

- **Vendor: Microsoft**
- **Exam Code: DP-100**
- **Exam Name: Designing and Implementing a Data Science Solution on Azure**
- **New Questions (Dec/2019)**

Visit PassLeader and Download Full Version DP-100 Exam Dumps

NEW QUESTION 1

You are conducting feature engineering to prepuce data for further analysis. The data includes seasonal patterns on inventory requirements. You need to select the appropriate method to conduct feature engineering on the data. Which method should you use?

- A. Exponential Smoothing (ETS) function.
- B. One Class Support Vector Machine module.
- C. Time Series Anomaly Detection module.
- D. Finite Impulse Response (FIR) Filter module.

Answer: D

NEW QUESTION 2

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

- A. Intel Software Guard Extensions (Intel SGX) technology.
- B. Solid State Drives (SSD).
- C. Graphic Processing Unit (GPU).
- D. Computer Processing Unit (CPU) speed increase by using overclocking.
- E. High Random Access Memory (RAM) configuration.

Answer: B

NEW QUESTION 3

You are creating a new experiment in Azure Machine Learning Studio. You have a small dataset that has missing values in many columns. The data does not require the application of predictors for each column. You plan to use the Clean Missing Data module to handle the missing data. You need to select a data cleaning method. Which method should you use?

- A. Synthetic Minority Oversampling Technique (SMOTE)
- B. Replace using MICE
- C. Replace using Probabilistic PCA
- D. Normalization

Answer: A

NEW QUESTION 4

You are a data scientist using Azure Machine Learning Studio. You need to normalize values to

produce an output column into bins to predict a target column.

Solution: Apply an Equal Width with Custom Start and Stop binning mode.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Use the Entropy MDL binning mode which has a target column.

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins>

NEW QUESTION 5

You are using Azure Machine Learning Studio to perform feature engineering on a dataset. You need to normalize values to produce a feature column grouped into bins.

Solution: Apply an Entropy Minimum Description Length (MDL) binning mode.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

Entropy MDL binning mode: This method requires that you select the column you want to predict and the column or columns that you want to group into bins. It then makes a pass over the data and attempts to determine the number of bins that minimizes the entropy. In other words, it chooses a number of bins that allows the data column to best predict the target column.

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins>

NEW QUESTION 6

You are a data scientist building a deep convolutional neural network (CNN) for image classification. The CNN model you built shows signs of overfitting. You need to reduce overfitting and converge the model to an optimal fit. Which two actions should you perform? (Each correct answer presents a complete solution. Choose two.)

- A. Reduce the amount of training data.
- B. Add an additional dense layer with 64 input units.
- C. Add L1/L2 regularization.
- D. Use training data augmentation.
- E. Add an additional dense layer with 512 input units.

Answer: BE

NEW QUESTION 7

Drag and Drop

You are building an intelligent solution using machine learning models. The environment must support the following requirements:

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

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

Actions

- Install the Azure Machine Learning SDK for Python on the cluster.
- When the cluster is ready, export Zeppelin notebooks to a local environment.
- Create and execute a Jupyter notebook by using automated machine learning (AutoML) on the cluster.
- Install Microsoft Machine Learning for Apache Spark.
- When the cluster is ready and has processed the notebook, export your Jupyter notebook to a local environment.
- Create an Azure HDInsight cluster to include the Apache Spark Mlib library.
- Create and execute the Zeppelin notebooks on the cluster.
- Create an Azure Databricks cluster.

Answer area



Answer:

Actions

- Run the `isinteg.exe` command and specify the `-fix` parameter.
- Run the `New-MailboxRestoreRequest` cmdlet.
- Install Exchange Server 2019 on a new server.

Answer Area



- Restore the database backup.
- Run the `eseutil.exe` command and specify the `/r` parameter.
- Create a recovery database.
- Run the `New-MailboxMoveRequest` cmdlet.

Explanation:

Step 1: Create an Azure HDInsight cluster to include the Apache Spark Mlib library.
 Step 2: Install Microsoft Machine Learning for Apache Spark. You install AzureML on your Azure HDInsight cluster. Microsoft Machine Learning for Apache Spark (MMLSpark) provides a number of deep learning and data science tools for Apache Spark, including seamless integration of Spark Machine Learning pipelines with Microsoft Cognitive Toolkit (CNTK) and OpenCV, enabling you to quickly create powerful, highly-scalable predictive and analytical models for large image and text datasets.
 Step 3: Create and execute the Zeppelin notebooks on the cluster.

Step 4: When the cluster is ready, export Zeppelin notebooks to a local environment. Notebooks must be exportable to be version controlled locally.

<https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-zeppelin-notebook>

<https://azuremlbuild.blob.core.windows.net/pysparkapi/intro.html>

NEW QUESTION 8

HotSpot

You create an experiment in Azure Machine Learning Studio. You add a training dataset that contains 10,000 rows. The first 9,000 rows represent class 0 (90 percent). The remaining 1,000 rows represent class 1 (10 percent). The training set is imbalances between two classes. You must increase the number of training examples for class 1 to 4,000 by using 5 data rows. You add the Synthetic Minority Oversampling Technique (SMOTE) module to the experiment. You need to configure the module. Which values should you use? (To answer, select the appropriate options in the dialog box in the answer area.)

SMOTE

Label column

Selected columns:
All labels

Launch column selector

SMOTE percentage

	▼
0	
300	
3000	
4000	

Number of nearest neighbors

	▼
0	
1	
5	
4000	

Random seed

0

Answer:

▲ SMOTE

Label column

Selected columns:

All labels

Launch column selector

SMOTE percentage

	▼
0	
300	
3000	
4000	

Number of nearest neighbors

	▼
0	
1	
5	
4000	

Random seed

0

Explanation:

Box 1: 300. You type 300 (%), the module triples the percentage of minority cases (3000) compared to the original dataset (1000).

Box 2: 5. We should use 5 data rows. Use the Number of nearest neighbors option to determine the size of the feature space that the SMOTE algorithm uses when in building new cases. A nearest neighbor is a row of data (a case) that is very similar to some target case. The distance between any two cases is measured by combining the weighted vectors of all features. By increasing the number of nearest neighbors, you get features from more cases. By keeping the number of nearest neighbors low, you use features that are more like those in the original sample.

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

NEW QUESTION 9

You are creating a machine learning model. You have a dataset that contains null rows. You need to use the Clean Missing Data module in Azure Machine Learning Studio to identify and resolve the null and missing data in the dataset. Which parameter should you use?

- A. Replace with mean
- B. Remove entire column
- C. Remove entire row
- D. Hot Deck

Answer: C

Explanation:

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

NEW QUESTION 10

You use the Two-Class Neural Network module in Azure Machine Learning Studio to build a binary classification model. You use the Tune Model Hyperparameters module to tune accuracy for the model. You need to select the hyperparameters that should be tuned using the Tune Model Hyperparameters module. Which two hyperparameters should you use? (Each correct answer presents part of the solution. Choose two.)

- A. Number of hidden nodes.
- B. Learning Rate.
- C. The type of the normalizer.
- D. Number of learning iterations.
- E. Hidden layer specification.

Answer: DE

Explanation:

D: For Number of learning iterations, specify the maximum number of times the algorithm should process the training cases.

E: For Hidden layer specification, select the type of network architecture to create. Between the input and output layers you can insert multiple hidden layers. Most predictive tasks can be accomplished easily with only one or a few hidden layers.

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

NEW QUESTION 11

.....

Case Study 1 - Sporting Events

You are a data scientist in a company that provides data science for professional sporting events. Models will be global and local market data to meet the following business goals:

.....

NEW QUESTION 101

You need to resolve the local machine learning pipeline performance issue. What should you do?

- A. Increase Graphic Processing Units (GPUs).
- B. Increase the learning rate.
- C. Increase the training iterations.
- D. Increase Central Processing Units (CPUs).

Answer: A

NEW QUESTION 102

You need to select an environment that will meet the business and data requirements. Which environment should you use?

- A. Azure HDInsight with Spark MLlib
- B. Azure Cognitive Services
- C. Azure Machine Learning Studio
- D. Microsoft Machine Learning Server

Answer: D

NEW QUESTION 103

You need to implement a scaling strategy for the local penalty detection data. Which normalization type should you use?

- A. Streaming
- B. Weight
- C. Batch
- D. Cosine

Answer: C

Explanation:

Post batch normalization statistics (PBN) is the Microsoft Cognitive Toolkit (CNTK) version of how to evaluate the population mean and variance of Batch Normalization which could be used in inference Original Paper. In CNTK, custom networks are defined using the BrainScriptNetworkBuilder and described in the CNTK network description language "BrainScript".
<https://docs.microsoft.com/en-us/cognitive-toolkit/post-batch-normalization-statistics>

NEW QUESTION 104

.....

Case Study 2 - Fabrikam

You are a data scientist for Fabrikam Residences, a company specializing in quality private and commercial property in the United States. Fabrikam Residences is considering expanding into Europe and has asked you to investigate prices for private residences in major European cities. You use Azure Machine Learning Studio to measure the median value of properties. You produce a regression model to predict property prices by using the Linear Regression and Bayesian Linear Regression modules.

.....

NEW QUESTION 107

You need to select a feature extraction method. Which method should you use?

- A. Mutual information
- B. Mood's median test
- C. Kendall correlation
- D. Permutation Feature Importance

Answer: C

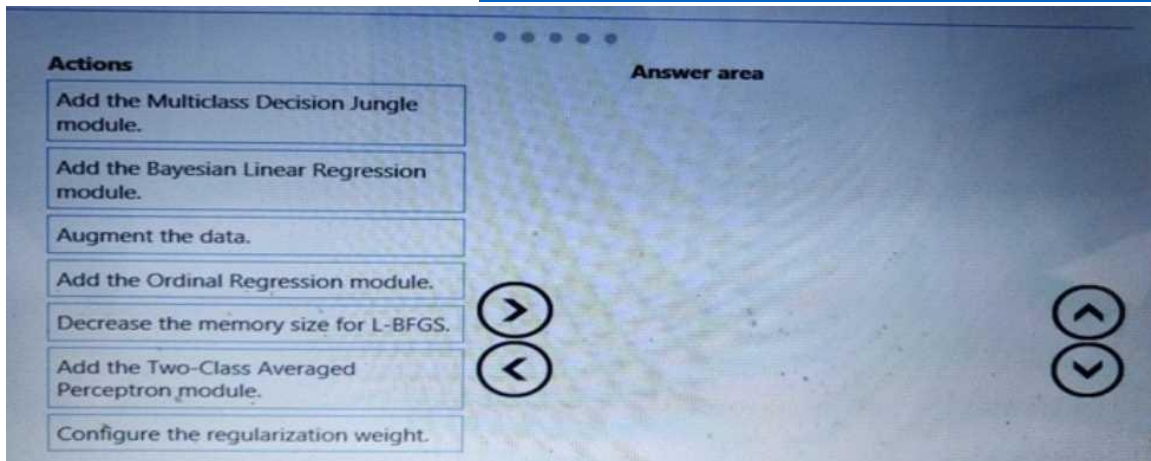
Explanation:

In statistics, the Kendall rank correlation coefficient, commonly referred to as Kendall's tau coefficient (after the Greek letter τ), is a statistic used to measure the ordinal association between two measured quantities. It is a supported method of the Azure Machine Learning Feature selection.
https://en.wikipedia.org/wiki/Kendall_rank_correlation_coefficient
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/feature-selection-modules>

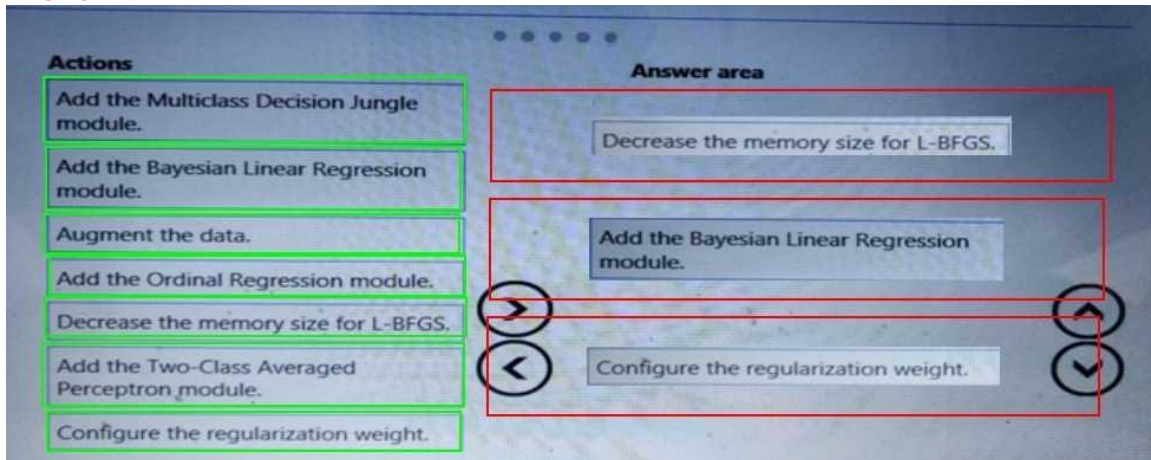
NEW QUESTION 108

Drag and Drop

You need to correct the model fit issue. Which three actions should you perform in sequence? (To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.)



Answer:



NEW QUESTION 109

HotSpot

You need to configure the Feature Based Feature Selection module based on the experiment requirements and datasets. How should you configure the module properties? (To answer, select the appropriate options in the dialog box in the answer area.)

▲ Filter Based Feature Selection

Feature scoring method

	▼
Fisher Score	
Chi-squared	
Mutual information	
Counts	

☒ Operate on feature columns only

Target column

	▼
MedianValue	
AvgRooms/nHouse	

Launch column selector

Number of desired features

1

Answer:

▲ Filter Based Feature Selection

Feature scoring method

	▼
Fisher Score	
Chi-squared	
Mutual information	
Counts	

☒ Operate on feature columns only

Target column

	▼
MedianValue	
AvgRooms/nHouse	

Launch column selector

Number of desired features

1

NEW QUESTION 110

.....

[Visit PassLeader and Download Full Version DP-100 Exam Dumps](#)