

70-774.exam.27q

Number: 70-774
Passing Score: 800
Time Limit: 120 min



Website: <https://vceplus.com>
VCE to PDF Converter: <https://vceplus.com/vce-to-pdf/>
Facebook: <https://www.facebook.com/VCE.For.All.VN/>
Twitter : https://twitter.com/VCE_Plus

<https://vceplus.com/>

70-774

Perform Cloud Data Science with Azure Machine Learning

Exam A

QUESTION 1

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 sections, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are working on an Azure Machine Learning experiment.

You have the dataset configured as shown in the following table.

Model	Mean absolute error (MAE)
Boosted decision tree	.2
Relative absolute error (RAE)	.43

You need to ensure that you can compare the performance of the models and add annotations to the results.

Solution: You consolidate the output of the Score Model modules by using the Add Rows module, and then use the Execute R Script module.

Does this meet the goal?

- A. Yes
- B. No

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

Explanation:

References: <https://msdn.microsoft.com/en-us/library/azure/dn905915.aspx>

QUESTION 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 sections, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are working on an Azure Machine Learning experiment.

You have the dataset configured as shown in the following table.

Model	Mean absolute error (MAE)
Boosted decision tree	.2
Relative absolute error (RAE)	.43

You need to ensure that you can compare the performance of the models and add annotations to the results.

Solution: You connect the Score Model modules from each trained model as inputs for the Evaluate Model module, and use the Execute R Script module.

Does this meet the goal?



<https://vceplus.com/>

- A. Yes
- B. No

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

Explanation:

References: <https://msdn.microsoft.com/en-us/library/azure/dn905915.aspx>

QUESTION 3

You have the following three training datasets for a restaurant:

- User features
- Item features
- Ratings of items by users

You must recommend restaurant to a particular user based only on the users features.

You need to use a Matchbox Recommender to make recommendations.

How many input parameters should you specify?

- A. 1
- B. 2
- C. 3
- D. 4

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

Explanation:

References: <https://msdn.microsoft.com/en-us/library/azure/dn905987.aspx>

QUESTION 4

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.

A travel agency named Margie's Travel sells airline tickets to customers in the United States.

Margie's Travel wants you to provide insights and predictions on flight delays. The agency is considering implementing a system that will communicate to its customers as the flight departure nears about possible delays due to weather conditions. The flight data contains the following attributes: ▪

DepartureDate: The departure date aggregated at a per hour granularity

- Carrier: The code assigned by the IATA and commonly used to identify a carrier
- OriginAirportID: An identification number assigned by the USDOT to identify a unique airport (the flight's origin)
- DestAirportID: An identification number assigned by the USDOT to identify a unique airport (the flight's destination)
- DepDel: The departure delay in minutes

- DepDel30: A Boolean value indicating whether the departure was delayed by 30 minutes or more (a value of 1 indicates that the departure was delayed by 30 minutes or more)

The weather data contains the following attributes: AirportID, ReadingDate (YYYY/MM/DD HH), SkyConditionVisibility, WeatherType, WindSpeed, StationPressure, PressureChange, and HourlyPrecip.

You plan to predict flight delays that are 30 minutes or more.

You need to build a training model that accurately fits the data. The solution must minimize over fitting and minimize data leakage.

Which attribute should you remove?

- A. OriginAirportID
- B. DepDel
- C. DepDel30
- D. Carrier
- E. DestAirportID

Correct Answer: C

Section: (none)

Explanation



Explanation/Reference:

Explanation:

QUESTION 5

Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series. Information and details provided in a question apply only to that question.

You need to remove rows that have an empty value in a specific column. The solution must use a native module.

Which module should you use?

- A. Execute Python Script
- B. Tune Model Hyperparameters
- C. Normalize Data
- D. Select Columns in Dataset
- E. Import Data
- F. Edit Metadata

- G. Clip Values
- H. Clean Missing Data

Correct Answer: H

Section: (none)

Explanation

Explanation/Reference:

Explanation:

References: <https://blogs.msdn.microsoft.com/azuredev/2017/05/27/data-cleansing-tools-in-azure-machine-learning/>

QUESTION 6

Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series. Information and details provided in a question apply only to that question.

You have a dataset that contains a column named Column1. Column1 is empty.

You need to omit Column1 from the dataset. The solution must use a native module.

Which module should you use?

- A. Execute Python Script
- B. Tune Model Hyperparameters
- C. Normalize Data
- D. Select Columns in Dataset
- E. Import Data
- F. Edit Metadata
- G. Clip Values
- H. Clean Missing Data

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

Explanation:

References: <https://msdn.microsoft.com/en-us/library/azure/dn905883.aspx>

QUESTION 7

Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series. Information and details provided in a question apply only to that question.

You need to use only one percent of an Apache Hive data table by conducting random sampling by groups.

Which module should you use?

- A. Execute Python Script
- B. Tune Model Hyperparameters
- C. Normalize Data
- D. Select Columns in Dataset
- E. Import Data
- F. Edit Metadata
- G. Clip Values
- H. Clean Missing Data

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

Explanation:

References: <https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/sample-data-hive>

QUESTION 8

From the Cortana Intelligence Gallery, you deploy a solution.

You need to modify the solution.

What should you use?

- A. Azure Stream Analytics
- B. Microsoft Power BI Desktop



- C. Azure Machine Learning Studio
- D. R Tools for Visual Studio

Correct Answer: C

Section: (none)

Explanation

Explanation/Reference:

Explanation:

References: <https://docs.microsoft.com/en-us/azure/machine-learning/studio/gallery-experiments>

QUESTION 9

You plan to use the Data Science Virtual Machine for development, but you are unfamiliar with R scripts.

You need to generate R code for an experiment.



<https://vceplus.com/>

Which IDE should you use?

- A. XgBoost
- B. Rattle
- C. Vowpal Wabbit
- D. R Tools for Visual Studio

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

Explanation:

References: <https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/provision-vm>

QUESTION 10

You are performing exploratory analysis of files that are encoded in a complex proprietary format. The format requires disk intensive access to several dependent files in HDFS.

You need to build an Azure Machine Learning model by using a canopy clustering algorithm. You must ensure that changes to proprietary file formats can be maintained by using the least amount of effort.

Which Machine Learning library should you use?

- A. MicrosoftML
- B. scikit-learn
- C. SparkRD. Mahout

Correct Answer: D

Section: (none)

Explanation

**Explanation/Reference:**

Explanation:

QUESTION 11

DRAG DROP

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.

You plan to create a predictive analytics solution for credit risk assessment and fraud prediction in Azure Machine Learning. The Machine Learning workspace for the solution will be shared with other users in your organization. You will add assets to projects and conduct experiments in the workspace.

The experiments will be used for training models that will be published to provide scoring from web services.

The experiment for fraud prediction will use Machine Learning modules and APIs to train the models and will predict probabilities in an Apache Hadoop ecosystem.

You finish training the model and are ready to publish a predictive web service that will provide the users with the ability to specify the data source and the save location of the results. The model includes a Split Data module.

Which two actions should you perform to convert the Machine Learning experiment to a predictive web service? To answer, drag the appropriate actions to the correct targets. Each action 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:

Actions	Answer Area
Click Set Up Web Service for the training experiment.	First action: <input type="text" value="Action"/>
Configure a web service endpoint for input and output, and then specify the parameters.	Second action: <input type="text" value="Action"/>
Remove the Split Data module.	
Replace the Machine Learning algorithm and the train model by using a saved training model.	
Save the trained model.	

Correct Answer:

Actions

Remove the Split Data module.

Replace the Machine Learning algorithm and the train model by using a saved training model.

Save the trained model.

Answer Area

First action:

Click Set Up Web Service for the training experiment.

Second action:

Configure a web service endpoint for input and output, and then specify the parameters.

Section: (none)

Explanation

Explanation/Reference:

References: <https://docs.microsoft.com/en-us/azure/machine-learning/studio/convert-training-experiment-to-scoring-experiment>

QUESTION 12

You have an Azure Machine Learning experiment.

You discover that a model causes many errors in a production dataset. The model causes only few errors in the training data.

What is the cause of the errors?

A. overfitting

- B. generalization
- C. underfitting
- D. a simple predictor

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

Explanation:

QUESTION 13

You need to integrate code and formatted text into an Azure Machine Learning experiment that enables interactive execution.

What should you use?

- A. a Jupyter notebook
- B. Azure Stream Analytics
- C. an Execute Python Script module
- D. an Execute R Script module



Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

Explanation:

QUESTION 14

You have a dataset that is missing values in a column named Column3. Column3 is correlated to two columns named Column4 and Column5.

You need to improve the accuracy of the dataset, while minimizing data loss.

What should you do?

- A. Replace the missing values in Column3 by using probabilistic Principal Component Analysis (PCA).
- B. Remove all of the rows that have the missing values in Column4 and Column5.
- C. Replace the missing values in Column3 with a mean value.

D. Remove the rows that have the missing values in Column3.

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

Explanation:

QUESTION 15

You are building an Azure Machine Learning experiment.

You need to transform 47 numeric columns into a set of 10 linearly uncorrelated features.

Which module should you add to the experiment?

- A. Principal Component Analysis
- B. K-Means Clustering
- C. Normalize Data
- D. Group Data into Bins



Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

Explanation:

QUESTION 16

You are building an Azure Machine Learning experiment.

You need to transform a string column into a label column for a Multiclass Decision Jungle module.

Which module should you use?

- A. Select Columns Transform
- B. Group Categorical Values
- C. Convert to Indicator Values
- D. Edit Metadata

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

Explanation:

QUESTION 17

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 ML experiment that contains an intermediate dataset.

You need to explore data from the intermediate dataset by using Jupyter.

Solution: You add a Convert to CSV module to the Azure ML experiment and then open the module output in a new notebook.

Does this meet the goal?

A. Yes

B. No



Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

Explanation:

Access intermediate datasets from Machine Learning experiments

The following steps show an example that creates an experiment, runs it and accesses the intermediate dataset.

1. Create a new experiment.
2. Insert an Adult Census Income Binary Classification dataset module.
3. Insert a Split module, and connect its input to the dataset module output.
4. Insert a Convert to CSV module and connect its input to one of the Split module outputs.
5. Save the experiment, run it, and wait for it to finish running.
6. Click the output node on the Convert to CSV module.

7. When the context menu appears, select **Generate Data Access Code**
8. Select the code snippet and copy it to your clipboard from the window that appears..
9. Paste the code in your notebook.

Note: After an experiment is run in the Machine Learning Studio, it is possible to access the intermediate datasets from the output nodes of modules. Intermediate datasets are data that has been created and used for intermediate steps when a model tool has been run.

Intermediate datasets can be accessed as long as the data format is compatible with the Python client library.

The following formats are supported (constants for these are in the azureml.DataTypes class): ▪

PlainText

- GenericCSV
- GenericTSV
- GenericCSVNoHeader
- GenericTSVNoHeader

References: <https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/python-data-access>

QUESTION 18

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 ML experiment that contains an intermediate dataset.

You need to explore data from the intermediate dataset by using Jupyter.

Solution: You add a web service input to retrieve the data for the data source, and then add the Execute R Script module.

Does this meet the goal?

- A. Yes
- B. No

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

Explanation:

Use a Convert to CSV module.

References: <https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/python-data-access>

QUESTION 19

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 ML experiment that contains an intermediate dataset.

You need to explore data from the intermediate dataset by using Jupyter.

Solution: You add a Convert to ARFF module, and then add the Execute R Script module.



<https://vceplus.com/>

Does this meet the goal?

- A. Yes
- B. No

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

Explanation:

Use a Convert to CSV module.

References: <https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/python-data-access>

QUESTION 20

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.

Start of repeated scenario

You plan to use Azure platform tools to detect and analyze food items in smart refrigerators. To provide families with an integrated experience for grocery shopping and cooking, the refrigerators will connect to other smart appliances, such as stoves and microwave ovens, on a LAN.

You plan to build an object recognition model by using the Microsoft Cognitive Toolkit. The object recognition model will receive input from the connected devices and send results to applications.

The training data will be derived from more than 10 TB of images. You will convert the raw images to the sparse format.

End of repeated scenario.

You need to preprocess the training data by using a Principal Component Analysis (PCA) algorithm in the least amount of time possible.

Which implementation method should you use?

- A. Azure HDInsight using HiveML
- B. Azure Machine Learning Studio and a custom Execute Python Script module
- C. Azure HDInsight using Microsoft R Server
- D. Azure Machine Learning Studio with a custom Execute R Script module

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

References: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/principal-component-analysis>

QUESTION 21

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.

Start of repeated scenario

You plan to use Azure platform tools to detect and analyze food items in smart refrigerators. To provide families with an integrated experience for grocery shopping and cooking, the refrigerators will connect to other smart appliances, such as stoves and microwave ovens, on a LAN.

You plan to build an object recognition model by using the Microsoft Cognitive Toolkit. The object recognition model will receive input from the connected devices and send results to applications.

The training data will be derived from more than 10 TB of images. You will convert the raw images to the sparse format.

End of repeated scenario.

You need to deploy a multiple-service solution that was developed already and published by other users in the Microsoft development community.

What should you use?

- A. the edX Data Science Learning Dashboard
- B. the Data Science Virtual Machine
- C. an Azure Machine Learning experiment
- D. Cortana Intelligence Gallery

Correct Answer: D

Section: (none)

Explanation



Explanation/Reference:

Explanation:

Azure AI Gallery enables our growing community of developers and data scientists to share their analytics solutions.

References: <https://gallery.azure.ai/>

QUESTION 22

Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series. Information and details provided in a question apply only to that question.

You process some data by using Azure Machine Learning Studio. You have an intermediate dataset. The dataset has a column that contains date values stored in a format of MM/DD/YYYY.

You need to split the column into three separate columns by year, month, and day.

Which module should you use?

- A. Edit Metadata
- B. Normalize Data
- C. Clean Missing Data
- D. Import Data
- E. Execute Python Script
- F. Clip Values

- G. Tune Model Hyperparameters
- H. Select Columns in Dataset

Correct Answer: E

Section: (none)

Explanation

Explanation/Reference:

Explanation:

You can use Python code to create a new column.

Incorrect Answers:

F: The Clip Values module in Azure Machine Learning Studio, is used to identify and optionally replace data values that are above or below a specified threshold.

References: <https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-transform-data>

QUESTION 23

Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series. Information and details provided in a question apply only to that question.

You need to transform the columns in a dataset. The resulting data must be mean centered and have a variance of 1. The solution must use a native module.

Which module should you use?

- A. Execute Python Script
- B. Import Data
- C. Edit Metadata
- D. Select Columns in Dataset
- E. Clean Missing Data
- F. Tune Model Hyperparameters

- G. Clip Values
- H. Normalize Data

Correct Answer: H

Section: (none)

Explanation

Explanation/Reference:

QUESTION 24

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.

Start of repeated scenario

You plan to create a predictive analytics solution for credit risk assessment and fraud prediction in Azure Machine Learning. The Machine Learning workspace for the solution will be shared with other users in your organization. You will add assets to projects and conduct experiments in the workspace.

The experiments will be used for training models that will be published to provide scoring from web services.

The experiment for fraud prediction will use Machine Learning modules and APIs to train the models and will predict probabilities in an Apache Hadoop ecosystem.

End of repeated scenario.

You plan to share the Machine Learning workspace with the other users.

You are evaluating whether to assign the User role or the Owner role to several of the users.

Which three actions can be performed by the users who are assigned the User role? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. Create, open, modify, and delete datasets.
- B. Create, open, modify, and delete experiments.
- C. Invite users to the workspace.
- D. Delete users from the workspace.
- E. Create, open, modify, and delete web services.
- F. Access notebooks.

Correct Answer: ABE

Section: (none)

Explanation

Explanation/Reference:

Explanation:

Once a Machine Learning workspace is created, you can invite users to your workspace and share access to your workspace and all of its experiments. We support two roles of users:

- User - A workspace user can create, open, modify and delete datasets, experiments and web services in the workspace.
- Owner - An owner can invite, remove, and list users with access to the workspace, in addition to what a user can do. He/she also have access to Notebooks.

References: <https://github.com/anthonychu/azure-content/blob/master/articles/machine-learning/machine-learning-create-workspace.md>

QUESTION 25

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.

Start of repeated scenario

You plan to create a predictive analytics solution for credit risk assessment and fraud prediction in Azure Machine Learning. The Machine Learning workspace for the solution will be shared with other users in your organization. You will add assets to projects and conduct experiments in the workspace.

The experiments will be used for training models that will be published to provide scoring from web services.

The experiment for fraud prediction will use Machine Learning modules and APIs to train the models and will predict probabilities in an Apache Hadoop ecosystem.

End of repeated scenario.

The users will use different data sources that follow a standard format. The users will receive results in a standard format by using the fraud prediction web service. The results will be saved to a location specified by the users.

You need to provide the users with the ability to get results for different risk tolerances without affecting the calculation of the model.

Which three modules should be configured to use the Web Service Parameters? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Evaluate Model
- B. Import Data
- C. Select Columns in Dataset

- D. Export Data
- E. Time Series Anomaly Detection

Correct Answer: ABD

Section: (none)

Explanation

Explanation/Reference:

Explanation:

A: From scenario: The experiment for fraud prediction will use Machine Learning modules and APIs to train the models and will predict probabilities in an Apache Hadoop ecosystem.

BD: An Azure Machine Learning web service is created by publishing an experiment that contains modules with configurable parameters. In some cases, you may want to change the module behavior while the web service is running. Web Service Parameters allow you to do this task.

A common example is setting up the Import Data module so that the user of the published web service can specify a different data source when the web service is accessed. Or configuring the Export Data module so that a different destination can be specified

References: <https://docs.microsoft.com/en-us/azure/machine-learning/studio/web-service-parameters>

QUESTION 26

You need to identify which columns are more predictive by using a statistical method. Which module should you use?

- A. Filter Based Feature Selection
- B. Principal Component Analysis
- C. Group Data into Bins
- D. Tune Model Hyperparameters

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

Explanation:

In machine learning and statistics, feature selection is the process of selecting a subset of relevant, useful features to use in building an analytical model. Feature selection helps narrow the field of data to the most valuable inputs. Narrowing the field of data helps reduce noise and improve training performance.

Feature selection is an important tool in machine learning. Machine Learning Studio provides multiple methods for performing feature selection. Choose a feature selection method based on the type of data that you have, and the requirements of the statistical technique that's applied.

References: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/feature-selection-modules>

QUESTION 27

You are building an Azure Machine Learning workflow by using Azure Machine Learning Studio.

You create an Azure notebook that supports the Microsoft Cognitive Toolkit.

You need to ensure that the stochastic gradient descent (SGD) configuration maximizes the samples per second and supports parallel modeling that is managed by a parameter server.

Which SGD algorithm should you use?

- A. DataParallelASGD
- B. DataParallelSGD
- C. BlockMomentumSGD
- D. ModelAveragingSGD

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

Explanation:

Data-Parallel Training with Parameter Server

Parameter server is a widely used framework in distributed machine learning. The most important benefit it brings is the asynchronous parallel training with many workers.

You can use Data-Parallel ASGD.

Incorrect ANswers:

D: Model-Averaging SGD generally converges more slowly and to a worse optimum, compared to 1-bit SGD and Block-Momentum SGD, so it is no longer recommended.

References: <https://docs.microsoft.com/en-us/cognitive-toolkit/multiple-gpus-and-machines>





<https://vceplus.com/>

