

Microsoft

Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure



NEW QUESTION 1

- (Exam Topic 3)

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.

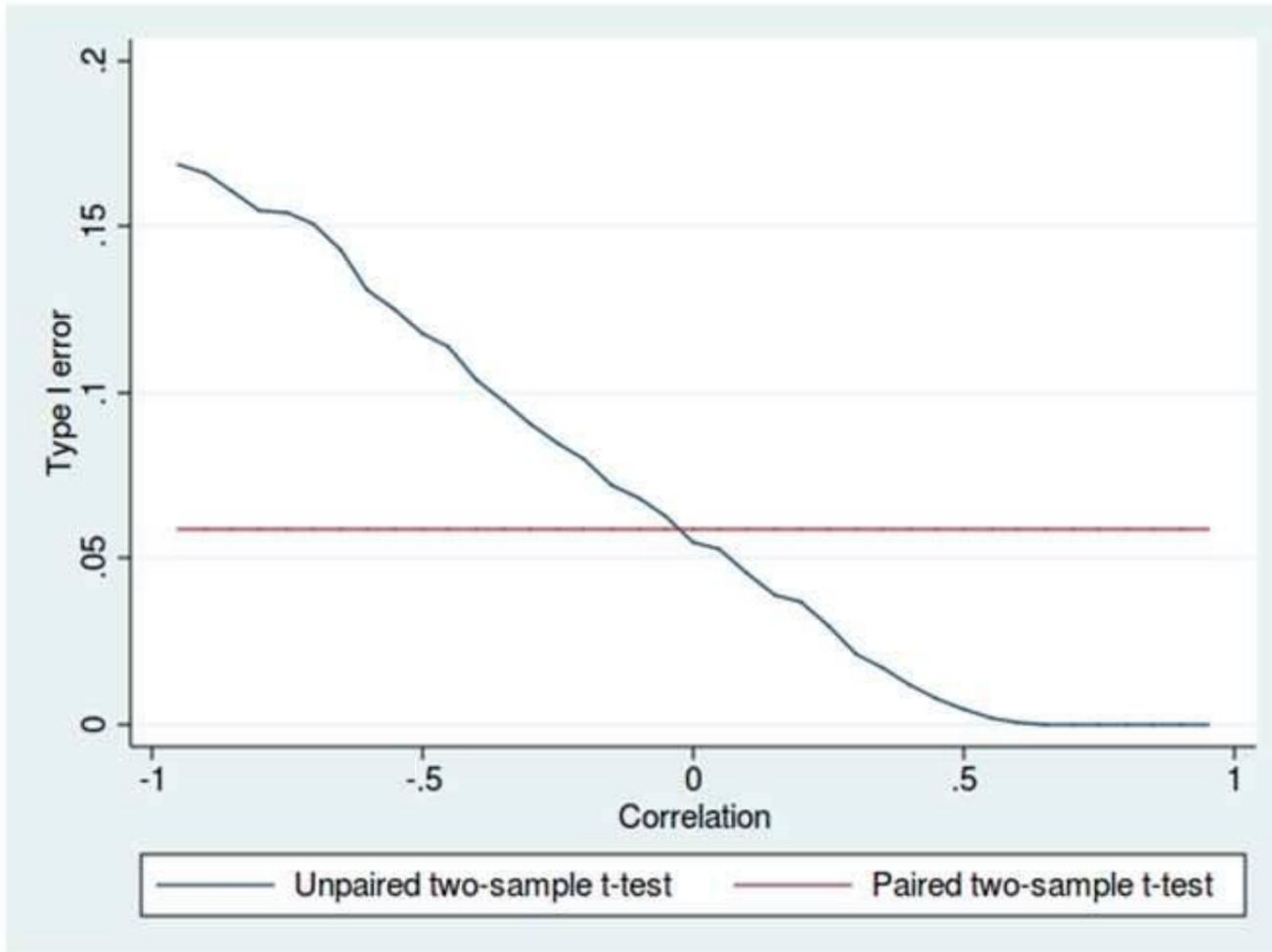
Which type of distribution should you produce?

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

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

NEW QUESTION 2

- (Exam Topic 3)

You are tuning a hyperparameter for an algorithm. The following table shows a data set with different hyperparameter, training error, and validation errors.

Hyperparameter (H)	Training error (TE)	Validation error (VE)
1	105	95
2	200	85
3	250	100
4	105	100
5	400	50

Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.

Question

Answer Choice

Which H value should you select based on the data?

	▼
1	
2	
3	
4	
5	

What H value displays the poorest training result?

	▼
1	
2	
3	
4	
5	

- A. Mastered
- B. Not Mastered

Answer: A

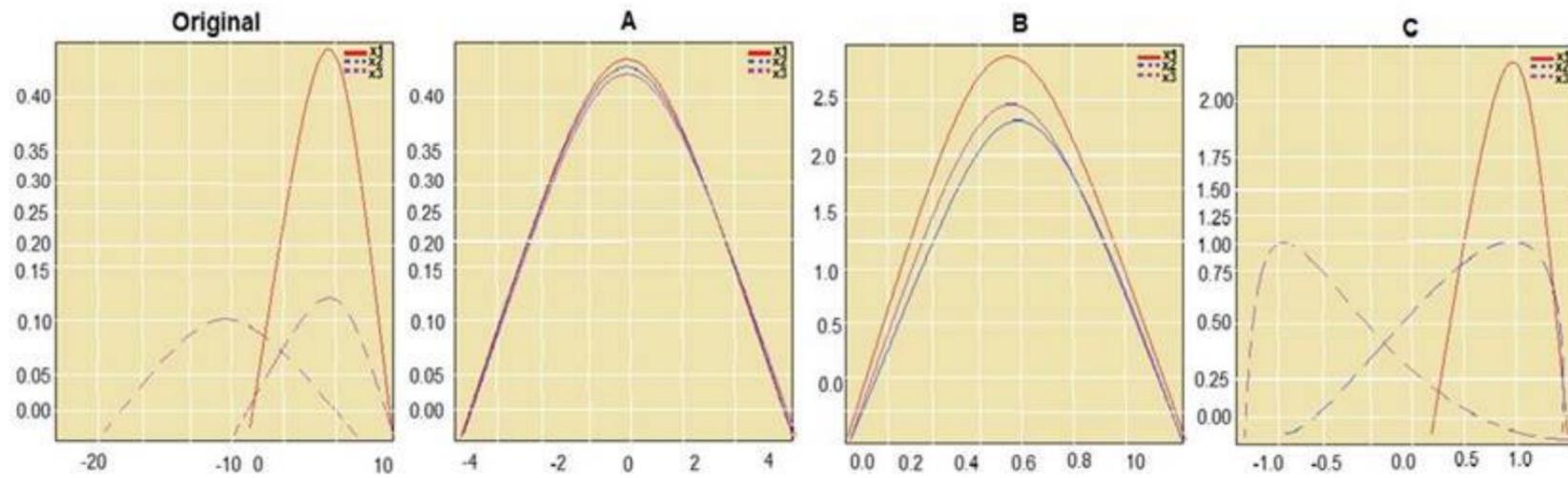
Explanation:

Box 1: 4
 Choose the one which has lower training and validation error and also the closest match. Minimize variance (difference between validation error and train error).
 Box 2: 5
 Minimize variance (difference between validation error and train error). Reference:
<https://medium.com/comet-ml/organizing-machine-learning-projects-project-management-guidelines-2d2b8565>

NEW QUESTION 3

- (Exam Topic 3)

You are performing feature scaling by using the scikit-learn Python library for x1, x2, and x3 features. Original and scaled data is shown in the following image.



Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.
 NOTE: Each correct selection is worth one point.

Question

Answer choice

Which scaler is used in graph A?

▼
Standard Scaler
Min Max Scale
Normalizer

Which scaler is used in graph B?

▼
Standard Scaler
Min Max Scale
Normalizer

Which scaler is used in graph C?

▼
Standard Scaler
Min Max Scale
Normalizer

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: StandardScaler

The StandardScaler assumes your data is normally distributed within each feature and will scale them such that the distribution is now centred around 0, with a standard deviation of 1.

Example:

All features are now on the same scale relative to one another. Box 2: Min Max Scaler

Notice that the skewness of the distribution is maintained but the 3 distributions are brought into the same scale so that they overlap.

Box 3: Normalizer References:

<http://benalexkeen.com/feature-scaling-with-scikit-learn/>

NEW QUESTION 4

- (Exam Topic 3)

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: BE

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.

E: Summarize Data statistics are useful when you want to understand the characteristics of the complete dataset. For example, you might need to know: How many missing values are there in each column? How many unique values are there in a feature column?

What is the mean and standard deviation for each column?

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

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>

NEW QUESTION 5

- (Exam Topic 3)

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

NEW QUESTION 6

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are a data scientist using Azure Machine Learning Studio.

You need to normalize values to produce an output column into bins to predict a target column. Solution: Apply a Quantiles normalization with a QuantileIndex normalization.

Does the solution meet the GOAL?

- A. Yes
- B. No

Answer: B

Explanation:

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

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

NEW QUESTION 7

- (Exam Topic 3)

You are performing feature engineering on a dataset.

You must add a feature named CityName and populate the column value with the text London.

You need to add the new feature to the dataset.

Which Azure Machine Learning Studio module should you use?

- A. Edit Metadata
- B. Preprocess Text
- C. Execute Python Script
- D. Latent Dirichlet Allocation

Answer: A

Explanation:

Typical metadata changes might include marking columns as features. References:

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

NEW QUESTION 8

- (Exam Topic 3)

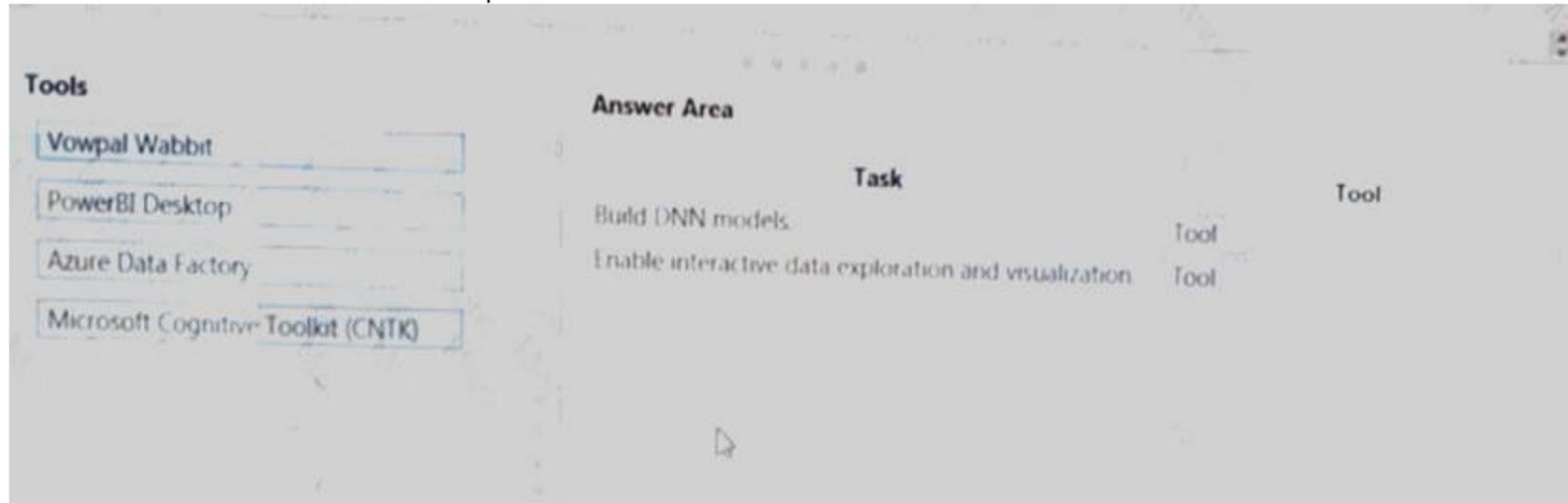
You configure a Deep Learning Virtual Machine for Windows.

You need to recommend tools and frameworks to perform the following: Build deep neural network (DNN) models.

Perform interactive data exploration and visualization.

Which tools and frameworks should you recommend? To answer, drag the appropriate tools to the correct tasks. Each tool may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

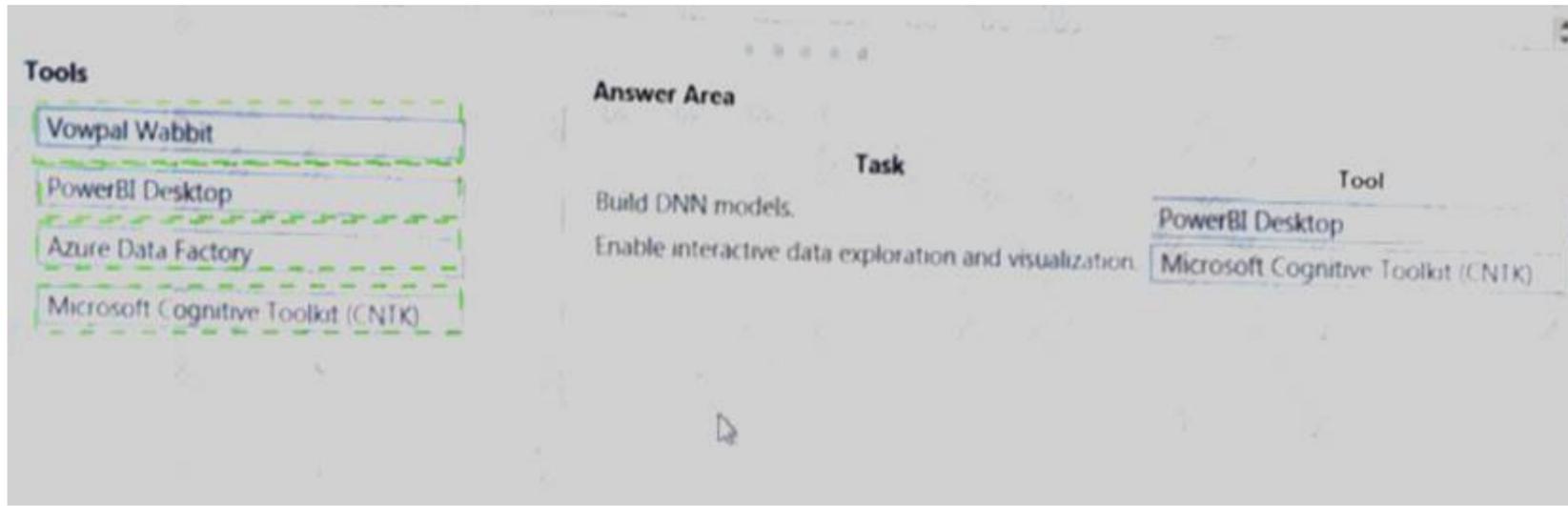
NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 9

- (Exam Topic 3)

You use Azure Machine Learning Studio to build a machine learning experiment. You need to divide data into two distinct datasets. Which module should you use?

- A. Partition and Sample
- B. Assign Data to Clusters
- C. Group Data into Bins
- D. Test Hypothesis Using t-Test

Answer: A

Explanation:

Partition and Sample with the Stratified split option outputs multiple datasets, partitioned using the rules you specified.

References:

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

NEW QUESTION 10

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are 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.

References:

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

NEW QUESTION 10

- (Exam Topic 2)

You need to set up the Permutation Feature Importance module according to the model training requirements.

Which properties should you select? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

▾ Tune Model Hyperparameters

Specify parameter sweeping mode

Random sweep ▼

Maximum number of runs on random sweep

5

Random seed

0

Label column

Selected columns:
 Column names: MedianValue

Launch column selector

Metric for measuring performance for classification

▼
 F-score
 Precision
 Recall
 Accuracy

Metric for measuring performance for regression

▼
 Root of mean squared error
 R-squared
 Mean zero one error
 Mean absolute error

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Accuracy

Scenario: You want to configure hyperparameters in the model learning process to speed the learning phase by using hyperparameters. In addition, this configuration should cancel the lowest performing runs at each evaluation interval, thereby directing effort and resources towards models that are more likely to be successful.

Box 2: R-Squared

NEW QUESTION 11

- (Exam Topic 2)

You need to produce a visualization for the diagnostic test evaluation according to the data visualization requirements.

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

Modules	Answer Area
Score Matchbox Recommender	
Apply Transformation	
Evaluate Recommender	
Evaluate Model	⏪
Train Model	⏩
Sweep Clustering	
Score Model	⏴
Load Trained Model	⏵

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Sweep Clustering

Start by using the "Tune Model Hyperparameters" module to select the best sets of parameters for each of the models we're considering.

One of the interesting things about the "Tune Model Hyperparameters" module is that it not only outputs the results from the Tuning, it also outputs the Trained Model.

Step 2: Train Model Step 3: Evaluate Model

Scenario: You need to provide the test results to the Fabrikam Residences team. You create data visualizations to aid in presenting the results.

You must produce a Receiver Operating Characteristic (ROC) curve to conduct a diagnostic test evaluation of the model. You need to select appropriate methods for producing the ROC curve in Azure Machine Learning Studio to compare the Two-Class Decision Forest and the Two-Class Decision Jungle modules with one another.

References:

<http://breaking-bi.blogspot.com/2017/01/azure-machine-learning-model-evaluation.html>

NEW QUESTION 13

- (Exam Topic 2)

You need to configure the Permutation Feature Importance module for the model training requirements. What should you do? To answer, select the appropriate options in the dialog box in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

Permutation Feature importance

Random seed

	▼
0	
500	

	▼
Regression – Root Mean Square Error	
Regression – R-squared	
Regression – Mean Zero One Error	
Regression – Mean Absolute Error	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: 500

For Random seed, type a value to use as seed for randomization. If you specify 0 (the default), a number is generated based on the system clock.

A seed value is optional, but you should provide a value if you want reproducibility across runs of the same experiment.

Here we must replicate the findings. Box 2: Mean Absolute Error

Scenario: Given a trained model and a test dataset, you must compute the Permutation Feature Importance scores of feature variables. You need to set up the Permutation Feature Importance module to select the correct metric to investigate the model's accuracy and replicate the findings.

Regression. Choose one of the following: Precision, Recall, Mean Absolute Error, Root Mean Squared Error, Relative Absolute Error, Relative Squared Error, Coefficient of Determination

References:

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

NEW QUESTION 18

- (Exam Topic 2)

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.

Actions	Answer Area	
Add the Ordinal Regression module.		
Add the Two-Class Averaged Perception module.		
Augment the data.		
Add the Bayesian Linear Regression module.		
Decrease the memory size for L-BFGS.		
Add the Multiclass Decision Jungle module.		
Configure the regularization weight.		

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Augment the data

Scenario: Columns in each dataset contain missing and null values. The datasets also contain many outliers.

Step 2: Add the Bayesian Linear Regression module.

Scenario: You produce a regression model to predict property prices by using the Linear Regression and Bayesian Linear Regression modules.

Step 3: Configure the regularization weight.

Regularization typically is used to avoid overfitting. For example, in L2 regularization weight, type the value to use as the weight for L2 regularization. We recommend that you use a non-zero value to avoid overfitting.

Scenario:

Model fit: The model shows signs of overfitting. You need to produce a more refined regression model that reduces the overfitting.

NEW QUESTION 19

- (Exam Topic 1)

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."

Scenario:

Local penalty detection models must be written by using BrainScript. References:

<https://docs.microsoft.com/en-us/cognitive-toolkit/post-batch-normalization-statistics>

NEW QUESTION 21

- (Exam Topic 1)

You need to use the Python language to build a sampling strategy for the global penalty detection models. How should you complete the code segment? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

```
import pytorch as deeplearninglib
import tensorflow as deeplearninglib
import cntk as deeplearninglib
```

```
train_sampler = deeplearninglib.DistributedSampler(penalty_video_dataset)
train_sampler = deeplearninglib.log_uniform_candidate_sampler(penalty_video_dataset)
train_sampler = deeplearninglib.WeightedRandomSampler(penalty_video_dataset)
train_sampler = deeplearninglib.all_candidate_sampler(penalty_video_dataset)
```

```
...
train_loader =
...
(train_sampler, penalty_video_dataset)
```

```
optimizer = deeplearninglib.optim.SGD(model.parameters(), lr=0.01)
optimizer = deeplearninglib.train.GradientDescentOptimizer(learning_rate=0.10)
```

```
model = deeplearninglib.parallel.Distributed(DataParallel(model))
model = deeplearninglib.nn.parallel.DistributedDataParallel(model)
model = deeplearninglib.keras.Model([
model = deeplearninglib.keras.Sequential([
```

```
...
train_sampler.set_epoch(epoch)
for data, target in train_loader:
    data, target = data.to(device), target.to(device)
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: import pytorch as deeplearninglib Box 2: ..DistributedSampler(Sampler).. DistributedSampler(Sampler): Sampler that restricts data loading to a subset of the dataset.

It is especially useful in conjunction with class: `torch.nn.parallel.DistributedDataParallel`. In such case, each process can pass a DistributedSampler instance as a DataLoader sampler, and load a subset of the original dataset that is exclusive to it.

Scenario: Sampling must guarantee mutual and collective exclusivity between local and global segmentation models that share the same features.

Box 3: optimizer = deeplearninglib.train.GradientDescentOptimizer(learning_rate=0.10)

NEW QUESTION 26

- (Exam Topic 1)

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 30

- (Exam Topic 1)

You need to build a feature extraction strategy for the local models.

How should you complete the code segment? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

```
with C.layers.default_options(init=C.glorot_uniform(), activation=C.relu):
h = features
```

```
h = C.layers.Convolution2D(num_filters=8...)(h)
h = C.layers.MaxPooling(filter_shape=(3,3)...)(h)
h = C.layers.Convolution2D(num_filters=16...)(h)
h = C.layers.MaxPooling(filter_shape=(2,2)...)(h)
```

```
r = C.layers.Dense...
```

```
h = C.layers.MaxPooling(filter_shape=(3,3)...)(h)
h = C.layers.MaxPooling(filter_shape=(2,2)...)(h)
h = C.layers.Convolution2D(num_filters=8...)(h)
h = C.layers.Convolution2D(num_filters=16...)(h)
```

```
h = C.layers.Convolution2D(num_filters=16...)(h)
h = C.layers.Convolution2D(num_filters=8...)(h)
h = C.layers.MaxPooling(filter_shape=(2,2)...)(h)
h = C.layers.MaxPooling(filter_shape=(3,3)...)(h)
```

```
h = C.layers.MaxPooling(filter_shape=(3,3)...)(h)
h = C.layers.MaxPooling(filter_shape=(2,2)...)(h)
h = C.layers.Convolution2D(num_filters=8...)(h)
h = C.layers.Convolution2D(num_filters=16...)(h)
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

```
with C.layers.default_options(init=C.glorot_uniform(), activation=C.relu):
h = features
```

```
h = C.layers.Convolution2D(num_filters=8...)(h)
h = C.layers.MaxPooling(filter_shape=(3,3)...)(h)
h = C.layers.Convolution2D(num_filters=16...)(h)
h = C.layers.MaxPooling(filter_shape=(2,2)...)(h)
```

```
r = C.layers.Dense...
```

```
h = C.layers.MaxPooling(filter_shape=(3,3)...)(h)
h = C.layers.MaxPooling(filter_shape=(2,2)...)(h)
h = C.layers.Convolution2D(num_filters=8...)(h)
h = C.layers.Convolution2D(num_filters=16...)(h)
```

```
h = C.layers.Convolution2D(num_filters=16...)(h)
h = C.layers.Convolution2D(num_filters=8...)(h)
h = C.layers.MaxPooling(filter_shape=(2,2)...)(h)
h = C.layers.MaxPooling(filter_shape=(3,3)...)(h)
```

```
h = C.layers.MaxPooling(filter_shape=(3,3)...)(h)
h = C.layers.MaxPooling(filter_shape=(2,2)...)(h)
h = C.layers.Convolution2D(num_filters=8...)(h)
h = C.layers.Convolution2D(num_filters=16...)(h)
```

NEW QUESTION 34

- (Exam Topic 1)

You need to define an evaluation strategy for the crowd sentiment models.

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.

Actions

Answer Area

- Define a cross-entropy function activation.
- Add cost functions for each target state.
- Evaluate the classification error metric.
- Evaluate the distance error metric.
- Add cost functions for each component metric.
- Define a sigmoid loss function activation.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Define a cross-entropy function activation

When using a neural network to perform classification and prediction, it is usually better to use cross-entropy error than classification error, and somewhat better to use cross-entropy error than mean squared error to evaluate the quality of the neural network.

Step 2: Add cost functions for each target state. Step 3: Evaluated the distance error metric. References:
<https://www.analyticsvidhya.com/blog/2018/04/fundamentals-deep-learning-regularization-techniques/>

NEW QUESTION 36

- (Exam Topic 3)

You are working on a classification task. You have a dataset indicating whether a student would like to play soccer and associated attributes. The dataset includes the following columns:

You need to classify variables by type.

Which variable should you add to each category? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Category

Variables

Categorical variables	<div style="border: 1px solid gray; padding: 5px;"> <p>Gender, IsPlaySoccer</p> <p>Gender, PrevExamMarks, Height, Weight</p> <p>PrevExamMarks, Height, Weight</p> <p>IsPlaySoccer</p> </div>
Continuous variables	<div style="border: 1px solid gray; padding: 5px;"> <p>Gender, IsPlaySoccer</p> <p>Gender, PrevExamMarks, Height, Weight</p> <p>PrevExamMarks, Height, Weight</p> <p>IsPlaySoccer</p> </div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

References: <https://www.edureka.co/blog/classification-algorithms/>

NEW QUESTION 38

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are creating a new experiment in Azure Learning learning Studio.

One class has a much smaller number of observations than the other classes in the training

You need to select an appropriate data sampling strategy to compensate for the class imbalance. Solution: You use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode. Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

SMOTE is used to increase the number of underrepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.

References:

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

NEW QUESTION 43

- (Exam Topic 3)

You create a binary classification model. You need to evaluate the model performance.

Which two metrics can you use? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. relative absolute error
- B. precision
- C. accuracy
- D. mean absolute error
- E. coefficient of determination

Answer: BC

Explanation:

The evaluation metrics available for binary classification models are: Accuracy, Precision, Recall, F1 Score, and AUC.

Note: A very natural question is: 'Out of the individuals whom the model, how many were classified correctly (TP)?'

This question can be answered by looking at the Precision of the model, which is the proportion of positives that are classified correctly.

References:

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

NEW QUESTION 45

- (Exam Topic 3)

You are developing a hands-on workshop to introduce Docker for Windows to attendees. You need to ensure that workshop attendees can install Docker on their devices.

Which two prerequisite components should attendees install on the devices? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Microsoft Hardware-Assisted Virtualization Detection Tool
- B. Kitematic
- C. BIOS-enabled virtualization
- D. VirtualBox
- E. Windows 10 64-bit Professional

Answer: E

Explanation:

C: Make sure your Windows system supports Hardware Virtualization Technology and that virtualization is enabled.

Ensure that hardware virtualization support is turned on in the BIOS settings. For example:



E: To run Docker, your machine must have a 64-bit operating system running Windows 7 or higher. References:

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

NEW QUESTION 50

- (Exam Topic 3)

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

- A. Data Science Virtual Machine for Linux (CentOS)
- B. Data Science Virtual Machine for Windows 2012
- C. Data Science Virtual Machine for Windows 2016
- D. Geo AI Data Science Virtual Machine with ArcGIS
- E. Data Science Virtual Machine for Linux (Ubuntu)

Answer: E

NEW QUESTION 54

- (Exam Topic 3)

You use Data Science Virtual Machines (DSVMs) for Windows and Linux in Azure. You need to access the DSVMs.

Which utilities should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Requirement	Utility
Use terminal sessions to access a DSVM for Linux.	<input type="checkbox"/> SSH Client <input checked="" type="checkbox"/> X2Go <input type="checkbox"/> JupyterLab <input type="checkbox"/> Remote Desktop
Access Jupyter notebooks on a DSVM for Linux.	<input type="checkbox"/> SSH Client <input type="checkbox"/> X2Go <input checked="" type="checkbox"/> JupyterLab <input type="checkbox"/> Remote Desktop
Access Jupyter notebooks on a DSVM for Linux.	<input type="checkbox"/> SSH Client <input type="checkbox"/> X2Go <input checked="" type="checkbox"/> JupyterLab <input type="checkbox"/> Remote Desktop
Access a DSVM for Windows.	<input type="checkbox"/> SSH Client <input type="checkbox"/> X2Go <input type="checkbox"/> JupyterLab <input checked="" type="checkbox"/> Remote Desktop

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Requirement	Utility
Use terminal sessions to access a DSVM for Linux.	<input type="checkbox"/> SSH Client <input checked="" type="checkbox"/> X2Go <input type="checkbox"/> JupyterLab <input type="checkbox"/> Remote Desktop
Access Jupyter notebooks on a DSVM for Linux.	<input type="checkbox"/> SSH Client <input type="checkbox"/> X2Go <input checked="" type="checkbox"/> JupyterLab <input type="checkbox"/> Remote Desktop
Access Jupyter notebooks on a DSVM for Linux.	<input type="checkbox"/> SSH Client <input type="checkbox"/> X2Go <input checked="" type="checkbox"/> JupyterLab <input type="checkbox"/> Remote Desktop
Access a DSVM for Windows.	<input type="checkbox"/> SSH Client <input type="checkbox"/> X2Go <input type="checkbox"/> JupyterLab <input checked="" type="checkbox"/> Remote Desktop

NEW QUESTION 58

- (Exam Topic 3)

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

NEW QUESTION 60

- (Exam Topic 3)

You are creating a machine learning model. You need to identify outliers in the data.

Which two visualizations can you use? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point. NOTE: Each correct selection is worth one point.

- A. box plot
- B. scatter
- 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. References:

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

NEW QUESTION 64

- (Exam Topic 3)

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 66

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are creating a new experiment in Azure Machine Learning Studio.

One class has a much smaller number of observations than tin- other classes in the training set. You need to select an appropriate data sampling strategy to compensate for the class imbalance. Solution: You use the Principal Components Analysis (PCA) sampling mode.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

NEW QUESTION 68

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are creating a model to predict the price of a student's artwork depending on the following variables: the student's length of education, degree type, and art form.

You start by creating a linear regression model. You need to evaluate the linear regression model.

Solution: Use the following metrics: Accuracy, Precision, Recall, F1 score and AUC. Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Those are metrics for evaluating classification models, instead use: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error, Relative Squared Error, and the Coefficient of Determination.

References:

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

NEW QUESTION 69

- (Exam Topic 3)

You are developing a machine learning, experiment by using Azure. The following images show the input and output of a machine learning experiment:

rows	columns	Server ID	Risk Level	Cost
4	3	N102696	High	4500
		N102874	Low	5000
		N107027	Medium	4000
		N106548	High	4800

rows	columns	Server ID	Risk Level-High	Risk Level-Low	Risk Level-Medium	Cost
4	5	N102696	1	0	0	4500
		N102874	0	1	0	5000
		N107027	0	0	1	4000
		N106548	1	0	0	4800

Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.
 NOTE: Each correct selection is worth one point.

You need to perform the data transformation applied to the Risk Level column. Which module should you use?

What is the expected input column type for this transformation?

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

You need to perform the data transformation applied to the Risk Level column. Which module should you use?

What is the expected input column type for this transformation?

NEW QUESTION 71

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are creating a model to predict the price of a student's artwork depending on the following variables: the student's length of education, degree type, and art form.

You start by creating a linear regression model. You need to evaluate the linear regression model.

Solution: Use the following metrics: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error, Accuracy, Precision, Recall, F1 score, and AUC. Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Accuracy, Precision, Recall, F1 score, and AUC are metrics for evaluating classification models. Note: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error are OK for the linear regression model.

References:

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

NEW QUESTION 73

- (Exam Topic 3)

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	Answer area
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.	

- A. Mastered
- B. Not Mastered

Answer: A

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.

References:

<https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-zeppelin-notebook> <https://azuremlbuild.blob.core.windows.net/pysparkapi/intro.html>

NEW QUESTION 74

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are analyzing a numerical dataset which contains missing values in several columns.

You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set.

You need to analyze a full dataset to include all values.

Solution: Replace each missing value using the Multiple Imputation by Chained Equations (MICE) method. Does the solution meet the goal?

- A. Yes
- B. NO

Answer: A

Explanation:

Replace using MICE: For each missing value, this option assigns a new value, which is calculated by using a method described in the statistical literature as "Multivariate Imputation using Chained Equations" or "Multiple Imputation by Chained Equations". With a multiple imputation method, each variable with missing data is modeled conditionally using the other variables in the data before filling in the missing values.

Note: Multivariate imputation by chained equations (MICE), sometimes called "fully conditional specification" or "sequential regression multiple imputation" has emerged in the statistical literature as one principled method of addressing missing data. Creating multiple imputations, as opposed to single imputations, accounts for the statistical uncertainty in the imputations. In addition, the chained equations approach is very flexible and can handle variables of varying types (e.g., continuous or binary) as well as complexities such as bounds or survey skip patterns.

References: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/>

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

NEW QUESTION 76

- (Exam Topic 3)

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. NOTE: Each correct selection is worth one point.

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

- A. Mastered
- B. Not Mastered

Answer: A

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.

References:

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

NEW QUESTION 80

- (Exam Topic 3)

You are training a deep learning model to identify cats and dogs. You have 25,000 color images.

You must meet the following requirements:

- Reduce the number of training epochs.
- Reduce the size of the neural network.
- Reduce over-fitting of the neural network.

You need to select the image modification values.

Which value should you use? To answer, select the appropriate Options in the answer area. NOTE: Each correct selection is worth one point.

Image Modification

Convert the image color.

Resize the image size.

grayscale

color RGB

color CMYK

GaussianBlur

15x15 pixels

70x70 pixels

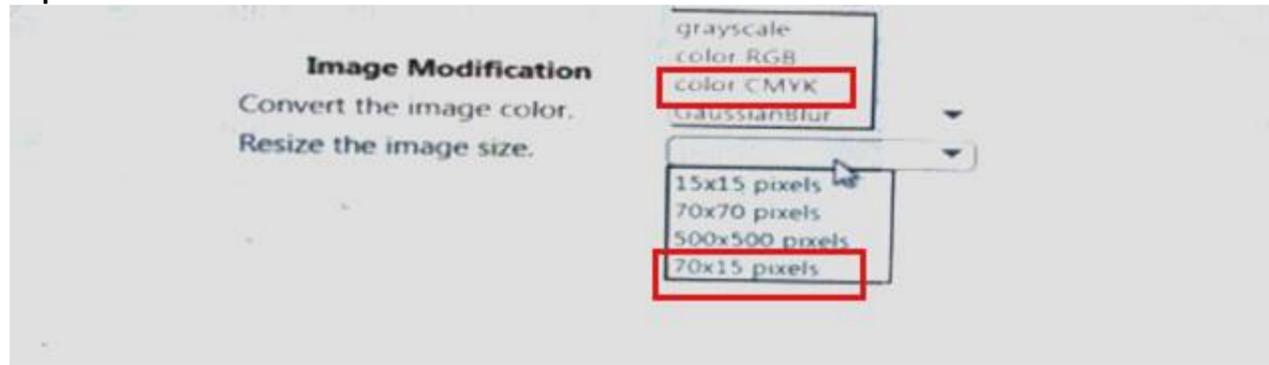
500x500 pixels

70x15 pixels

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 82

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are analyzing a numerical dataset which contains missing values in several columns.

You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set.

You need to analyze a full dataset to include all values.

Solution: Remove the entire column that contains the missing data point. Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Use the Multiple Imputation by Chained Equations (MICE) method. References: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/>
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data>

NEW QUESTION 84

- (Exam Topic 3)

You create a binary classification model by using Azure Machine Learning Studio.

You must tune hyperparameters by performing a parameter sweep of the model. The parameter sweep must meet the following requirements:

- > iterate all possible combinations of hyperparameters
- > minimize computing resources required to perform the sweep
- > You need to perform a parameter sweep of the model.

Which parameter sweep mode should you use?

- A. Random sweep
- B. Sweep clustering
- C. Entire grid
- D. Random grid
- E. Random seed

Answer: D

Explanation:

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

If you are training a model that supports an integrated parameter sweep, you can also set a range of seed values to use and iterate over the random seeds as well. This is optional, but can be useful for avoiding bias introduced by seed selection.

NEW QUESTION 89

- (Exam Topic 3)

You are building a machine learning model for translating English language textual content into French language textual content.

You need to build and train the machine learning model to learn the sequence of the textual content. Which type of neural network should you use?

- A. Multilayer Perceptions (MLPs)
- B. Convolutional Neural Networks (CNNs)
- C. Recurrent Neural Networks (RNNs)
- D. Generative Adversarial Networks (GANs)

Answer: C

Explanation:

To translate a corpus of English text to French, we need to build a recurrent neural network (RNN).

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

References:

<https://towardsdatascience.com/language-translation-with-rnns-d84d43b40571>

NEW QUESTION 90

- (Exam Topic 3)

You are retrieving data from a large datastore by using Azure Machine Learning Studio.

You must create a subset of the data for testing purposes using a random sampling seed based on the system clock.

You add the Partition and Sample module to your experiment. You need to select the properties for the module.

Which values should you select? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Partition and Sample

Partition or sample mode

▼

Assign to Folds
Pick Fold
Sampling
Head

Rate of sampling

.2

Random seed for sampling

▼

0
1
time.clock()
utcNow()

Stratified split for sampling

False ▼

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Sampling Create a sample of data

This option supports simple random sampling or stratified random sampling. This is useful if you want to create a smaller representative sample dataset for testing.

1. Add the Partition and Sample module to your experiment in Studio, and connect the dataset.

2. Partition or sample mode: Set this to Sampling.

3. Rate of sampling. See box 2 below. Box 2: 0

3. Rate of sampling. Random seed for sampling: Optionally, type an integer to use as a seed value.

This option is important if you want the rows to be divided the same way every time. The default value is 0, meaning that a starting seed is generated based on the system clock. This can lead to slightly different results each time you run the experiment.

References:

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

NEW QUESTION 93

- (Exam Topic 3)

You have a feature set containing the following numerical features: X, Y, and Z.

The Poisson correlation coefficient (r-value) of X, Y, and Z features is shown in the following image:

Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.

NOTE: Each correct selection is worth one point.

What is the r-value for the correlation of Y to Z?

▼

-0.106276
0.149676
0.859122
1

Which type of relationship exists between Z and Y in the feature set?

▼

a positive linear relationship
a negative linear relationship
no linear relationship

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: 0.859122

Box 2: a positively linear relationship

+1 indicates a strong positive linear relationship

-1 indicates a strong negative linear correlation

0 denotes no linear relationship between the two variables. References:

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

NEW QUESTION 95

- (Exam Topic 3)

You are evaluating a completed binary classification machine learning model. You need to use the precision as the valuation metric.

Which visualization should you use?

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

Answer: A

Explanation:

References:

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

NEW QUESTION 99

- (Exam Topic 3)

You are creating a binary classification by using a two-class logistic regression model. You need to evaluate the model results for imbalance.

Which evaluation metric should you use?

- A. Relative Absolute Error
- B. AUC Curve
- C. Mean Absolute Error
- D. Relative Squared Error

Answer: B

Explanation:

One can inspect the true positive rate vs. the false positive rate in the Receiver Operating Characteristic (ROC) curve and the corresponding Area Under the Curve (AUC) value. The closer this curve is to the upper left corner, the better the classifier's performance is (that is maximizing the true positive rate while minimizing the false positive rate). Curves that are close to the diagonal of the plot, result from classifiers that tend to make predictions that are close to random guessing.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio/evaluate-model-performance#evaluating-a-bina>

NEW QUESTION 103

.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

DP-100 Practice Exam Features:

- * DP-100 Questions and Answers Updated Frequently
- * DP-100 Practice Questions Verified by Expert Senior Certified Staff
- * DP-100 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- * DP-100 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
[Order The DP-100 Practice Test Here](#)