

Guarantee All Exams 100% Pass One Time!

# 2019 Microsoft 70-774: Perform Cloud Data Science with Azure Machine Learning Exam Questions and Answers RELEASED in Geekcert.com Online IT Study Website Today!

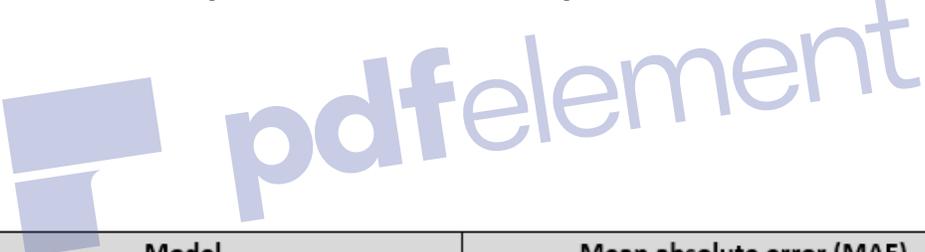
<https://www.geekcert.com/70-774.html>

## 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 connect the Score Model modules from each trained model as inputs for the Evaluate Model module, and then save the results as a dataset.

Does this meet the goal?

- A. Yes
- B. No

**Answer:** A

**Explanation:**

<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

## Guarantee All Exams 100% Pass One Time!

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?

- A. Yes
- B. No

**Answer: B**

**Explanation:**

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

### QUESTION 3

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 norigin)
- `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 need to use historical data about on-time flight performance and the weather data to predict whether the departure of a scheduled flight will be delayed by more than 30 minutes.

Which method should you use?

**Guarantee All Exams 100% Pass One Time!**

- A. clustering
- B. linear regression
- C. classification
- D. anomaly detection

**Answer: C**

**Explanation:**

<https://gallery.cortanaintelligence.com/Experiment/Binary-Classification-Flight-delay-prediction-3>

**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 norigin)
- **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 have an untrained Azure Machine Learning model that you plan to train to predict flight delays.

You need to assess the variability of the dataset and the reliability of the predictions from the model.

Which module should you use?

- A. Cross-Validate Model
- B. Evaluate Model
- C. Tune Model Hyperparameters
- D. Train Model
- E. Score Model

**Answer: A**

**Explanation:**

<https://msdn.microsoft.com/en-us/library/azure/dn905852.aspx>

**Guarantee All Exams 100% Pass One Time!****QUESTION 5**

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 norigin)
- `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`

**Answer: C**

**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 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`

**Guarantee All Exams 100% Pass One Time!**

- D. Select Columns in Dataset
- E. Import Data
- F. Edit Metadata
- G. Clip Values
- H. Clean Missing Data

**Answer:** H

**Explanation:**

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

**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 have a non-tabular file that is saved in Azure Blob storage.

You need to download the file locally, access the data in the file, and then format the data as a dataset.

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

**Answer:** E

**Explanation:**

<https://msdn.microsoft.com/en-us/library/azure/mt674698.aspx>

**QUESTION 8**

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

**Guarantee All Exams 100% Pass One Time!**

- G. Clip Values
- H. Clean Missing Data

**Answer: D**

**Explanation:**

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

**QUESTION 9**

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

**Answer: A**

**Explanation:**

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

**QUESTION 10**

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

**Answer: C**

**Explanation:**

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

**QUESTION 11**

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?

**Guarantee All Exams 100% Pass One Time!**

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

**Answer: B**



**Guarantee All Exams 100% Pass One Time!**

**QUESTION 12**

You are analyzing taxi trips in New York City. You leverage the Azure Data Factory to create data pipelines and to orchestrate data movement.

You plan to develop a predictive model for 170 million rows (37 GB) of raw data in Apache Hive by using Microsoft R Server to identify which factors contribute to the passenger tipping behavior. All of the platforms that are used for the analysis are the same. Each worker node has eight processor cores and 26 GB of memory.

Which type of Azure HDInsight cluster should you use to produce results as quickly as possible?

- A. Hadoop
- B. HBase
- C. Interactive Hive
- D. Spark

**Answer: C**

**Explanation:**

<https://azure.microsoft.com/en-gb/blog/general-availability-of-hdinsight-interactive-query-blazing-fast-data-warehouse-style-queries-on-hyper-scale-data-2/>

**QUESTION 13**

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.

Which IDE should you use?

- A. XgBoost
- B. Rattle

pdfelement

**Guarantee All Exams 100% Pass One Time!**

- C. Vowpal Wabbit
- D. R Tools for Visual Studio

**Answer: B**

**Explanation:**

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

**QUESTION 14**

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. SparkR
- D. Mahout

**Answer: D**

**QUESTION 15**

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 plan to configure the resources for part of a workflow that will be used to preprocess data from files stored in Azure Blob storage. You plan to use Python to preprocess and store the data in Hadoop.

You need to get the data into Hadoop as quickly as possible.

Which three actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Create an Azure virtual machine (VM), and then configure MapReduce on the VM.
- B. Create an Azure HDInsight Hadoop cluster.
- C. Create an Azure virtual machine (VM), and then install an IPython Notebook server.
- D. Process the files by using Python to store the data to a Hadoop instance.
- E. Create the Machine learning experiment, and then add an Execute Python Script module.

**Answer: BDE**

**Guarantee All Exams 100% Pass One Time!****QUESTION 16**

You are building an Azure Machine Learning experiment.

You need to transform a string column that has 47 distinct values into a binary indicator column.

The solution must use the One-vs-All Multiclass model.

Which module should you use?

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

**Answer: B**

**QUESTION 17**

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

**Answer: A**

**QUESTION 18**

You have an Azure Machine Learning environment.

You are evaluating whether to use R code or Python.

Which three actions can you perform by using both R code and Python in the Machine Learning environment? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. Preprocess, cleanse, and group data.
- B. Score a training model.
- C. Create visualizations.
- D. Create an untrained model that can be used with the Train Model module.
- E. Implement feature ranking.

**Answer: ABC**

**Guarantee All Exams 100% Pass One Time!**

**QUESTION 19**

You plan to use Azure Machine Learning to develop a predictive model. You plan to include an Execute Python Script module.

What capability does the module provide?

- A. Outputting a file to a network location.
- B. Performing interactive debugging of a Python script.
- C. Saving the results of a Python script run in a Machine Learning environment to a local file.
- D. Visualizing univariate and multivariate summaries by using Python code.

**Answer: D**

**Explanation:**

<https://docs.microsoft.com/en-us/azure/machine-learning/studio/execute-python-scripts>

**QUESTION 20**

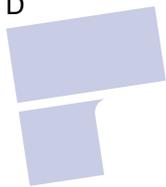
You are building an Azure Machine Learning experiment.

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

**Answer: D**



pdfelement

**Guarantee All Exams 100% Pass One Time!****QUESTION 21**

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 designing an Azure Machine Learning workflow.

You have a dataset that contains two million large digital photographs.

You plan to detect the presence of trees in the photographs.

You need to ensure that your model supports the following:

Hidden layers that support a directed graph structure

User-defined core components on the GPU

Solution: You create an endpoint to the Computer vision API.

Does this meet the goal?

- A. Yes
- B. No

**Answer: B**

**QUESTION 22**

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 designing an Azure Machine Learning workflow.

You have a dataset that contains two million large digital photographs.

You plan to detect the presence of trees in the photographs.

You need to ensure that your model supports the following:

Hidden layers that support a directed graph structure

User-defined core components on the GPU

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

Does this meet the goal?

- A. Yes
- B. No

**Answer: B**

**QUESTION 23**

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 designing an Azure Machine Learning workflow.

You have a dataset that contains two million large digital photographs.

You plan to detect the presence of trees in the photographs.

You need to ensure that your model supports the following:

Hidden layers that support a directed graph structure

**Guarantee All Exams 100% Pass One Time!**

User-defined core components on the GPU

Solution: You create a Machine Learning experiment that implements the Multiclass Neural Network module.

Does this meet the goal?

- A. Yes
- B. No

**Answer: A**

**QUESTION 24**

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 designing an Azure Machine Learning workflow.

You have a dataset that contains two million large digital photographs.

You plan to detect the presence of trees in the photographs.

You need to ensure that your model supports the following:

Hidden layers that support a directed graph structure

User-defined core components on the GPU

Solution: You create a Machine Learning experiment that implements the Multiclass Decision Jungle module.

Does this meet the goal?

- A. Yes
- B. No

**Answer: B**

**QUESTION 25**

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

**Guarantee All Exams 100% Pass One Time!**

B. No

**Answer: B**

**Explanation:**

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

#### **QUESTION 26**

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 need to alter the list of columns that will be used for predicting fraud for an input web service endpoint. The columns from the original data source must be retained while running the Machine Learning experiment.

Which module should you add after the web service input module and before the prediction module?

- A. Edit Metadata
- B. Import Data
- C. SMOTE
- D. Select Columns in Dataset

**Answer: D**

pdfelement

**Guarantee All Exams 100% Pass One Time!**

**QUESTION 27**

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

**Answer: A**

**QUESTION 28**

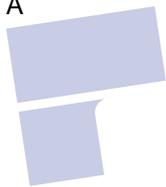
You are building an Azure Machine Learning solution for an online retailer.

When a customer selects a product, you need to recommend products that the customer might like to purchase at the same time. The recommendation should be based on what other customers purchased when they purchased the same product.

Which model should you use?

- A. Collaborative filtering
- B. Boosted Decision Tree Regression model
- C. Two-Class boosted decision tree
- D. K-Means Clustering

**Answer: A**



pdfelement

**Guarantee All Exams 100% Pass One Time!****QUESTION 29**

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

**Answer: A**

**QUESTION 30**

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.

**Answer: A**

**QUESTION 31**

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 save the output of the Score Model modules as a combined set, and then use the Project Columns module to select the MAE.

Does this meet the goal?

- A. Yes
- B. No

**Answer: B**

**Explanation:**

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

**Guarantee All Exams 100% Pass One Time!**

**QUESTION 32**

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

**Answer: C**

**Explanation:**

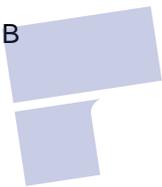
<https://docs.microsoft.com/en-us/azure/machine-learning/studio/gallery-experiments>

**QUESTION 33**

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. ModelAveragingSGD
- D. BlockMomentumSGD

**Answer: B**



pdfelement

**Guarantee All Exams 100% Pass One Time!**

**QUESTION 34**

You have data about the following:

Users

Movies

User ratings of the movies

You need to predict whether a user will like a particular movie.

Which Matchbox recommender should you use?

- A. Item Recommendation
- B. Related Items
- C. Rating Prediction
- D. Related Users

**Answer: C**

**Explanation:**

<https://msdn.microsoft.com/en-us/library/azure/dn905970.aspx#RatingPredictionOptions>

**QUESTION 35**

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

- 
- 
-

**Guarantee All Exams 100% Pass One Time!**

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

**Answer: B**

**Explanation:**

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

#### **QUESTION 36**

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. SparkR
- D. Mahout

**Answer: D**

#### **QUESTION 37**

Drag and Drop Questions

You have an Execute R Script module that has one input from either a Partition and Sample module or a Web service input module.

You need to preprocess tweets by using R. The solution must meet the following requirements:

- Remove digits.
- Remove punctuation.
- Convert to lowercase.

How should you complete the R code? To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar panes or scroll to view content. NOTE: Each correct selection is worth one point.

Guarantee All Exams 100% Pass One Time!

## Values

```
dataset[[1]]
```

```
dataset[[2]]
```

```
gsub("[^a-z]", " ", tweet_text, ignore.case = FALSE)
```

```
maml.mapInputPort(1)
```

```
sapply(tweet_text, tolower)
```

••••

## Answer area

```
dataset <-
```

Value

```
tweet_text <-
```

Value

```
tweet_text <-
```

Value

```
tweet_text <-
```

Value

```
data.set <- as.data.frame(tweet_text, stringsAsFactors=FALSE)
```

```
maml.mapOutputPort("data.set")
```

**Answer:**

Guarantee All Exams 100% Pass One Time!

## Values

```
dataset[[2]]
```



## Answer area

```
dataset <- maml.mapInputPort(1)

tweet_text <- dataset[[1]]

tweet_text <- sapply(tweet_text, tolower)

tweet_text <- gsub("[^a-z]", " ", tweet_text, ignore.case = FALSE)

data.set <- as.data.frame(tweet_text, stringsAsFactors=FALSE)

maml.mapOutputPort("data.set")
```

### QUESTION 38

#### Drag and Drop Questions

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

## Guarantee All Exams 100% Pass One Time!

- **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 need to remove the bias and to identify the columns in the input dataset that have the greatest predictive power.

Which module should you use for each requirement? To answer, drag the appropriate modules to the correct requirements. 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.

### Modules

Cross-validate Model
Evaluate Model
Filter and Sample
Filter Based Feature Selection Module
Parameter Sweep
Tune Model Hyperparameters

### Answer Area

Remove bias:

Identify the columns that have the greatest predictive power:

**Answer:**

### Modules

Evaluate Model
Filter and Sample
Filter Based Feature Selection Module
Parameter Sweep

### Answer Area

Remove bias:

Identify the columns that have the greatest predictive power:

**Guarantee All Exams 100% Pass One Time!**

**Explanation:**

<https://gallery.cortanaintelligence.com/Experiment/Binary-Classification-Flight-delay-prediction-3>  
<https://msdn.microsoft.com/library/azure/038d91b6-c2f2-42a1-9215-1f2c20ed1b40>



**QUESTION 41**

**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

**Answer:** A

**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

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

^p

#### QUESTION 42

**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

**Answer:** B

**Explanation:**

Use a Convert to CSV module.

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

#### QUESTION 43

**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.

**Guarantee All Exams 100% Pass One Time!**

**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: In Azure ML Studio, you use the Save as dataset option, and then open the output in a new notebook?

Does this meet the goal?

- A. Yes
- B. No

**Answer: B**

**Explanation:**

Use a Convert to CSV module.

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

**QUESTION 44**

**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.

Does this meet the goal?

- A. Yes
- B. No

**Answer: B**

**Explanation:**

Use a Convert to CSV module.

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

**QUESTION 45**

**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

**Guarantee All Exams 100% Pass One Time!**

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

**Answer:** D

**Explanation:**

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

**QUESTION 46**

**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 ensure that a web service endpoint can receive image data and use an object recognition model to return the expected object and the confidence level of the model.

The solution must minimize the effort required to generate the client code to access the web service.

Which resource should you use?

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

**Guarantee All Exams 100% Pass One Time!**

**Answer: B**

**Explanation:**

There are two ways to add an endpoint to a Web service.

Programmatically

Through the Azure Machine Learning Web Services portal

Adding an endpoint using the Azure Machine Learning Web Services portal

1. In Machine Learning Studio, on the left navigation column, click Web Services.
2. At the bottom of the Web service dashboard, click Manage endpoints. The Azure Machine Learning Web Services portal opens to the endpoints page for the Web service.
3. Click New.
4. Type a name and description for the new endpoint.

Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio/create-endpoint>

#### **QUESTION 47**

**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

**Answer: D**

**Explanation:**

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

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

#### **QUESTION 48**

**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**

**Guarantee All Exams 100% Pass One Time!**

**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

**Answer: E**

**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.

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

**QUESTION 49**

**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

**Answer: H**

**QUESTION 50**

**Guarantee All Exams 100% Pass One Time!**

**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 change a column name to a friendly name. 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

**Answer: D**

**Explanation:**

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/select-columns-in-dataset>

**QUESTION 51**

**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. Some of the values in Column1 are empty.

You need to replace the empty values by using probabilistic Principal Component Analysis (PCA). 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

**Answer: E**

**Explanation:**

You can use the Clean Missing Data module in Azure Machine Learning Studio, to remove,

**Guarantee All Exams 100% Pass One Time!**

replace, or infer missing values.

This module supports multiple type 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

You can choose to 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.

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



**Guarantee All Exams 100% Pass One Time!**

## QUESTION 52

**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.

**Guarantee All Exams 100% Pass One Time!**

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

**Answer:** ABE

**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.

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

**QUESTION 53**

**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

**Guarantee All Exams 100% Pass One Time!**

**Answer:** ABD

**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

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

#### **QUESTION 54**

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

**Answer:** A

**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.

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

#### **QUESTION 55**

You are building a classification experiment in Azure Machine Learning.

You need to ensure that you can use the Evaluate Model module in the experiment.

Which two actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Connect the input of the Score Model modules to the output of the Evaluate Model module.
- B. Connect the input of the Score Model modules to the output of the Train Model modules and the output Split Data modules.
- C. Connect the output of the Score Model modules to the input of the Evaluate Model module.
- D. Connect the output of the Score Model modules to the input of the Train Model modules and the

## Guarantee All Exams 100% Pass One Time!

input of the Split Data modules.

**Answer:** BC

**Explanation:**

Use the training data.

To evaluate a model, you must connect a dataset that contains a set of input columns and scores. If no other data is available, you can use your original dataset.

1. Connect the Scored dataset output of the Score Model to the input of Evaluate Model.
2. Click Evaluate Model module, and select Run selected to generate the evaluation scores.

Use testing data

A common scenario in machine learning is to separate your original data set into training and testing datasets, using the Split module, or the Partition and Sample module.

1. Connect the Scored dataset output of the Score Model to the input of Evaluate Model.
2. Connect the output of the Split Data module that contains the testing data to the right-hand input of Evaluate Model.

3. Click Evaluate Model module, and select Run selected to generate the evaluation scores.

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model>

### QUESTION 56

You are working on an Azure Machine Learning experiment that uses four different logistic regression algorithms.

You are evaluating the algorithms based on the data in the following table.

Metric	Model 1	Model 2	Model 3	Model 4
Mean absolute error (MAE)	.2	.63	.41	.46
Relative absolute error (RAE)	.40	.30	.60	.65
Root mean squared error (RMSE)	.25	.37	.15	.46

Which model produces predictions that are the closest to the actual outcomes?

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

**Answer:** C

**Explanation:**

Mean absolute error (MAE) measures how close the predictions are to the actual outcomes; thus, a lower score is better.

Relative absolute error (RAE) is the relative absolute difference between expected and actual values; relative because the mean difference is divided by the arithmetic mean.

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.

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate->

**Guarantee All Exams 100% Pass One Time!**

model

**QUESTION 57**

The manager of a call center reports that staffing the center is difficult because the number of calls is unpredictable.

You have historical data that contains information about the calls.

You need to build an Azure Machine Learning experiment to predict the number of total calls each hour.

Which model should you use?

- A. Multiclass Logistic Regression
- B. Boosted Decision Tree Regression
- C. Decision Forest Regression
- D. Poisson Regression

**Answer: D**

**Explanation:**

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.

Incorrect Answers:

A: Logistic regression is a well-known method in statistics that is used to predict the probability of an outcome, and is particularly popular for classification tasks. The algorithm predicts the probability of occurrence of an event by fitting data to a logistic function.

B: Boosting means that each tree is dependent on prior trees. The algorithm learns by fitting the residual of the trees that preceded it. Thus, boosting in a decision tree ensemble tends to improve accuracy with some small risk of less coverage.

C: Decision trees are non-parametric models that perform a sequence of simple tests for each instance, traversing a binary tree data structure until a leaf node (decision) is reached.

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/poisson-regression>

**QUESTION 58**

You have an Apache Spark cluster in Azure HDInsight. The cluster includes 200 TB in five Apache Hive tables that have multiple foreign key relationships.

You have an Azure Machine Learning model that was built by using SPARK Accelerated Failure Time (AFT) Survival Regression Model (spark.survreg).

You need to prepare the Hive data into a single table as input for the Machine Learning model.

The Hive data must be prepared in the least amount of time possible. What should you use to prepare the data?

**Guarantee All Exams 100% Pass One Time!**

- A. a Hive user-defined function (UDF)
- B. Spark SQL
- C. the GPU
- D. Java Mapreduce jobs

**Answer: A**

**Explanation:**

Create features for data in an HDInsight Hadoop cluster using Hive queries. Feature engineering attempts to increase the predictive power of learning algorithms by creating features from raw data that facilitate the learning process. You can run HiveQL queries from Azure ML, and access data processed in Hive and stored in blob storage, by using the Import Data module.

<https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-machine-learning-overview#azure-machine-learning-and-hive>

#### QUESTION 59

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

**Answer: A**

**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.

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

#### QUESTION 60

You plan to use Azure Machine Learning to develop a predictive model. You plan to include an Execute Python Script module.

What capability does the module provide?

- A. importing Python modules from a ZIP file for execution in a Machine Learning experiment
- B. performing interactive debugging of a Python script
- C. saving the results of a Python script run in a Machine Learning environment to a local file

**Guarantee All Exams 100% Pass One Time!**

D. returning multiple data frames

**Answer:** A

**Explanation:**

A common use-case for many data scientists is to incorporate existing Python scripts into Azure ML experiments. Instead of requiring that all code be concatenated and pasted into a single script box, the Execute Python Script module accepts a zip file that contains Python modules at the third input port. The file is unzipped by the execution framework at runtime and the contents are added to the library path of the Python interpreter.

<https://docs.microsoft.com/en-us/azure/machine-learning/studio/execute-python-scripts>

