADLS) Gen 2	

Suggested Answer:





jmh106 Most Recent 7 months, 3 weeks ago

Azure Blob Storage → Shared Access Signature (SAS) Un SAS permite acceso temporal y sin credenciales permanentes. Es el mecanismo correcto cuando el blob es “sin credenciales”. ADLS Gen2 → Service principal ADLS Gen2 requiere autenticación con Azure AD. El service principal es el método estándar y soportado para acceso desde Azure Machine Learning datastores. User identity passthrough no aplica aquí porque el acceso es desde el datastore, no desde un usuario interactivo. Por qué no las otras opciones Access key → Credencial permanente (no cumple el requisito de “sin credenciales”). User identity passthrough → Solo válido para escenarios interactivos, no para datastores compartidos.

upvoted 1 times

HOTSPOT

-

You have an Azure subscription that contains a resource group named rg-ml.

You plan to create an Azure Machine Learning workspace named workspace1 in rg-ml by using Azure Machine Learning Python SDK v2.

You need to ensure workspace is configured to prevent the collection of potentially sensitive data by Microsoft diagnostics.

How should you complete the provided code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Azure Machine Learning Python SDK v2 code

```
from [ ▼ ] import Workspace
```

[▼] options:
- azure.ai.ml
- azure.ai.ml.entities
- azure.ai.ml.constants

```
ws = Workspace(
    name='workspace1',
    location="eastus",
    display_name="workspace1",
    resource_group="rg-ml",
    [ ▼ ] =False
)
ml_client.workspaces.begin_create(ws)
```

[▼] options:
- hbi_workspace
- public_network_access
- customer_managed_key

Suggested Answer:

Answer Area

Azure Machine Learning Python SDK v2 code

```
from [azure.ai.ml | azure.ai.ml.entities | azure.ai.ml.constants] import Workspace
ws = Workspace(
    name='workspace1',
    location="eastus",
    display_name="workspace1",
    resource_group="rg-ml",
    [hbi_workspace | public_network_access | customer_managed_key] =False
)
ml_client.workspaces.begin_create(ws)
```


Answer Area

Azure Machine Learning Python SDK v2 code

```
from [azure.ai.ml | azure.ai.ml.entities | azure.ai.ml.constants] import Workspace
ws = Workspace(
    name='workspace1',
    location="eastus",
    display_name="workspace1",
    resource_group="rg-ml",
    [hbi_workspace | public_network_access | customer_managed_key] =False
)
ml_client.workspaces.begin_create(ws)
```


b805b03 Most Recent 1 year, 2 months ago

Shouldn't it be "hbi_workspace = True" ?

upvoted 2 times

Question #1 *Topic 3*

You are analyzing a dataset containing historical data from a local taxi company. You are developing a regression model.
You must predict the fare of a taxi trip.
You need to select performance metrics to correctly evaluate the regression model.
Which two metrics can you use? Each correct answer presents a complete solution?
NOTE: Each correct selection is worth one point.

A. a Root Mean Square Error value that is low

B. an R-Squared value close to 0

C. an F1 score that is low

D. an R-Squared value close to 1

E. an F1 score that is high

F. a Root Mean Square Error value that is high

Suggested Answer: *A, D*
RMSE and R2 are both metrics for regression models. A: Root mean squared error (RMSE) creates a single value that summarizes the error in the model. By squaring the difference, the metric disregards the difference between over-prediction and under-prediction. D: Coefficient of determination, often referred to as R2, represents the predictive power of the model as a value between 0 and 1. Zero means the model is random (explains nothing); 1 means there is a perfect fit. However, caution should be used in interpreting R2 values, as low values can be entirely normal and high values can be suspect. Incorrect Answers: C, E: F-score is used for classification models, not for regression models.
Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model

Community vote distribution
AD (100%)

ipindado2020 Highly Voted 5 years, 9 months ago
Agree with AD
upvoted 14 times

GodReign Most Recent 1 year, 10 months ago
Given answer is correct. RMSE of zero means perfect prediction, so we aim for a RMSE score very close to zero. R-Sqaured for Linear Regression or Adjusted R-Squared for multiple regression close to 1 or 100% indicates that the features used are good for explaining the variability in the dependent variable.
upvoted 1 times

evangelist 2 years, 2 months ago
F1 =>classification model, given answer is correct
upvoted 1 times

krishna1818 3 years, 2 months ago
for Linear regression ideally rmse =0 and R2 = 1
upvoted 1 times

ZIMARAKI 3 years, 6 months ago
The given answer is correct.
upvoted 1 times

Edriv 3 years, 8 months ago
Agree AD
upvoted 1 times

Nav727 4 years, 2 months ago
why is E wrong here? "an F1 score that is high" is a good metric right?
upvoted 2 times

Nav727 4 years, 2 months ago
why is E wrong here? "an F1 score that is high" is a good metric right?the f1 sore, precision, recall used for classification problemsf1 score is mainly used for classification problems while rms and r1 for regression

upvoted 2 times

ZIMARAKI 3 years, 6 months ago

the f1 sore, precision, recall used for classification problems

upvoted 2 times

sdz 4 years ago

f1 score is mainly used for classification problems while rms and r1 for regression

upvoted 3 times

mischelin11 4 years, 7 months ago

Agree with AD

upvoted 2 times

kisskeo 4 years, 10 months ago

On Exam 01 Oct 2021

upvoted 3 times

ljljljlj 5 years, 1 month ago

On exam 2021/7/10

upvoted 4 times

dev2dev 5 years, 5 months ago

Coefficient of determination is the actual metric name for R squared in aml

upvoted 2 times

Vipuls 5 years, 7 months ago

Yes, Answer is D

upvoted 2 times

Question #2 *Topic 3*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are using Azure Machine Learning to run an experiment that trains a classification model.
You want to use Hyperdrive to find parameters that optimize the AUC metric for the model. You configure a HyperDriveConfig for the experiment by running the following code:
You plan to use this configuration to run a script that trains a random forest model and then tests it with validation data. The label values for the validation data are stored in a variable named y_test variable, and the predicted probabilities from the model are stored in a variable named y_predicted.
You need to add logging to the script to allow Hyperdrive to optimize hyperparameters for the AUC metric.
Solution: Run the following code:
Does the solution meet the goal?

```
hyperdrive = HyperDriveConfig(estimator=your_estimator,
 hyperparameter_sampling=your_params,
 policy=policy,
 primary_metric_name='AUC',
 primary_metric_goal=PrimaryMetricGoal.MAXIMIZE,
 max_total_runs=6,
 max_concurrent_runs=4)

from sklearn.metrics import roc_auc_score
import logging
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
logging.info("AUC: " + str(auc))
```

A. Yes

B. No

Suggested Answer: *A*
Python printing/logging example: logging.info(message) Destination: Driver logs, Azure Machine Learning designer Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines

Community vote distribution
B (100%)

DennisWitjes Highly Voted 4 years, 6 months ago
To allow optimize hyperdrive you require to use Azure ML run context to log and use it in the hyperdrive config. logging.info() logs it but does not allow for optimizing hyperdrive See: https://docs.microsoft.com/en-us/learn/modules/tune-hyperparameters-with-azure-machine-learning/6a-knowledge-checkagreed. correct answer is B.The question is "You need to add logging to the script to allow Hyperdrive to optimize hyperparameters for the AUC metric." so it should be correct, since it only ask for log. Answer: A. Yes
upvoted 20 times

DennisWitjes Highly Voted 4 years, 6 months ago
To allow optimize hyperdrive you require to use Azure ML run context to log and use it in the hyperdrive config.

logging.info() logs it but does not allow for optimizing hyperdrive

See: https://docs.microsoft.com/en-us/learn/modules/tune-hyperparameters-with-azure-machine-learning/6a-knowledge-check
upvoted 20 times

spaceykacey 4 years, 3 months ago
agreed. correct answer is B.

upvoted 2 times

JohnDup 4 years, 2 months ago

The question is "You need to add logging to the script to allow Hyperdrive to optimize hyperparameters for the AUC metric." so it should be correct, since it only ask for log. Answer: A. Yes

upvoted 1 times

phdykd 3 years ago

B. Logging the AUC metric alone is not enough to allow Hyperdrive to optimize hyperparameters for that metric. The HyperDriveConfig object needs to be used to initiate the Hyperdrive run, and the script needs to pass the HyperDriveConfig object as an argument to the Azure Machine Learning experiment run function. This will enable Hyperdrive to monitor the logged AUC metric and adjust the hyperparameters during the run to optimize the AUC metric.

upvoted 6 times

evangelist Most Recent 1 year, 8 months ago

Get the current run context run = Run.get_context() # Log the AUC score run.log("AUC", auc)

upvoted 1 times

PradhanManva 2 years, 4 months ago

B is the answer

upvoted 1 times

abhishekm94 2 years, 8 months ago

"The training script for your model must log the primary metric during model training using the same corresponding metric name so that the SweepJob can access it for hyperparameter tuning." - link :: https://learn.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters?view=azureml-api-2 clearly says mlflow logging is required. Any other logging will not work.

upvoted 1 times

Tommo565 2 years, 11 months ago

B is correct

upvoted 1 times

klowqw 3 years, 5 months ago

run.log

upvoted 6 times

synapse 3 years, 11 months ago

Copying a good explanation: The question is not about just logging AUC but logging to allow Hyperdrive to optimize hyperparameters for the AUC metric. So you must log using run instance. That way the Hyperdrive has access to that metric to compare with other runs. SO the correct answer is "No"

upvoted 6 times

prasad06 4 years, 5 months ago

The answer should be B

upvoted 5 times

dev2dev 4 years, 11 months ago

technically this might work if filemode was w+ instead of w but logger module is simpleIt still won't work. The logging should be done on run instance so that the hyperdrive experiment can actually get that value. Writing it to a file or printing it to stdout will not help.

upvoted 1 times

dev2dev 4 years, 11 months ago

technically this might work if filemode was w+ instead of w but logger module is simple

upvoted 1 times

chaudha4 4 years, 9 months ago

It still won't work. The logging should be done on run instance so that the hyperdrive experiment can actually get that value. Writing it to a file or printing it to stdout will not help.

upvoted 7 times

Question #3 *Topic 3*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are using Azure Machine Learning to run an experiment that trains a classification model.
You want to use Hyperdrive to find parameters that optimize the AUC metric for the model. You configure a HyperDriveConfig for the experiment by running the following code:
You plan to use this configuration to run a script that trains a random forest model and then tests it with validation data. The label values for the validation data are stored in a variable named y_test variable, and the predicted probabilities from the model are stored in a variable named y_predicted.
You need to add logging to the script to allow Hyperdrive to optimize hyperparameters for the AUC metric.
Solution: Run the following code:
Does the solution meet the goal?

```
hyperdrive = HyperDriveConfig(estimator=your_estimator,
 hyperparameter_sampling=your_params,
 policy=policy,
 primary_metric_name='AUC',
 primary_metric_goal=PrimaryMetricGoal.MAXIMIZE,
 max_total_runs=6,
 max_concurrent_runs=4)

import json, os
from sklearn.metrics import roc_auc_score
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
os.makedirs("outputs", exist_ok = True)
with open("outputs/AUC.txt", "w") as file_cur:
    file_cur.write(auc)
```

A. Yes

B. No

Suggested Answer: *B*
Explanation - Use a solution with logging.info(message) instead. Note: Python printing/logging example: logging.info(message) Destination: Driver logs, Azure Machine Learning designer Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines

Community vote distribution
B (100%)

scott_klein_12345 Highly Voted 4 years, 9 months ago
The answer is No. The exact question appears in the learning material provided by Microsoft. Check out this link - https://docs.microsoft.com/en-us/learn/modules/tune-hyperparameters-with-azure-machine-learning/6a-knowledge-check
upvoted 38 times

chaudha4 Highly Voted 4 years, 9 months ago
I think the answer is wrong. The logging is printing with "AUC: " while the primary_metric_name is "AUC". How is hyperdrive experiment suppose to know that they are the same !! Moreover logging should be done on "run" instance. So many mistakes in this code !!
upvoted 9 times

chaudha4 4 years, 9 months ago
I think the answer is wrong. The logging is printing with "AUC: " while the primary_metric_name is "AUC". How is hyperdrive experiment suppose to know that they are the same !! Moreover logging should be done on "run" instance. So many mistakes in this code !!Have you ever used python for development , its the logging mechnism where u print any string along with generated log "AUC:" or "XYCCC" anything will be there with str(auc). the solution is correct hereYes, but the question is asking to allow Hyperdrive to optimize hyperparameters. You need to use: run.log('AUC', np.float(acc)) https://docs.microsoft.com/en-us/learn/modules/tune-hyperparameters-with-azure-machine-learning/5-hyperdrive-experiment Correct answer is NO.
upvoted 9 times

aakash_0086 4 years, 8 months ago

Have you ever used python for development , its the logging mechnism where u print any string along with generated log "AUC:" or "XYCCC" anything will be there with str(auc). the solution is correct here

upvoted 1 times

aakash_0086 4 years, 8 months ago

Have you ever used python for development , its the logging mechnism where u print any string along with generated log "AUC:" or "XYCCC" anything will be there with str(auc). the solution is correct hereYes, but the question is asking to allow Hyperdrive to optimize hyperparameters. You need to use: run.log('AUC', np.float(acc)) https://docs.microsoft.com/en-us/learn/modules/tune-hyperparameters-with-azure-machine-learning/5-hyperdrive-experiment Correct answer is NO.

upvoted 1 times

treadst0ne 4 years, 7 months ago

Yes, but the question is asking to allow Hyperdrive to optimize hyperparameters. You need to use: run.log('AUC', np.float(acc)) https://docs.microsoft.com/en-us/learn/modules/tune-hyperparameters-with-azure-machine-learning/5-hyperdrive-experiment Correct answer is NO.

upvoted 4 times

evangelist Most Recent 1 year, 8 months ago

Get the current run context run = Run.get_context() # Log the AUC score run.log("AUC", auc)

upvoted 1 times

phdykd 3 years ago

B. No. The solution does not meet the goal as it only saves the AUC score to a text file in the "outputs" directory. It does not log the AUC metric in a way that can be monitored by Azure Machine Learning and Hyperdrive. In order to allow Hyperdrive to optimize hyperparameters for the AUC metric, the AUC score needs to be logged using the run.log method, as follows: import json, os from sklearn.metrics import roc_auc_score import logging auc=roc_auc_score(y_test, y_predicted) run.log("AUC", auc)

upvoted 2 times

Tj87 3 years, 9 months ago

It seems this question has been repeated for several times. Looking at section "Log metrics for hyperparameter tuning" of Microsoft doc(https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters), we need to use the following codes: from azureml.core.run import Run run_logger = Run.get_context() run_logger.log("accuracy", float(val_accuracy)) So log.info doesn't seem to be the correct option. The hyperDrive needs to know the metric value of each run to be able to compare the values

upvoted 2 times

zehraoneexam 3 years, 11 months ago

I think the answer is yes. logging info can print the results. We can have more than one solution...

upvoted 1 times

synapse 3 years, 11 months ago

Answer is B. No. Get a reference to the Azure ML run context and use a run.log() statement to write the AUC value to the run log https://docs.microsoft.com/en-us/learn/modules/tune-hyperparameters-with-azure-machine-learning/6a-knowledge-check

upvoted 1 times

Krupek 4 years ago

answer should be no

upvoted 2 times

BlindedByScience 4 years, 5 months ago

The answer should be 'YES'. The documentation says for print(val) the destination is driver logs.

upvoted 1 times

trickerk 4 years, 7 months ago

"You need to add logging", print() function doesn't add any log. So answer is NO.

upvoted 1 times

dtus 4 years, 8 months ago

the Answer is NO

upvoted 3 times

Question #4 *Topic 3*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are using Azure Machine Learning to run an experiment that trains a classification model.
You want to use Hyperdrive to find parameters that optimize the AUC metric for the model. You configure a HyperDriveConfig for the experiment by running the following code:
You plan to use this configuration to run a script that trains a random forest model and then tests it with validation data. The label values for the validation data are stored in a variable named y_test variable, and the predicted probabilities from the model are stored in a variable named y_predicted.
You need to add logging to the script to allow Hyperdrive to optimize hyperparameters for the AUC metric.
Solution: Run the following code:
Does the solution meet the goal?

```
hyperdrive = HyperDriveConfig(estimator=your_estimator,
 hyperparameter_sampling=your_params,
 policy=policy,
 primary_metric_name='AUC',
 primary_metric_goal=PrimaryMetricGoal.MAXIMIZE,
 max_total_runs=6,
 max_concurrent_runs=4)

import numpy as np
from sklearn.metrics import roc_auc_score
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
print(np.float(auc))
```

A. Yes

B. No

Suggested Answer: *B*
Explanation - Use a solution with logging.info(message) instead. Note: Python printing/logging example: logging.info(message) Destination: Driver logs, Azure Machine Learning designer Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines

Community vote distribution
B (100%)

chaudha4 Highly Voted 4 years, 9 months ago
THe question is not about just logging AUC but logging to allow Hyperdrive to optimize hyperparameters for the AUC metric. So you must log using run instance. That way the Hyperdrive has access to that metric to compare with other runs. SO the correct answer is "No"
upvoted 18 times

Narendra05 Highly Voted 4 years, 7 months ago
run.log() is the correct answer https://docs.microsoft.com/en-us/azure/machine-learning/how-to-log-view-metrics
upvoted 9 times

evangelist Most Recent 1 year, 8 months ago
Get the current run context run = Run.get_context() # Log the AUC score run.log("AUC", auc)
upvoted 1 times

synapse 3 years, 11 months ago
Copying: THe question is not about just logging AUC but logging to allow Hyperdrive to optimize hyperparameters for the AUC metric. So you must log using run instance. That way the Hyperdrive has access to that metric to compare with other runs. SO the correct answer is "No"
upvoted 1 times

azurecert2021 4 years, 7 months ago

question is about "You need to add logging to the script to allow Hyperdrive to optimize hyperparameters for the AUC metric." so if we go through following links we use run.log to log np.float(reg) whereas printf is used for general debugging. # Train a logistic regression model print('Training a logistic regression model with regularization rate of', reg) run.log('Regularization Rate', np.float(reg)) model = LogisticRegression(C=1/reg, solver="liblinear").fit(X_train, y_train) https://github.com/MicrosoftLearning/DP100/blob/master/08A%20-%20Tuning%20Hyperparameters.ipynb https://sites.google.com/view/raybellwaves/courses/build-ai-solutions-with-azure-machine-learning

upvoted 2 times

anjurad 4 years, 9 months ago

for hyperdrive to optimise, it has to extract the chosen metric from the experiment run, through what has been logged. The log name has to match the primary metric name specified in config. The values aren't being logged in the example script - and printing doesn't capture the key/value pairs required to do the matching and comparison

upvoted 3 times

levm39 4 years, 10 months ago

the print statement can be used to debug, but in this piece of code you are only printing np.float(AUC), so you are only printing the conversion of a value to float, you are not printing any debugging information from the algorithm.

upvoted 4 times

dev2dev 4 years, 11 months ago

Answer is Yes. We can use either logger or print as per the referenced document print(val) logging.info(message)

upvoted 3 times

dev2dev 4 years, 11 months ago

Answer is Yes. We can use either logger or print as per the referenced document print(val) logging.info(message)yes, agree. Correct answer should be A.Indeed. https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines - under "Logging options and behaviour".but this is not for debugging but to use on hyperdriveWhen we are able to debug, will you not be able to extract that as well ? documentation says, its logged to -Driver logs, Azure Machine Learning designer

upvoted 3 times

stonefl 4 years, 11 months ago

yes, agree. Correct answer should be A.Indeed. https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines - under "Logging options and behaviour".but this is not for debugging but to use on hyperdriveWhen we are able to debug, will you not be able to extract that as well ? documentation says, its logged to -Driver logs, Azure Machine Learning designer

upvoted 1 times

stonefl 4 years, 11 months ago

yes, agree. Correct answer should be A.

upvoted 1 times

Anty85 4 years, 11 months ago

Indeed. https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines - under "Logging options and behaviour".

upvoted 1 times

Anty85 4 years, 11 months ago

Indeed. https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines - under "Logging options and behaviour".but this is not for debugging but to use on hyperdriveWhen we are able to debug, will you not be able to extract that as well ? documentation says, its logged to -Driver logs, Azure Machine Learning designer

upvoted 1 times

cab123 4 years, 9 months ago

but this is not for debugging but to use on hyperdrive

upvoted 9 times

cab123 4 years, 9 months ago

but this is not for debugging but to use on hyperdriveWhen we are able to debug, will you not be able to extract that as well ? documentation says, its logged to -Driver logs, Azure Machine Learning designer

upvoted 9 times

VJPrakash 4 years, 6 months ago

When we are able to debug, will you not be able to extract that as well ? documentation says, its logged to -Driver logs, Azure Machine Learning designer

upvoted 1 times

Question #5 *Topic 3*

You use the following code to run a script as an experiment in Azure Machine Learning:
You must identify the output files that are generated by the experiment run.
You need to add code to retrieve the output file names.
Which code segment should you add to the script?

```
from azureml.core import Workspace, Experiment, Run
from azureml.core import RunConfig, ScriptRunConfig
ws = Workspace.from_config()
run_config = RunConfiguration()
run_config.target='local'
script_config = ScriptRunConfig(source_directory='./script', script='experiment.py', run_config=run_config)
experiment = Experiment(workspace=ws, name='script experiment')
run = experiment.submit(config=script_config)
run.wait_for_completion()
```

A. files = run.get_properties()

B. files= run.get_file_names()

C. files = run.get_details_with_logs()

D. files = run.get_metrics()

E. files = run.get_details()

Suggested Answer: *B*
You can list all of the files that are associated with this run record by called run.get_file_names() Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-track-experiments

Community vote distribution
B (100%)

ljljljlj Highly Voted 4 years, 1 month ago
On exam 2021/7/10
upvoted 5 times

fvil Highly Voted 2 years, 9 months ago
On exam 07/11/2022
upvoted 5 times

damaldon Most Recent 2 years, 1 month ago
Correct. https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.run.run?view=azure-ml-py
upvoted 1 times

Yuriy_Ch 2 years, 5 months ago
on exam 07/March/2023
upvoted 5 times

racnaoamo 3 years, 3 months ago
similar question on 18-5-22
upvoted 4 times

synapse 3 years, 5 months ago
given answer is correct. https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run.run?view=azure-ml-py#get-file-names
upvoted 2 times

hargur 3 years, 10 months ago
on 19Oct2021
upvoted 3 times

kisskeo 3 years, 10 months ago
On Exam 01 Oct 2021
upvoted 3 times

mthombenindhl84 3 years, 11 months ago
on exam 11/9/2021

upvoted 4 times

dushmantha 3 years, 11 months ago

On exam 2021/08/31

upvoted 3 times

Abdulraoufhakeem 4 years, 1 month ago

100% correct.

upvoted 4 times

azurecert2021 4 years, 1 month ago

yes given answer is correct https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run.run?view=azure-ml-py#get-file-names--

upvoted 4 times

HkIsCrazY 4 years, 1 month ago

The answer is correct! view the link: https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run.run?view=azure-ml-py

upvoted 3 times

SaulG 4 years, 2 months ago

B is the correct answer!

upvoted 3 times

Question #6 *Topic 3*

You write five Python scripts that must be processed in the order specified in Exhibit A `" which allows the same modules to run in parallel, but will wait for modules with dependencies.

You must create an Azure Machine Learning pipeline using the Python SDK, because you want to script to create the pipeline to be tracked in your version control system. You have created five PythonScriptSteps and have named the variables to match the module names.

You need to create the pipeline shown. Assume all relevant imports have been done.

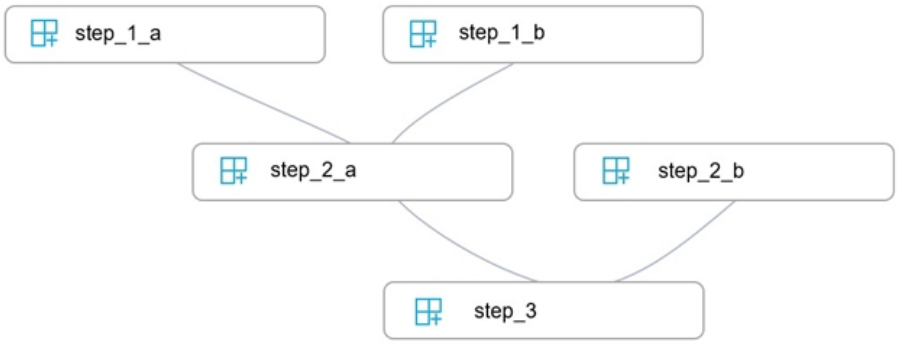
Which Python code segment should you use?

A.

B.

C.

D.




```
p = Pipeline(ws, steps=[[[[step_1_a, step_1_b], step_2_a],  step_2_b], step_3])
```

```
pipeline_steps = {
    "Pipeline": {
        "run": step_3,
        "run_after": {[
            {"run": step_2_a,
                "run_after":
                    [{"run": step_1_a},
                     {"run": step_1_b}]
                },
            {"run": step_2_b}]
        }
    }
}
p = Pipeline(ws, steps=pipeline_steps)
```

```
step_2_a.run_after(step_1_b)
step_2_a.run_after(step_1_a)
step_3.run_after(step_2_b)
step_3.run_after(step_2_a)
p = Pipeline(ws, steps=[step_3])
```

```
p = Pipeline(ws, steps=[step_1_a, step_1_b, step_2_a, step_2_b, step_3])
```

Suggested Answer:
The steps parameter is an array of steps. To build pipelines that have multiple steps, place the steps in order in this array. Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-parallel-run-step

david_Fdz Highly Voted 5 years, 4 months ago
it should be D
upvoted 33 times

chaudha4 Highly Voted 5 years, 3 months ago

Correct answer is D. The dependencies of steps is implicitly defined in the step itself (via arguments parameter that could be an output from some other step). Read https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinedata?view=azure-ml-py.

upvoted 18 times

chaudha4 5 years, 3 months ago

Correct answer is D. The dependencies of steps is implicitly defined in the step itself (via arguments parameter that could be an output from some other step). Read https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinedata?view=azure-ml-py.You could also use StepSequence but that is not an option in this question. See https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.builder.stepsequence?view=azure-ml-pyI check that same link you posted and it doesn't show in this method. This shows as an example when there's no flow control so this works for no flow control but this example seems to critically need A as the answer to control the flow

upvoted 18 times

chaudha4 5 years, 3 months ago

You could also use StepSequence but that is not an option in this question. See https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.builder.stepsequence?view=azure-ml-py

upvoted 5 times

jpalaci22 3 years, 6 months ago

I check that same link you posted and it doesn't show in this method. This shows as an example when there's no flow control so this works for no flow control but this example seems to critically need A as the answer to control the flow

upvoted 1 times

Antoh1978 Most Recent 1 year, 11 months ago

I would go for D

upvoted 1 times

deyoz 2 years, 5 months ago

I go for D, chaudha4 explanation below is spot-on.

upvoted 1 times

Plb2 2 years, 5 months ago

A would be correct if it had been p = Pipeline(ws, steps=StepSequence(steps=[initial_steps, step4])) https://learn.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.stepsequence?view=azure-ml-py#remarks

upvoted 2 times

vv_bb 2 years, 9 months ago

It should be C Check these links: 1) https://learn.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.builder.pipelinestep?view=azure-ml-py#azureml-pipeline-core-builder-pipelinestep-run-after 2) https://github.com/Azure/MachineLearningNotebooks/blob/master/how-to-use-azureml/machine-learning-pipelines/intro-to-pipelines/aml-pipelines-getting-started.ipynb -> section "Running a few steps in sequence"

upvoted 5 times

phdykd 3 years, 1 month ago

C could be

upvoted 3 times

iai 3 years, 2 months ago

A. is also wrong because it creates a non documented dependency of step_2_b on step_2_a

upvoted 1 times

ajay0011 3 years, 4 months ago

I will definitely go with D

upvoted 1 times

esimsek 3 years, 4 months ago

it should be C

upvoted 2 times

STEVTRANCE 3 years, 7 months ago

I don't want to create confusion but I was unable to find documentation about this, so I used chatGCP and it provides code samples, it seems that A is not an option cause the steps parameter should be a list of steps, not a nested list of steps. Option C the pipeline definition is missing the steps that step3 depends on, so like chaudha4 explains the dependencies are implicitly and I would go for option DI agree. I checked it with ChatGPT (for fun as well) but also verified with the https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinedata?view=azure-ml-py which is funny because someone used it earlier to justify D, which it doesn't seem to be correct via the documentation and Microsoft's Jupyter notebooks

upvoted 4 times

STEVTRANCE 3 years, 7 months ago

I don't want to create confusion but I was unable to find documentation about this, so I used chatGCP and it provides code samples, it seems that A is not an option cause the steps parameter should be a list of steps, not a nested list of steps. Option C the pipeline definition is missing the steps that step3 depends on, so like chaudha4 explains the dependencies are implicitly and I would go for option D

upvoted 4 times

jpalaci22 3 years, 6 months ago

I agree. I checked it with ChatGPT (for fun as well) but also verified with the https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinedata?view=azure-ml-py which is funny because someone used it earlier to justify D, which it doesn't seem to be correct via the documentation and Microsoft's Jupyter notebooks

upvoted 1 times

michaelmorar 3 years, 7 months ago

A gets my vote. D has absolutely no sequence and C fails to run step 2_a.

upvoted 2 times

FlexingD 3 years, 9 months ago

see the graph, so it's A

upvoted 1 times

ning 4 years, 3 months ago

A is correct, that is the syntax I used for years!Agree!

upvoted 8 times

ning 4 years, 3 months ago

A is correct, that is the syntax I used for years!

upvoted 8 times

JTWang 3 years, 10 months ago

Agree!

upvoted 1 times

Tj87 4 years, 3 months ago

D misses the StepSequence and it means it runs all the steps in parallel. Then C seems to be a better option. The only problem with C is that it doesn't run 1_a and 1_b in parallel and it runs them sequentially.

upvoted 4 times

JTWang 4 years, 4 months ago

it should be C Running a few steps in sequence https://github.com/Azure/MachineLearningNotebooks/blob/master/how-to-use-azureml/machine-learning-pipelines/intro-to-pipelines/aml-pipelines-getting-started.ipynb

upvoted 3 times

synapse 4 years, 5 months ago

I think the answer is D I think the key here is this sentence: "but will wait for modules with dependencies." This means the lines indicate data dependencies between steps not just the sequence. run_after() or any such constructs are for sequencing where there are no data dependencies between the steps. Text from the notebook linked below: We will reuse step1, step2, step3, but build the pipeline in such a way that we chain step3 after step2 and step2 after step1. Note that there is no explicit data dependency between these steps, but still steps can be made dependent by using the run_after construct. https://github.com/Azure/MachineLearningNotebooks/blob/master/how-to-use-azureml/machine-learning-pipelines/intro-to-pipelines/aml-pipelines-getting-started.ipynb

upvoted 1 times

Question #7 *Topic 3*

You create a datastore named training_data that references a blob container in an Azure Storage account. The blob container contains a folder named csv_files in which multiple comma-separated values (CSV) files are stored.

You have a script named train.py in a local folder named ./script that you plan to run as an experiment using an estimator. The script includes the following code to read data from the csv_files folder:

You have the following script.

You need to configure the estimator for the experiment so that the script can read the data from a data reference named data_ref that references the csv_files folder in the training_data datastore.

Which code should you use to configure the estimator?

A.

B.

C.

D.

E.

```
import os
import argparse
import pandas as pd

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from azureml.core import Run

run = Run.get_context()
parser = argparse.ArgumentParser()
parser.add_argument('--data-folder', type=str, dest='data_folder', help='data reference')
args = parser.parse_args()

data_folder = args.data_folder
csv_files = os.listdir(data_folder)
training_data = pd.concat((pd.read_csv(os.path.join(data_folder,csv_file)) for csv_file in csv_files))

# Code goes on to split the training data and train a logistic regression model
```

```
from azureml.core import Workspace, Datastore, Experiment
from azureml.train.sklearn import SKLearn

ws = Workspace.from_config()
exp = Experiment(workspace=ws, name='csv_training')
ds = Datastore.get(ws, datastore_name='training_data')
data_ref = ds.path('csv_files')

# Code to define estimator goes here

run = exp.submit(config=estimator)
run.wait_for_completion(show_output=True)
```

```
estimator = SKLearn(source_directory='./script',
 inputs=[data_ref.as_named_input('data-folder').to_pandas_dataframe()],
 compute_target='local',
 entry_script='train.py')
```

```
script_params = {
    '--data-folder': data_ref.as_mount()
}
estimator = SKLearn(source_directory='./script',
 script_params=script_params,
 compute_target='local',
 entry_script='train.py'
```

```
estimator = SKLearn(source_directory='./script',
 inputs=[data_ref.as_named_input('data-folder').as_mount()],
 compute_target='local',
 entry_script='train.py')
```

```
script_params = {
    '--data-folder': data_ref.as_download(path_on_compute='csv_files')
}
estimator = SKLearn(source_directory='./script',
 script_params=script_params,
 compute_target='local',
 entry_script='train.py'

estimator = SKLearn(source_directory='./script',
 inputs=[data_ref.as_named_input('data-folder').as_download(path_on_compute='csv_files')],
 compute_target='local',
 entry_script='train.py')
```

Suggested Answer:
Besides passing the dataset through the input parameters in the estimator, you can also pass the dataset through script_params and get the data path (mounting point) in your training script via arguments. This way, you can keep your training script independent of azureml-sdk. In other words, you will be able use the same training script for local debugging and remote training on any cloud platform. Example: from azureml.train.sklearn import SKLearn script_params = { # mount the dataset on the remote compute and pass the mounted path as an argument to the training script '--data-folder': mnist_ds.as_named_input('mnist').as_mount(), '--regularization': 0.5 } est = SKLearn(source_directory=script_folder, script_params=script_params, compute_target=compute_target, environment_definition=env, entry_script='train_mnist.py') # Run the experiment run = experiment.submit(est) run.wait_for_completion(show_output=True) Incorrect Answers: A: Pandas DataFrame not used. Reference: https://docs.microsoft.com/es-es/azure/machine-learning/how-to-train-with-datasets

chaudha4 Highly Voted 4 years, 3 months ago
The use of estimator is deprecated. Use the ScriptRunConfig object with your own defined environment. Hope we don't see this question going forward !! https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.estimator.estimator?view=azure-ml-pyYou're right, but if you replace the estimator with the ScriptRunConfig this question still holds, as the method to pass Dataset, mount vs. download, by argument, etc.. are relevant
upvoted 13 times

chaudha4 Highly Voted 4 years, 3 months ago
The use of estimator is deprecated. Use the ScriptRunConfig object with your own defined environment. Hope we don't see this question going forward !!

https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.estimator.estimator?view=azure-ml-py
upvoted 13 times

scipio 4 years, 3 months ago
You're right, but if you replace the estimator with the ScriptRunConfig this question still holds, as the method to pass Dataset, mount vs. download, by argument, etc.. are relevant
upvoted 5 times

vv_bb Most Recent 1 year, 9 months ago
Even though the Estimator is deprecated in favor for ScriptRunConfig (google - "Migrating from Estimators to ScriptRunConfig") , I tried to understand the correct answer for the question as it is defined here. 1) For Estimator class both "script_params" and "arguments" parameters are acceptable check here - https://learn.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.estimator.estimator?view=azure-ml-py 2) So how to define which of them is valid in our case? The answer is here: (be aware for PythonScriptStep "arguments" is the same as "script_params" for Estimator) https://learn.microsoft.com/en-us/azure/machine-learning/how-to-move-data-in-out-of-pipelines?view=azureml-api-1#access-datasets-within-your-script Meaning because in our script we use the ArgParser we have to pass the dataset using the "script_params"
upvoted 3 times

iai 2 years, 2 months ago
Shouldn't it be D.? for local compute_target not sure if as_mount will work. better as_download
upvoted 2 times

danishanis 2 years, 5 months ago
Answer is B. I typed the question as it is in ChatGPT and it gave the answer where the 'script_params' argument is configured to read data from 'data_ref' (and data_ref.as_mount() is being used to specify the file path in datastore) that references a 'csv_files' folder.
upvoted 2 times

jpalaci22 2 years, 5 months ago
Seen on the exam 20Feb2023

upvoted 3 times

Edriv 2 years, 8 months ago

can be A,C,E - what do you thing?

upvoted 1 times

ning 3 years, 3 months ago

B should be correct!

upvoted 3 times

TheYazan 3 years, 5 months ago

on march 2022

upvoted 4 times

kisskeo 3 years, 10 months ago

On Exam 01 Oct 2021

upvoted 3 times

ljljljlj 4 years, 1 month ago

On exam 2021/7/10

upvoted 3 times

sarahmoin 4 years, 2 months ago

what is the correct answer? Why its not D.as_download, which copies the files to a temporary location on the compute where the script is being run. as_mount to stream the files directly from their source.Notice however, that compute target is local, will mounting work?

upvoted 1 times

sarahmoin 4 years, 2 months ago

what is the correct answer? Why its not D.

upvoted 1 times

vhx 4 years, 2 months ago

as_download, which copies the files to a temporary location on the compute where the script is being run. as_mount to stream the files directly from their source.

upvoted 3 times

vhx 4 years, 2 months ago

as_download, which copies the files to a temporary location on the compute where the script is being run. as_mount to stream the files directly from their source.Notice however, that compute target is local, will mounting work?

upvoted 3 times

iai 2 years, 2 months ago

Notice however, that compute target is local, will mounting work?

upvoted 1 times

iuolu 4 years, 3 months ago

Nobody checked this question? The answer should be A, using to_pandas_dataframe() for tabular files instead

upvoted 2 times

iuolu 4 years, 3 months ago

Nobody checked this question? The answer should be A, using to_pandas_dataframe() for tabular files insteadNo, you are wrong. Several problems in A. 1) Parameter is being passed as named input. That is wrong since it is not being accessed using named input in t he script. 2) You convert to dataframe in the script not when you pass it. So A is definitely not the correct answer.

upvoted 2 times

chaudha4 4 years, 3 months ago

No, you are wrong. Several problems in A. 1) Parameter is being passed as named input. That is wrong since it is not being accessed using named input in t he script. 2) You convert to dataframe in the script not when you pass it. So A is definitely not the correct answer.

upvoted 10 times

DRAG DROP -

You create a multi-class image classification deep learning experiment by using the PyTorch framework. You plan to run the experiment on an Azure Compute cluster that has nodes with GPU's.

You need to define an Azure Machine Learning service pipeline to perform the monthly retraining of the image classification model. The pipeline must run with minimal cost and minimize the time required to train the model.

Which three pipeline steps should you run in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Select and Place:

Actions

- Configure a DataTransferStep() to fetch new image data from public web portal, running on the cpu-compute compute target.
- Configure an EstimatorStep() to run an estimator that runs the bird_classifier_train.py model training script on the gpu_compute compute target.
- Configure a PythonScriptStep() to run both image_fetcher.py and image_resize.py on the cpu-compute compute target.
- Configure an EstimatorStep() to run an estimator that runs the bird_classifier_train.py model training script on the cpu_compute compute target.
- Configure a PythonScriptStep() to run image_fetcher.py on the cpu-compute compute target.
- Configure a PythonScriptStep() to run image_resize.py on the cpu-compute compute target.
- Configure a PythonScriptStep() to run bird_classifier_train.py on the cpu-compute compute target.
- Configure a PythonScriptStep() to run bird_classifier_train.py on the gpu-compute compute target.

Answer Area

Suggested Answer:

Actions	Answer Area
Configure a DataTransferStep() to fetch new image data from public web portal, running on the cpu-compute compute target.	Configure a DataTransferStep() to fetch new image data from public web portal, running on the cpu-compute compute target.
Configure an EstimatorStep() to run an estimator that runs the bird_classifier_train.py model training script on the gpu_compute compute target.	Configure a PythonScriptStep() to run image_resize.py on the cpu-compute compute target.
Configure a PythonScriptStep() to run both image_fetcher.py and image_resize.py on the cpu-compute compute target.	Configure an EstimatorStep() to run an estimator that runs the bird_classifier_train.py model training script on the gpu_compute compute target.
Configure an EstimatorStep() to run an estimator that runs the bird_classifier_train.py model training script on the cpu_compute compute target.	
Configure a PythonScriptStep() to run image_fetcher.py on the cpu-compute compute target.	
Configure a PythonScriptStep() to run image_resize.py on the cpu-compute compute target.	
Configure a PythonScriptStep() to run bird_classifier_train.py on the cpu-compute compute target.	
Configure a PythonScriptStep() to run bird_classifier_train.py on the gpu-compute compute target.	

Actions	Answer Area
Configure a DataTransferStep() to fetch new image data from public web portal, running on the cpu-compute compute target.	Configure a DataTransferStep() to fetch new image data from public web portal, running on the cpu-compute compute target.
Configure an EstimatorStep() to run an estimator that runs the bird_classifier_train.py model training script on the gpu_compute compute target.	Configure a PythonScriptStep() to run image_resize.py on the cpu-compute compute target.
Configure a PythonScriptStep() to run both image_fetcher.py and image_resize.py on the cpu-compute compute target.	Configure an EstimatorStep() to run an estimator that runs the bird_classifier_train.py model training script on the gpu_compute compute target.
Configure an EstimatorStep() to run an estimator that runs the bird_classifier_train.py model training script on the cpu_compute compute target.	
Configure a PythonScriptStep() to run image_fetcher.py on the cpu-compute compute target.	
Configure a PythonScriptStep() to run image_resize.py on the cpu-compute compute target.	
Configure a PythonScriptStep() to run bird_classifier_train.py on the cpu-compute compute target.	
Configure a PythonScriptStep() to run bird_classifier_train.py on the gpu-compute compute target.	

Step 1: Configure a DataTransferStep() to fetch new image dataג€¦ Step 2: Configure a PythonScriptStep() to run image_resize.y on the cpu-compute compute target. Step 3: Configure the EstimatorStep() to run training script on the gpu_compute computer target. The PyTorch estimator provides a simple way of launching a PyTorch training job on a compute target. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-pytorch

chaudha4 Highly Voted 4 years, 9 months ago

Hope we don't see this question in exam since it is covering topics that are all deprecated.

upvoted 17 times

ning Highly Voted 3 years, 9 months ago

I think the answer is correct, though it might not in the scope of test any more 1. Fetch whole / big image 2. Break into small / tile images 3. Run training models

upvoted 6 times

haby Most Recent 2 years, 1 month ago

EstimatorStep is deprecated, it's recommended to use PythonSriptStep or CommandStep. For now(2023-12), I will take : 1. Configure a PythonScriptStep() to run image_fetcher.py on the cpu-compute compute target. 2. Configure a PythonScriptStep() to run image_resize.py on the cpu-compute compute target. 3. Configure an PythonScriptStep() to run an estimator that runs the bird_classifier_train.py model training script on the gpu_compute compute target.

upvoted 4 times

NullVoider_0 2 years, 2 months ago

Based on the requirements to minimize cost and training time for the image classification retraining pipeline, the three steps that should be configured are: 1. Configure a PythonScriptStep() to run image_fetcher.py on the cpu-compute compute target. 2. Configure a PythonScriptStep() to run image_resize.py on the cpu-compute compute target. 3. Configure an EstimatorStep() to run an estimator that runs the bird_classifier_train.py model training script on the gpu_compute compute target.

upvoted 1 times

giusecozza 3 years, 5 months ago

I'm afraid the first step is wrong, since DataTransferStep is used only to move data between datastores, not fetching data from the web. https://docs.microsoft.com/en-us/python/api/azureml-pipeline-steps/azureml.pipeline.steps.data_transfer_step.datatransferstep?view=azure-ml-py

upvoted 3 times

chevyli 3 years, 5 months ago

Look correct

upvoted 3 times

luisbenitez_14 3 years, 8 months ago

Estimator is deprecated. PythonScript is the correct 3rd step

upvoted 1 times

Question #9 *Topic 3*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
An IT department creates the following Azure resource groups and resources:
The IT department creates an Azure Kubernetes Service (AKS)-based inference compute target named aks-cluster in the Azure Machine Learning workspace.
You have a Microsoft Surface Book computer with a GPU. Python 3.6 and Visual Studio Code are installed.
You need to run a script that trains a deep neural network (DNN) model and logs the loss and accuracy metrics.
Solution: Attach the mlvm virtual machine as a compute target in the Azure Machine Learning workspace. Install the Azure ML SDK on the Surface Book and run
Python code to connect to the workspace. Run the training script as an experiment on the mlvm remote compute resource.
Does the solution meet the goal?

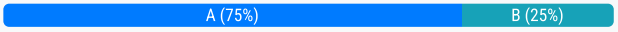
Resource group	Resources
ml_resources	• an Azure Machine Learning workspace named amlworkspace • an Azure Storage account named amlworkspace12345 • an Application Insights instance named amlworkspace54321 • an Azure Key Vault named amlworkspace67890 • an Azure Container Registry named amlworkspace09876
general_compute	A virtual machine named mlvm with the following configuration: • Operating system: Ubuntu Linux • Software installed: Python 3.6 and Jupyter Notebooks • Size: NC6 (6 vCPUs, 1 vGPU, 56 Gb RAM)

A. Yes

B. No

Suggested Answer: *A*
Use the VM as a compute target. Note: A compute target is a designated compute resource/environment where you run your training script or host your service deployment. This location may be your local machine or a cloud-based compute resource. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target

Community vote distribution



bkuchi Highly Voted 4 years, 8 months ago
Question was there in June 2021 Exam
upvoted 9 times

Yuriy_Ch Highly Voted 2 years, 11 months ago
Exactly this question was on exam 07/March/2023
upvoted 5 times

sl_mslconsulting Most Recent 1 year, 9 months ago
Remote VMs are considered as a unmanaged and can require extra steps for you to maintain or to improve performance for machine learning workloads. I interpret this as you won't be able to log the metrics out-of-the box. Link: require extra steps for you to maintain or to improve performance for machine learning workloads. https://learn.microsoft.com/en-us/azure/machine-learning/concept-compute-target?view=azureml-api-2#unmanaged-compute
upvoted 1 times

MarinaMijailovic 2 years, 8 months ago
Isn't inference cluster used for deployment and not for training?
upvoted 2 times

MarinaMijailovic 2 years, 8 months ago
Isn't inference cluster used for deployment and not for training?I believe what's currently recommended as the compute resource for deployment is a Kubernetes compute, but the question is simply asking if the given compute can work, not whether it's the most optimal.
upvoted 2 times

Deathking15 2 years, 3 months ago

I believe what's currently recommended as the compute resource for deployment is a Kubernetes compute, but the question is simply asking if the given compute can work, not whether it's the most optimal.

upvoted 3 times

krishna1818 2 years, 8 months ago

existing VM has to be attached

upvoted 1 times

robotcop 3 years, 2 months ago

no objection

upvoted 1 times

ning 3 years, 9 months ago

Attach existing VM as a target, sounds correct

upvoted 1 times

synapse 3 years, 11 months ago

Given answer is correct

upvoted 1 times

ranjsi01 4 years, 1 month ago

correct

upvoted 1 times

TheCyanideLancer 4 years, 1 month ago

Is the given answer correct?

upvoted 2 times

hargur 4 years, 4 months ago

on 19Oct2021

upvoted 2 times

snsnsnsn 4 years, 5 months ago

on exam 2/9/21

upvoted 3 times

dushmantha 4 years, 5 months ago

On exam 2021/08/31

upvoted 3 times

Question #10 *Topic 3*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
An IT department creates the following Azure resource groups and resources:
The IT department creates an Azure Kubernetes Service (AKS)-based inference compute target named aks-cluster in the Azure Machine Learning workspace.
You have a Microsoft Surface Book computer with a GPU. Python 3.6 and Visual Studio Code are installed.
You need to run a script that trains a deep neural network (DNN) model and logs the loss and accuracy metrics.
Solution: Install the Azure ML SDK on the Surface Book. Run Python code to connect to the workspace and then run the training script as an experiment on local compute.
Does the solution meet the goal?

Resource group	Resources
ml_resources	• an Azure Machine Learning workspace named amlworkspace • an Azure Storage account named amlworkspace12345 • an Application Insights instance named amlworkspace54321 • an Azure Key Vault named amlworkspace67890 • an Azure Container Registry named amlworkspace09876
general_compute	A virtual machine named mlvm with the following configuration: • Operating system: Ubuntu Linux • Software installed: Python 3.6 and Jupyter Notebooks • Size: NC6 (6 vCPUs, 1 vGPU, 56 Gb RAM)

A. Yes

B. No

Suggested Answer: *B*
Need to attach the mlvm virtual machine as a compute target in the Azure Machine Learning workspace. Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target

Community vote distribution
A (80%) B (20%)

dev2dev Highly Voted 4 years, 11 months ago
we can run locally. anwer is yes
upvoted 24 times

Manishi_VBA Highly Voted 4 years, 9 months ago
The answer should be yes. The local machine does have the capability to run DNN
upvoted 10 times

jmh106 Most Recent 7 months, 3 weeks ago
A ES LA CORRECTA
upvoted 1 times

f82411e 1 year, 2 months ago
la respuesta es no
upvoted 1 times

evangelist 1 year, 8 months ago
answer is A, yes, AKS is used as inference compute target, inference means hosting and deploying the model not training the model
upvoted 1 times

sl_mslconsulting 1 year, 9 months ago
Even the doc says you can use Local computer as a training target, but I can't find any detailed steps to do this. Besides, it will still be unmanaged compute and I doubt you can log the metrics that can be utilized to improve the performance of the model.ChatGPT said I was wrong. The answer should be A.
upvoted 1 times

sl_mslconsulting 1 year, 9 months ago

Selected Answer: B

Even the doc says you can use Local computer as a training target, but I can't find any detailed steps to do this. Besides, it will still be unmanaged compute and I doubt you can log the metrics that can be utilized to improve the performance of the model.

upvoted 1 times

sl_mslconsulting 1 year, 9 months ago

ChatGPT said I was wrong. The answer should be A.

upvoted 1 times

PI_Team 2 years, 6 months ago

The question only asks about running a script that trains a DNN model and logs the loss and accuracy metrics. In that case, the solution does meet the goal. Installing the Azure ML SDK on the Surface Book and running the training script as an experiment on local compute would allow you to train a DNN model and log the loss and accuracy metrics. SaM

upvoted 1 times

krishna1818 2 years, 8 months ago

yes we can run the DNN locally

upvoted 2 times

Yuriy_Ch 2 years, 11 months ago

Exactly this question was on exam 07/March/2023

upvoted 2 times

slcheng 3 years, 3 months ago

Ans : A

upvoted 1 times

majma 3 years, 8 months ago

if we should log the metrics into the workspace then answer is correct

upvoted 1 times

ning 3 years, 9 months ago

This question is so poorly written ... Can we run local DNN, yes, we can ... Does that use the VM listed, no it is not ... As question is asking for whether we can run ... So I will vote for YES

upvoted 2 times

JoshuaXu 4 years, 3 months ago

on 6 Nov 2021, the answer should be "yes"

upvoted 3 times

hargur 4 years, 4 months ago

on 19Oct2021

upvoted 1 times

kail85 4 years, 7 months ago

No indication if cudatoolkit is installed. Just have python and gpu is not enough to run dnn.technically you don't even need a GPU to run a DNN. Its just faster with one, but having python and a cpu is enough to run a DNN in normal cases.

upvoted 3 times

kail85 4 years, 7 months ago

No indication if cudatoolkit is installed. Just have python and gpu is not enough to run dnn.

upvoted 3 times

lewitt 3 years, 5 months ago

technically you don't even need a GPU to run a DNN. Its just faster with one, but having python and a cpu is enough to run a DNN in normal cases.

upvoted 2 times

treadst0ne 4 years, 8 months ago

The answer should be YES, you can log metrics running locally without attaching the computer. The only case to attach your computer, is to make it available as a ComputeTarget when running experiments.

upvoted 4 times

chaudha4 4 years, 9 months ago

The answer is Yes. The surface book has a GPU and even though there is no mention of RAM, it seems like the surface book is capable of training a DNN model.

upvoted 4 times

Question #11 *Topic 3*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
An IT department creates the following Azure resource groups and resources:
The IT department creates an Azure Kubernetes Service (AKS)-based inference compute target named aks-cluster in the Azure Machine Learning workspace.
You have a Microsoft Surface Book computer with a GPU. Python 3.6 and Visual Studio Code are installed.
You need to run a script that trains a deep neural network (DNN) model and logs the loss and accuracy metrics.
Solution: Install the Azure ML SDK on the Surface Book. Run Python code to connect to the workspace. Run the training script as an experiment on the aks- cluster compute target.
Does the solution meet the goal?

Resource group	Resources
ml_resources	• an Azure Machine Learning workspace named amlworkspace • an Azure Storage account named amlworkspace12345 • an Application Insights instance named amlworkspace54321 • an Azure Key Vault named amlworkspace67890 • an Azure Container Registry named amlworkspace09876
general_compute	A virtual machine named mlvm with the following configuration: • Operating system: Ubuntu Linux • Software installed: Python 3.6 and Jupyter Notebooks • Size: NC6 (6 vCPUs, 1 vGPU, 56 Gb RAM)

A. Yes

B. No

Suggested Answer: *B*
Need to attach the mlvm virtual machine as a compute target in the Azure Machine Learning workspace. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target

Community vote distribution
B (100%)

chaudha4 Highly Voted 4 years, 10 months ago
A very long question trying to ask if you can use an inference cluster for training purpose !!
upvoted 19 times

gamezone25 Highly Voted 4 years, 10 months ago
B is indeed the correct answer. AKS is for inference, not for training
upvoted 13 times

evangelist Most Recent 1 year, 8 months ago
answer is A, yes, AKS is used as inference compute target, inference means hosting and deploying the model not training the model
upvoted 1 times

a6cb3b0 1 year, 10 months ago
read this: Azure Machine Learning Kubernetes compute supports two kinds of Kubernetes cluster: AKS cluster in Azure. With your self-managed AKS cluster in Azure, you can gain security and controls to meet compliance requirement and flexibility to manage teams' ML workload. Arc Kubernetes cluster outside of Azure. With Arc Kubernetes cluster, you can train or deploy models in any infrastructure on-premises, across multicloud, or the edge. Since this question menthioned we have AKS Cluster in Azure, so it is not possible to train the model. the correct answer is B.
upvoted 1 times

Matt2000 2 years ago
AKS clusters can process training and allow for gpu usage. So it seems feasible. Reference: https://learn.microsoft.com/en-us/azure/machine-learning/how-to-attach-kubernetes-anywhere?view=azureml-api-2
upvoted 1 times

Yuriy_Ch 2 years, 11 months ago
Exactly this question was on exam 07/March/2023

upvoted 4 times

ning 3 years, 9 months ago

No need to over thinking ... AKS for deploying large scale applications ... No where in documentation mentioned as training target ... In practically, you can certainly train data on AKS, LOL ...

upvoted 1 times

synapse 3 years, 11 months ago

aks is for inferencing not for training

upvoted 1 times

hargur 4 years, 4 months ago

on 19Oct2021

upvoted 1 times

lander_c 4 years, 5 months ago

The answer should be No. But note the AKS has a preview feature that allows it to train models as well. Check whether it has been moved out of the preview feature set.

upvoted 3 times

snsnsnsn 4 years, 5 months ago

on exam 2/9/21

upvoted 1 times

HOTSPOT -

You plan to use Hyperdrive to optimize the hyperparameters selected when training a model. You create the following code to define options for the hyperparameter experiment:

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

Hot Area:

```
import azureml.train.hyperdrive.parameter_expressions as pe
from azureml.train.hyperdrive import GridParameterSampling, HyperDriveConfig

param_sampling = GridParameterSampling({
  "max_depth" : pe.choice(6, 7, 8, 9),
  "learning_rate" : pe.choice(0.05, 0.1, 0.15)
  })
hyperdrive_run_config = HyperDriveConfig(
  estimator = estimator,
  hyperparameter_sampling = param_sampling,
  policy = None,
  primary_metric_name = "auc",
  primary_metruc_goal = PrimaryMetricGoal.MAXIMIZE,
  max_total_runs = 50,
  max_concurrent_runs = 4)
```

Answer Area

	Yes	No
There will be 50 runs for this hyperparameter tuning experiment.	○	○
You can use the policy parameter in the HyperDriveConfig class to specify a security policy.	○	○
The experiment will create a run for every possible value for the learning rate parameter between 0.05 and 0.15.	○	○

Suggested Answer:

Answer Area

	Yes	No
There will be 50 runs for this hyperparameter tuning experiment.	○	○
You can use the policy parameter in the HyperDriveConfig class to specify a security policy.	○	○
The experiment will create a run for every possible value for the learning rate parameter between 0.05 and 0.15.	○	○

Answer Area

	Yes	No
There will be 50 runs for this hyperparameter tuning experiment.	○	○
You can use the policy parameter in the HyperDriveConfig class to specify a security policy.	○	○
The experiment will create a run for every possible value for the learning rate parameter between 0.05 and 0.15.	○	○

Box 1: No - max_total_runs (50 here) The maximum total number of runs to create. This is the upper bound; there may be fewer runs when the sample space is smaller than this value.

Box 2: Yes -

Policy EarlyTerminationPolicy - The early termination policy to use. If None - the default, no early termination policy will be used.

Box 3: No - Discrete hyperparameters are specified as a choice among discrete values. choice can be: ➪ one or more comma-separated values ➪ a range object ➪ any arbitrary list object Reference: https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.hyperdriveconfig https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters

ExamTaker177 Highly Voted 4 years, 11 months ago

Box 2 should be NO as 'Policy' parameter is for early termination policy not security policy.

upvoted 40 times

ExamTaker177 Highly Voted 4 years, 11 months ago

Box 2 should be NO as 'Policy' parameter is for early termination policy not security policy.agree 3 NO

upvoted 40 times

kty 4 years, 10 months ago

agree 3 NO

upvoted 14 times

JB246 4 years, 9 months ago

So Answer could be 3 NO s

upvoted 16 times

NullVoider_0 2 years, 2 months ago

Please verify the information in the solution section to this question. The actual answer is 1. Yes. The max_total_runs parameter is set to 50, so the hyperparameter tuning experiment will have up to 50 total iterations. 2. No. The policy parameter in HyperDriveConfig relates to the early termination policy, not security policies. It defines stopping rules for low-performing runs. 3. No. The learning rate parameter uses GridParameterSampling with discrete values specified (0.05, 0.1, 0.15). So only 3 options listed would be searched, not the full range.

upvoted 4 times

MarinaMijailovic 2 years, 7 months ago

Can someone explain why thrid box is a no when grid sampling tries every possible option?because there are only three options defined: 0.05, 0.1, 0.15 and they will be tested. Otherwise, there is an infinite number of options between 0.05 and 0.15

upvoted 1 times

MarinaMijailovic 2 years, 7 months ago

Can someone explain why thrid box is a no when grid sampling tries every possible option?

upvoted 1 times

Lapiduse 2 years, 6 months ago

because there are only three options defined: 0.05, 0.1, 0.15 and they will be tested. Otherwise, there is an infinite number of options between 0.05 and 0.15

upvoted 1 times

vish9 2 years, 9 months ago

First box Yes. Second Box No, Third Box No. For the first box, the total number of combinations of max_depth and learning rate is 81. However runs will be capped to 50, by max_runs parameter.

upvoted 1 times

vish9 2 years, 9 months ago

First box Yes. Second Box No, Third Box No. For the first box, the total number of combinations of max_depth and learning rate is 81. However runs will be capped to 50, by max_runs parameter.how it is 81 combinations, shouldn't be just 12 (4*3)

upvoted 1 times

deyoz Most Recent 1 year, 11 months ago

how it is 81 combinations, shouldn't be just 12 (4*3)

upvoted 1 times

ZIMARAKI 3 years ago

Its No, No and No

upvoted 2 times

ning 3 years, 9 months ago

No, No, No, Policy = Early Termination Policy

upvoted 4 times

racnaoamo 3 years, 9 months ago

similar question on 18-5-22

upvoted 1 times

kevinvbc 3 years, 9 months ago

Box 2 should be NO

upvoted 1 times

TheYazan 3 years, 11 months ago

On march 2022

upvoted 3 times

tom2hoso 4 years, 4 months ago

Box1 No Box2 No Box3 No

upvoted 8 times

dev2dev 4 years, 11 months ago

very funny answer. box 1 is yes, 2 is no, 3 is no

upvoted 10 times

dev2dev 4 years, 11 months ago

very funny answer. box 1 is yes, 2 is no, 3 is nobox 1 is no which is correct, because its max limit, expierment can finish before thatBox 1 should be yes as no early termination policy has been set.Yet there are only 12 options to test for the grid sampling.Correct, since there are only 12 options to test, there will only be 12 options (sample space is smaller than the max_total_runs). https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.hyperdriveconfig?view=azure-ml-py#parameters

upvoted 10 times

dev2dev 4 years, 10 months ago

box 1 is no which is correct, because its max limit, expierment can finish before thatBox 1 should be yes as no early termination policy has been set.Yet there are only 12 options to test for the grid sampling.Correct, since there are only 12 options to test, there will only be 12 options (sample space is smaller than the max_total_runs). https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.hyperdriveconfig?view=azure-ml-py#parameters

upvoted 7 times

dev2dev 4 years, 10 months ago

box 1 is no which is correct, because its max limit, expierment can finish before that

upvoted 7 times

Dasist 4 years, 10 months ago

Box 1 should be yes as no early termination policy has been set.

upvoted 8 times

Dasist 4 years, 10 months ago

Box 1 should be yes as no early termination policy has been set.Yet there are only 12 options to test for the grid sampling.Correct, since there are only 12 options to test, there will only be 12 options (sample space is smaller than the max_total_runs). https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.hyperdriveconfig?view=azure-ml-py#parameters

upvoted 8 times

Anty85 4 years, 10 months ago

Yet there are only 12 options to test for the grid sampling.Correct, since there are only 12 options to test, there will only be 12 options (sample space is smaller than the max_total_runs). https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.hyperdriveconfig?view=azure-ml-py#parameters

upvoted 14 times

Anty85 4 years, 10 months ago

Yet there are only 12 options to test for the grid sampling.

upvoted 14 times

gamezone25 4 years, 10 months ago

Correct, since there are only 12 options to test, there will only be 12 options (sample space is smaller than the max_total_runs). https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.hyperdriveconfig?view=azure-ml-py#parameters

upvoted 5 times

BleadFast 4 years, 11 months ago

box 2 is No https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.hyperdriveconfig?view=azure-ml-py Policy: The early termination policy to use. If None - the default, no early termination policy will be used.

upvoted 5 times

Question #13 *Topic 3*

HOTSPOT -

You are using Azure Machine Learning to train machine learning models. You need a compute target on which to remotely run the training script.

You run the following Python code:

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

Hot Area:

```
from azureml.core.compute import ComputeTarget, AmlCompute
from azureml.core.compute_target import ComputeTargetException
the_cluster_name = "NewCompute"
config = AmlCompute.provisioning_configuration(vm_size= 'STANDARD_D2', max_nodes=3)
the_cluster = ComputeTarget.create(ws, the_cluster_name, config)
```

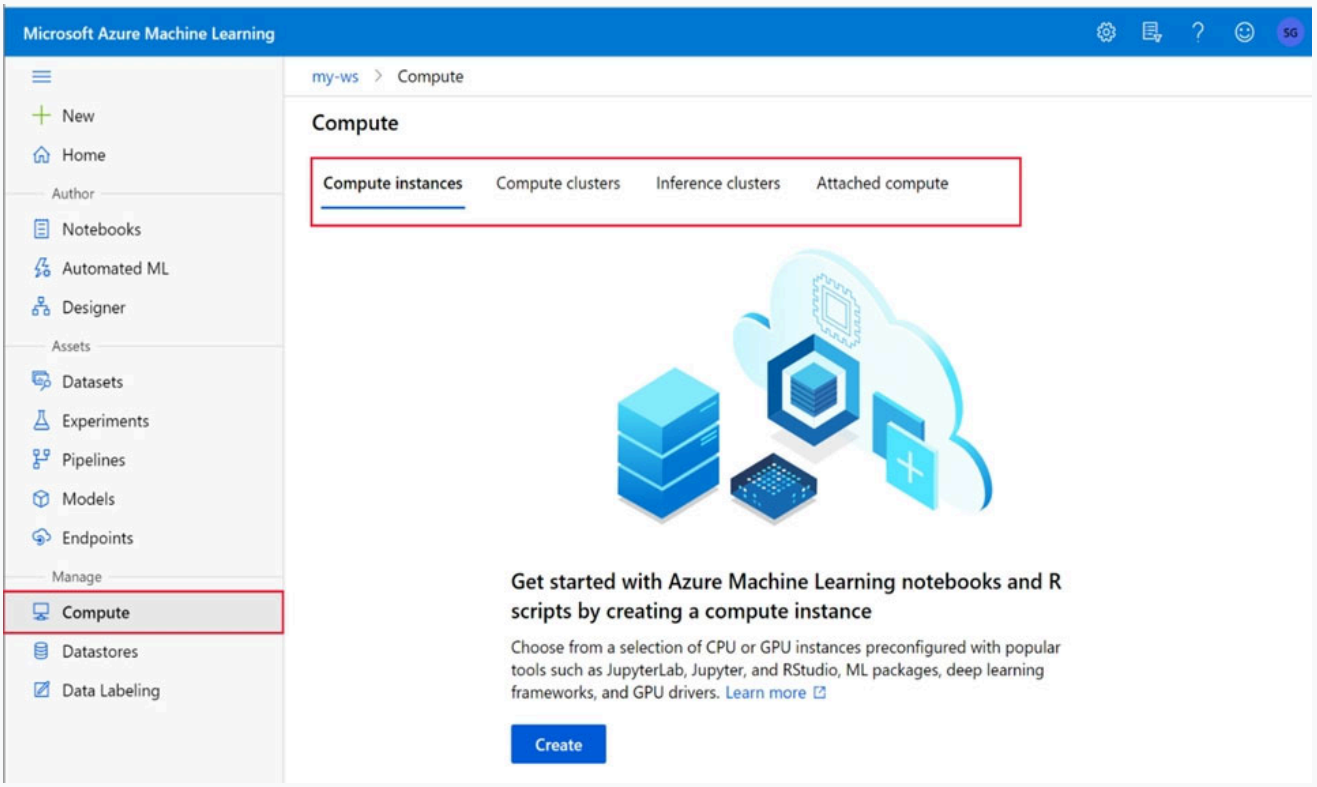
Answer Area

	Yes	No
The compute is created in the same region as the Machine Learning service workspace.	○	○
The compute resource created by the code is displayed as a compute cluster in Azure Machine Learning studio.	○	○
The minimum number of nodes will be zero.	○	○

Suggested Answer:

Answer Area

	Yes	No
The compute is created in the same region as the Machine Learning service workspace.	○	○
The compute resource created by the code is displayed as a compute cluster in Azure Machine Learning studio.	○	○
The minimum number of nodes will be zero.	○	○



Answer Area

	Yes	No
The compute is created in the same region as the Machine Learning service workspace.	○	○
The compute resource created by the code is displayed as a compute cluster in Azure Machine Learning studio.	○	○
The minimum number of nodes will be zero.	○	○

Box 1: Yes - The compute is created within your workspace region as a resource that can be shared with other users.

Box 2: Yes - It is displayed as a compute cluster.

View compute targets - 1. To see all compute targets for your workspace, use the following steps: 2. Navigate to Azure Machine Learning studio. 3. Under Manage, select Compute. 4. Select tabs at the top to show each type of compute target.

Box 3: Yes - min_nodes is not specified, so it defaults to 0. Reference: https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute.amlcomputeprovisioningconfiguration https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-studio

trickerk Highly Voted 4 years, 6 months ago

Given answer is correct!

upvoted 11 times

ljljljlj 4 years, 7 months ago

On exam 2021/7/10

upvoted 5 times

Matt2000 2 years ago

2 should be no. 'AmlCompute' refers to compute clusters, not instances: https://learn.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-cluster?view=azureml-api-2&tabs=python Instances require 'ComputeInstance', see https://learn.microsoft.com/en-us/azure/machine-learning/how-to-create-compute-instance?view=azureml-api-2&tabs=pythonMy mistake. The answer should be 'yes'

upvoted 1 times

Matt2000 2 years ago

2 should be no. 'AmlCompute' refers to compute clusters, not instances: https://learn.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-cluster?view=azureml-api-2&tabs=python

Instances require 'ComputeInstance', see https://learn.microsoft.com/en-us/azure/machine-learning/how-to-create-compute-instance?view=azureml-api-2&tabs=python

upvoted 1 times

Matt2000 Most Recent 2 years ago

My mistake. The answer should be 'yes'

upvoted 1 times

haby 2 years, 2 months ago

2 should be No, this script is for training a ml model, that must be a computer instance, the first item in the pic shows. You can't say it's a cluster coz you assign it 'the_cluster_name'.

upvoted 1 times

SunilB 2 years, 11 months ago

Box 3 is correct - 0 The minimum number of nodes to use on the cluster. If not specified, defaults to 0. https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute(class)?view=azure-ml-py

upvoted 3 times

danishanis 2 years, 11 months ago

The correct answer is Yes, Yes and No The minimum number of nodes for the compute target created in the above code will default to 1 if you do not specify the min_nodes parameter, not 0

upvoted 1 times

danishanis 2 years, 11 months ago

The correct answer is Yes, Yes and No The minimum number of nodes for the compute target created in the above code will default to 1 if you do not specify the min_nodes parameter, not 0Additional explanation: For box 3, it is 1 and not 0; This means that the compute target will always have at least one node running, even if no jobs are currently scheduled to run on it. The max_nodes parameter, on the other hand, does default to 1 if not specified. This means that by default, the compute target will have a single node, and will not scale up to more nodes automatically.The default is zero https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute.amlcomputeprovisioningconfiguration?view=azure-ml-py#parameters

upvoted 1 times

danishanis 2 years, 11 months ago

Additional explanation: For box 3, it is 1 and not 0; This means that the compute target will always have at least one node running, even if no jobs are currently scheduled to run on it. The max_nodes parameter, on the other hand, does default to 1 if not specified. This means that by default, the compute target will have a single node, and will not scale up to more nodes automatically.The default is zero https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute.amlcomputeprovisioningconfiguration?view=azure-ml-py#parameters

upvoted 1 times

danishanis 2 years, 11 months ago

Additional explanation: For box 3, it is 1 and not 0; This means that the compute target will always have at least one node running, even if no jobs are currently scheduled to run on it.

The max_nodes parameter, on the other hand, does default to 1 if not specified. This means that by default, the compute target will have a single node, and will not scale up to more nodes automatically.

upvoted 1 times

GaryEl 2 years, 9 months ago

The default is zero https://learn.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute.amlcomputeprovisioningconfiguration?view=azure-ml-py#parameters

upvoted 6 times

Suman_512 2 years, 11 months ago

For 2 it will be rather "Compute Instance" than "Compute Cluster" so - NO

upvoted 1 times

Suman_512 2 years, 11 months ago

For 2 it will be rather "Compute Instance" than "Compute Cluster" so - NO"We guess" its a cluster by the varable name, I agree with YES and a not misleading question had to Exibit the Compute configuration ...

upvoted 1 times

InversaRadice 2 years, 2 months ago

"We guess" its a cluster by the varable name, I agree with YES and a not misleading question had to Exibit the Compute configuration ...

upvoted 1 times

racnaoamo 3 years, 9 months ago

similar question on 18-5-22

upvoted 4 times

hargur 4 years, 4 months ago

on 19Oct2021

upvoted 3 times

mthombenindhl84 4 years, 5 months ago

on exam 11/9/2021

upvoted 3 times

dushmantha 4 years, 5 months ago

On exam 2021/08/31

upvoted 2 times

Skandinov 4 years, 9 months ago

how's 1 yes?

upvoted 1 times

Skandinov 4 years, 9 months ago

how's 1 yes?"The compute is created within your workspace region as a resource that can be shared with other users in your workspace."
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-cluster?tabs=python#what-is-a-compute-cluster

upvoted 1 times

gamezone25 Highly Voted 4 years, 9 months ago

"The compute is created within your workspace region as a resource that can be shared with other users in your workspace."
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-cluster?tabs=python#what-is-a-compute-cluster

upvoted 7 times

HOTSPOT -

You have an Azure blob container that contains a set of TSV files. The Azure blob container is registered as a datastore for an Azure Machine Learning service workspace. Each TSV file uses the same data schema.

You plan to aggregate data for all of the TSV files together and then register the aggregated data as a dataset in an Azure Machine Learning workspace by using the Azure Machine Learning SDK for Python.

You run the following code.

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

Hot Area:

```
from azureml.core.workspace import Workspace
from azureml.core.datastore import Datastore
from azureml.core.dataset import Dataset
import pandas as pd
datastore_paths = (datastore, './data/*.tsv')
myDataset_1 = Dataset.File.from_files(path=datastore_paths)
myDataset_2 = Dataset.Tabular.from_delimited_files(path=datastore_paths, separator='\t')
```

Answer Area

	Yes	No
The myDataset_1 dataset can be converted into a pandas dataframe by using the following method: `using myDataset_1.to_pandas_dataframe()`	○	○
The myDataset_1.to_path() method returns an array of file paths for all of the TSV files in the dataset.	○	○
The myDataset_2 dataset can be converted into a pandas dataframe by using the following method: `myDataset_2.to_pandas_dataframe()`	○	○

Suggested Answer:

Answer Area

	Yes	No
The myDataset_1 dataset can be converted into a pandas dataframe by using the following method: `using myDataset_1.to_pandas_dataframe()`	○	●
The myDataset_1.to_path() method returns an array of file paths for all of the TSV files in the dataset.	●	○
The myDataset_2 dataset can be converted into a pandas dataframe by using the following method: `myDataset_2.to_pandas_dataframe()`	●	○

Answer Area

	Yes	No
The myDataset_1 dataset can be converted into a pandas dataframe by using the following method: `using myDataset_1.to_pandas_dataframe()`	○	●
The myDataset_1.to_path() method returns an array of file paths for all of the TSV files in the dataset.	●	○
The myDataset_2 dataset can be converted into a pandas dataframe by using the following method: `myDataset_2.to_pandas_dataframe()`	●	○

Box 1: No - FileDataset references single or multiple files in datastores or from public URLs. The TSV files need to be parsed.

Box 2: Yes - to_path() gets a list of file paths for each file stream defined by the dataset.

Box 3: Yes - TabularDataset.to_pandas_dataframe loads all records from the dataset into a pandas DataFrame. TabularDataset represents data in a tabular format created by parsing the provided file or list of files. Note: TSV is a file extension for a tab-delimited file used with spreadsheet software. TSV stands for Tab Separated Values. TSV files are used for raw data and can be imported into and exported from spreadsheet software. TSV files are essentially text files, and the raw data can be viewed by text editors, though they are often used when moving raw data between spreadsheets. Reference: https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.data.tabulardataset

SaulG Highly Voted 3 years, 8 months ago
Answer No, Yes, Yes is correct!
upvoted 18 times

Dilesha Most Recent 1 year, 12 months ago
On Exam 17 Feb 2023
upvoted 4 times

majma 2 years, 8 months ago
Should be 3 No
upvoted 1 times

kkkk_jjjj 2 years, 11 months ago
on exam 18/03/2022
upvoted 3 times

TheYazan 2 years, 11 months ago
On march 2022
upvoted 2 times

JoshuaXu 3 years, 3 months ago
on exam 6 Nov 2021
upvoted 1 times

AI247 3 years, 3 months ago
Was in exam 11/05/2021
upvoted 1 times

pkal 3 years, 4 months ago

on exam 9/24/2021

upvoted 1 times

ljljljlj 3 years, 7 months ago

On exam 2021/7/10

upvoted 4 times

ljljljlj 3 years, 7 months ago

On exam 2021/7/10how many questions came from this dump ?

upvoted 4 times

slash_nyk Highly Voted 3 years, 7 months ago

how many questions came from this dump ?

upvoted 5 times

Question #15 *Topic 3*

You create a batch inference pipeline by using the Azure ML SDK. You configure the pipeline parameters by executing the following code:
You need to obtain the output from the pipeline execution.
Where will you find the output?

```
from azureml.contrib.pipeline.steps import ParallelRunConfig
parallel_run_config = ParallelRunConfig(
     source_directory=scripts_folder,
     entry_script= "batch_pipeline.py",
     mini_batch_size= "5",
     error_threshold=10,
     output_action= "append_row",
     environment=batch_env,
     compute_target=compute_target,
     logging_level= "DEBUG",
     node_count=4)
```

A. the digit_identification.py script

B. the debug log

C. the Activity Log in the Azure portal for the Machine Learning workspace

D. the Inference Clusters tab in Machine Learning studio

E. a file named parallel_run_step.txt located in the output folder

Suggested Answer: *E*
output_action (str): How the output is to be organized. Currently supported values are 'append_row' and 'summary_only'. 'append_row' ɹ€" All values output by run() method invocations will be aggregated into one unique file named parallel_run_step.txt that is created in the output location. 'summary_only' Reference: https://docs.microsoft.com/en-us/python/api/azureml-contrib-pipeline-steps/azureml.contrib.pipeline.steps.parallelrunconfig

Community vote distribution
E (100%)

ljljljlj Highly Voted 4 years, 1 month ago
On exam 2021/7/10
upvoted 11 times

PI_Team Most Recent 2 years ago
By default, the ParallelRunStep class writes its output to a file named parallel_run_step.txt in the specified output directory. This file contains the concatenated output from all the mini-batches processed by the pipeline. Each line in the file represents the output from a single mini-batch.
upvoted 2 times

Edriv 2 years, 8 months ago
https://learn.microsoft.com/en-us/azure/machine-learning/migrate-to-v2-execution-parallel-run-step
upvoted 1 times

Edriv 2 years, 8 months ago
https://learn.microsoft.com/en-us/azure/machine-learning/migrate-to-v2-execution-parallel-run-stephttps://learn.microsoft.com/en-us/python/api/azureml-pipeline-steps/azureml.pipeline.steps.parallelrunconfig?view=azure-ml-py
upvoted 1 times

Edriv 2 years, 7 months ago
https://learn.microsoft.com/en-us/python/api/azureml-pipeline-steps/azureml.pipeline.steps.parallelrunconfig?view=azure-ml-py
upvoted 2 times

JTWang 3 years, 3 months ago
on exam 04/22/2022
upvoted 3 times

synapse 3 years, 5 months ago
Given answer is correct

upvoted 2 times

TheYazan 3 years, 6 months ago

Correct

upvoted 2 times

DRAG DROP -

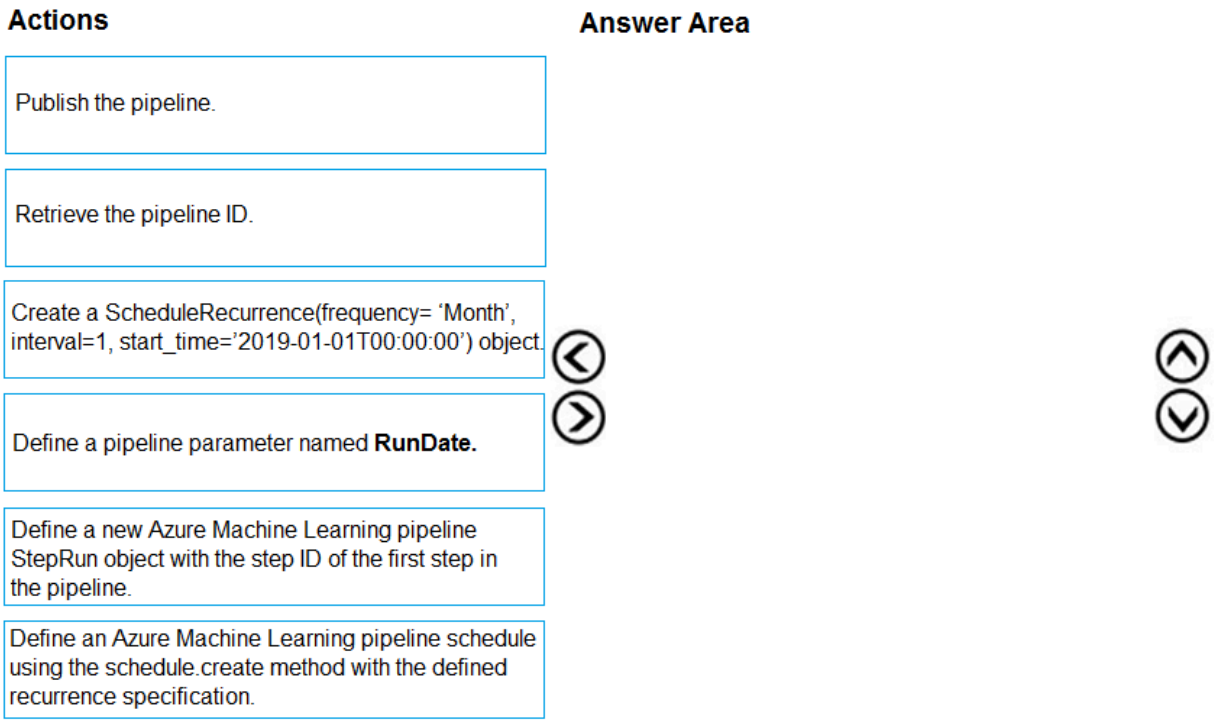
You create a multi-class image classification deep learning model.

The model must be retrained monthly with the new image data fetched from a public web portal. You create an Azure Machine Learning pipeline to fetch new data, standardize the size of images, and retrain the model.

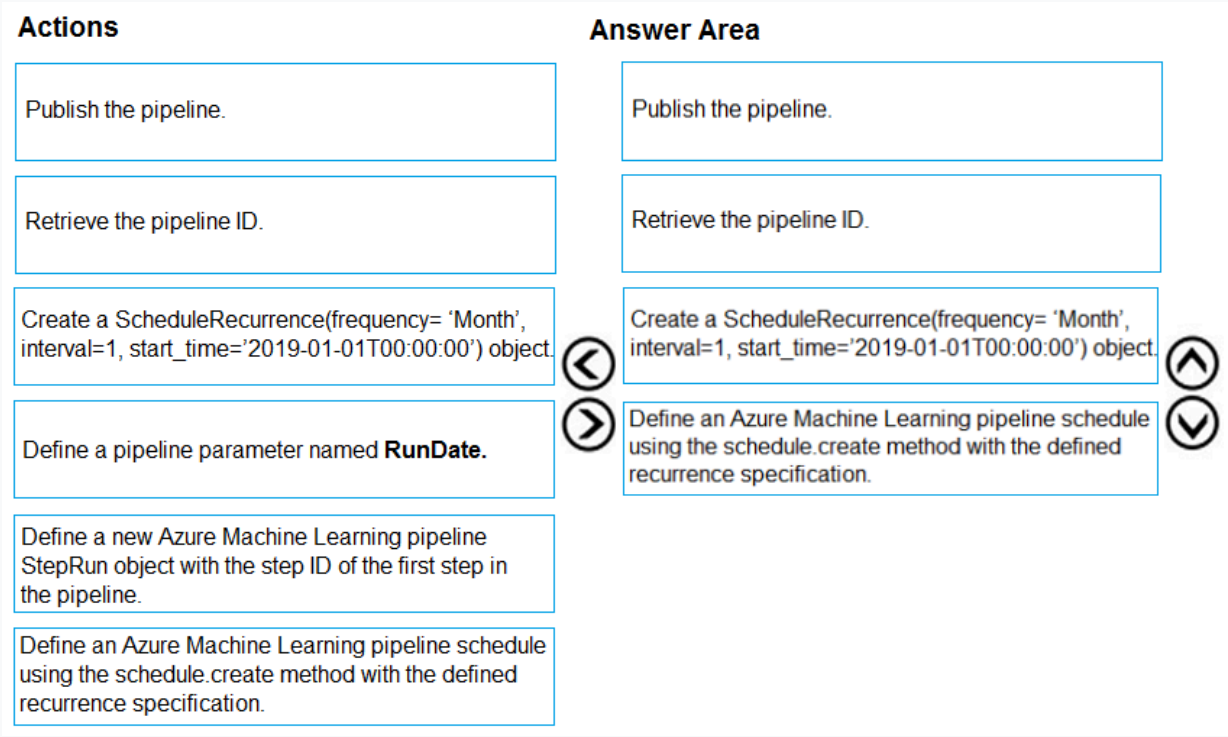
You need to use the Azure Machine Learning SDK to configure the schedule for the pipeline.

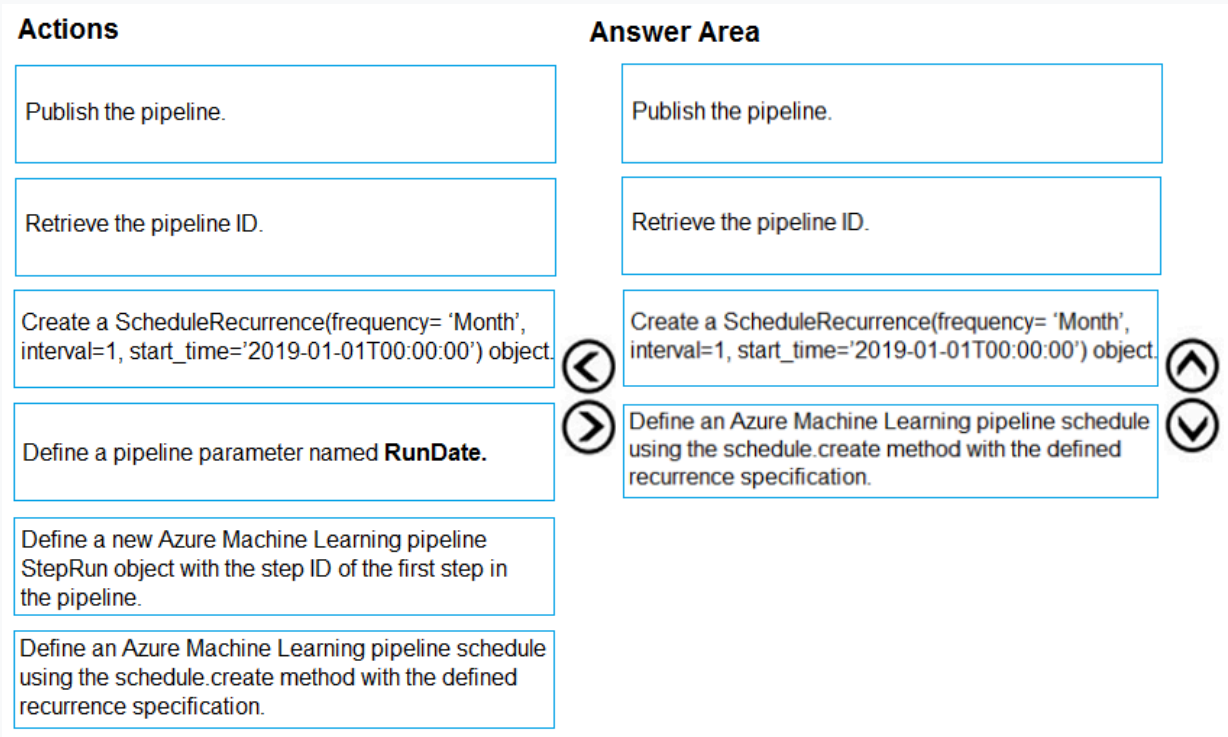
Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Select and Place:



Suggested Answer:





Step 1: Publish the pipeline. To schedule a pipeline, you'll need a reference to your workspace, the identifier of your published pipeline, and the name of the experiment in which you wish to create the schedule. Step 2: Retrieve the pipeline ID. Needed for the schedule. Step 3: Create a ScheduleRecurrence.. To run a pipeline on a recurring basis, you'll create a schedule. A Schedule associates a pipeline, an experiment, and a trigger. First create a schedule. Example: Create a Schedule that begins a run every 15 minutes: recurrence = ScheduleRecurrence(frequency="Minute", interval=15) Step 4: Define an Azure Machine Learning pipeline schedule.. Example, continued: recurring_schedule = Schedule.create(ws, name="MyRecurringSchedule", description="Based on time", pipeline_id=pipeline_id, experiment_name=experiment_name, recurrence=recurrence) Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-schedule-pipelines

BilJon Highly Voted 4 years, 10 months ago

Answer is corrrect

upvoted 17 times

evangelist Most Recent 1 year, 8 months ago

Answer is correct and also the sequence

upvoted 1 times

ning 3 years, 9 months ago

Correct, given that you have the model and pipeline ready, and you want to schedule a run

upvoted 1 times

racnaoamo 3 years, 9 months ago

on exam 18-5-22

upvoted 4 times

snsnsnsn 4 years, 5 months ago

on 2/9/21

upvoted 2 times

datamijn 4 years, 6 months ago

on 2/8/2021

upvoted 4 times

azurecert2021 4 years, 7 months ago

given answer is correct so that you can initiate the scheduled runs of a pipeline, the pipelinemust be created and published first. Then you have to create a ScheduleRecurrence (trigger) objectand, finally the Schedule itself, connecting the pipeline (via its id) and the trigger object must becreated and run. You can watch the scheduled pipelines in the Endpoints section of the ML Studio. https://docs.microsoft.com/en-us/azure/machine-learning/how-to-schedule-pipelines https://github.com/Azure/MachineLearningNotebooks/blob/master/how-to-use-azureml/machine-learning-pipelines/intro-to-pipelines/aml-pipelines-setup-schedule-for-a-published-pipeline.ipynb

upvoted 4 times