

Braindump2go Guarantee All Exams 100% Pass One Time!

Vendor: Microsoft

> Exam Code: DP-100

Exam Name: Designing and Implementing a Data Science Solution on Azure

> New Updated Questions from <u>Braindump2go</u> (Updated in <u>May/2020</u>)

Visit Braindump2go and Download Full Version DP-100 Exam Dumps

QUESTION 83

You are developing a data science workspace that uses an Azure Machine Learning service. You need to select a compote target to deploy the workspace. What should you use?

- A. Azure Data Lake Analytics
- B. Azure Databrick .
- C. Apache Spark for HDInsight.
- D. Azure Container Service

Answer: D

Explanation:

Azure Container Instances can be used as compute target for testing or development. Use for low-scale CPU-based workloads that require less than 48 GB of RAM.

Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-deploy-and-where

QUESTION 84

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

QUESTION 85

You are determining if two sets of data are significantly different from one another by using Azure Machine Learning Studio.

Estimated values in one set of data may be more than or less than reference values in the other set of data. You must produce a distribution that has a constant. Type I error as a function of the correlation.

You need to produce the distribution.

Which type of distribution should you produce?

A. Paired t-test with a two-tail option

DP-100 Exam Dumps DP-100 Exam Questions DP-100 PDF Dumps DP-100 VCE Dumps



- B. Unpaired t-test with a two tail option
- C. Paired t-test with a one-tail option
- D. Unpaired t-test with a one-tail option

Answer: D

Explanation:

Choose a one-tail or two-tail test. The default is a two-tailed test. This is the most common type of test, in which the expected distribution is symmetric around zero.

Example: Type I error of unpaired and paired two-sample t-tests as a function of the correlation. The simulated random numbers originate from a bivariate normal distribution with a variance of 1.



Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/test-hypothesis-using-t-test https://en.wikipedia.org/wiki/Student%27s_t-test

QUESTION 86

You are creating a machine learning model.

You need to identify outliers data.

Which two visualizations can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. box plot
- B. Scatter plot
- C. random forest diagram
- D. Venn diagram
- E. ROC curve

Answer: AB

Explanation:

The box-plot algorithm can be used to display outliers. One other way to quickly identify Outliers visually is to create scatter plots.

DP-100 Exam Dumps DP-100 Exam Questions DP-100 PDF Dumps DP-100 VCE Dumps



One Time!

References:

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

QUESTION 87

You are analyzing a dataset by using Azure Machine Learning Studio. YOU need to generate a statistical summary that contains the p value and the unique value count for each feature column.

Which two modules can you users? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. Execute Python Script
- B. Export Count Table
- C. Convert to Indicator Values
- D. Summarize Data
- E. Compute linear Correlation

Answer: BD

Explanation:

The Export Count Table module is provided for backward compatibility with experiments that use the Build Count Table (deprecated) and Count Featurizer (deprecated) modules.

D: Summarize Data statistics are useful when you want to understand the characteristics of the complete dataset. For example, you might need to know:

How many missing values are there in each column?

How many unique values are there in a feature column?

What is the mean and standard deviation for each column?

The module calculates the important scores for each column, and returns a row of summary statistics for each variable (data column) provided as input.

References:

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/export-count-table https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/summarize-data

QUESTION 88

You are building a binary classification model by using a supplied training set.

The training set is imbalanced between two classes.

You need to resolve the data imbalance.

What are three possible ways to achieve this goal? Each correct answer presents a complete solution NOTE: Each correct selection is worth one point.

- A. Penalize the classification
- B. Resample the data set using under sampling or oversampling
- C. Generate synthetic samples in the minority class.
- D. Use accuracy as the evaluation metric of the model.
- E. Normalize the training feature set.

Answer: BCD

QUESTION 89

You are analyzing a dataset containing historical data from a local taxi company. You arc developing a regression a regression model.

You must predict the fare of a taxi trip.

You need to select performance metrics to correctly evaluate the- regression model. Which two metrics can you use? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. an F1 score that is high
- B. an R Squared value dose to 1
- C. an R-Squared value close to 0
- D. a Root Mean Square Error value that is high

DP-100 Exam Dumps DP-100 Exam Questions DP-100 PDF Dumps DP-100 VCE Dumps



- E. a Root Mean Square Error value that is tow
- F. an F 1 score that is low.

Answer: DF

QUESTION 90

You are evaluating a completed binary classification machine learning model. You need to use the precision as the evaluation metric. Which visualization should you use?

- A. Binary classification confusion matrix
- B. box plot
- C. Gradient descent
- D. coefficient of determination

Answer: A

Explanation:

https://machinelearningknowledge.ai/confusion-matrix-and-performance-metrics-machine-learning/

QUESTION 91

You create a classification model with a dataset that contains 100 samples with Class A and 10,000 samples with Class B

The variation of Class B is very high. You need to resolve imbalances. Which method should you use?

- A. Partition and Sample
- B. Cluster Centroids
- C. Tomek links
- D. Synthetic Minority Oversampling Technique (SMOTE)

Answer: D

QUESTION 92

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. NOTE: Each correct selection is worth one point.

- 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: AC

Explanation:

C: Weight regularization provides an approach to reduce the overfitting of a deep learning neural network model on the training data and improve the performance of the model on new data, such as the holdout test set.

Keras provides a weight regularization API that allows you to add a penalty for weight size to the loss function.

Three different regularizer instances are provided; they are:

- L1: Sum of the absolute weights.
- L2: Sum of the squared weights.
- L1L2: Sum of the absolute and the squared weights.



Braindump2go Guarantee All Exams 100% Pass

One Time!

A: Because a fully connected layer occupies most of the parameters, it is prone to overfitting. One method to reduce overfitting is dropout. At each training stage, individual nodes are either "dropped out" of the net with probability 1-p or kept with probability p, so that a reduced network is left; incoming and outgoing edges to a dropped-out node are also removed.

By avoiding training all nodes on all training data, dropout decreases overfitting. References:

https://machinelearningmastery.com/how-to-reduce-overfitting-in-deep-learning-with-weight-regularization/ https://en.wikipedia.org/wiki/Convolutional_neural_network

QUESTION 93

You are with a time series dataset in Azure Machine Learning Studio.

You need to split your dataset into training and testing subsets by using the Split Data module. Which splitting mode should you use?

- A. Regular Expression Split
- B. Split Rows with the Randomized split parameter set to true
- C. Relative Expression Split
- D. Recommender Split

Answer: B

Explanation:

Split Rows: Use this option if you just want to divide the data into two parts. You can specify the percentage of data to put in each split, but by default, the data is divided 50-50.

Incorrect Answers:

B: Regular Expression Split: Choose this option when you want to divide your dataset by testing a single column for a value.

C: Relative Expression Split: Use this option whenever you want to apply a condition to a number column.

References:

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