Microsoft

DP-100

Microsoft Designing and Implementing a Data Science Solution on Azure

QUESTION & ANSWERS

QUESTION 1

Case Study	Number of Questions	Total Question
Case Study: 1	19	1 – 19
Case Study: 2	24	20 – 43
Case Study: 3	210	44 - 253
Total		253

Case Study: 1

Overview

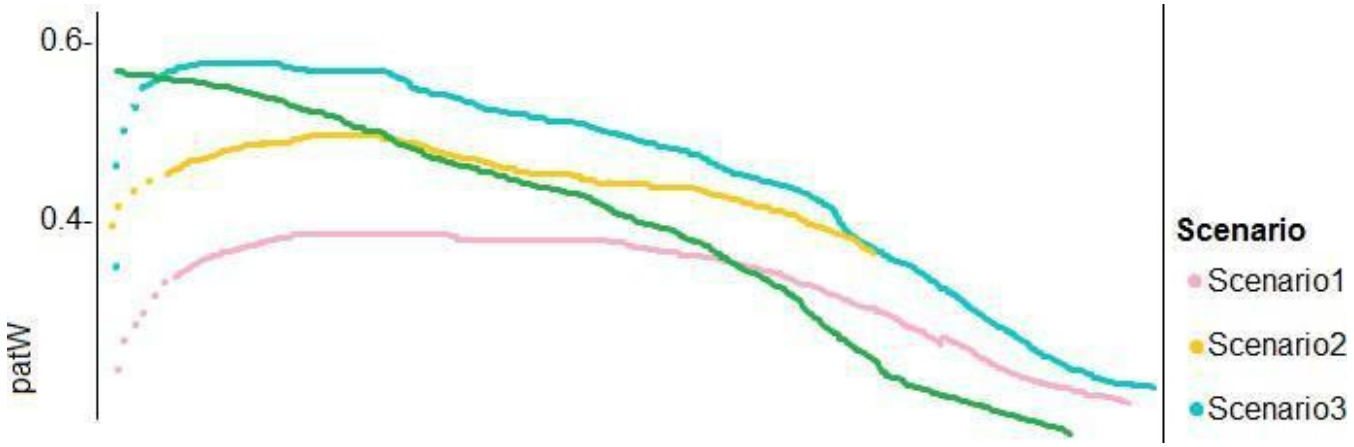
Requirements
* Media used for penalty event detection will be provided by consumer devices. Media may include images and videos captured during the sporting event and snared using social media. The images and videos will have varying sizes and formats.
* The data available for model building comprises of seven years of sporting event media. The sporting event media includes: recorded videos, transcripts of radio commentary, and logs from related social media feeds feeds captured during the sporting events.
*Crowd sentiment will include audio recordings submitted by event attendees in both mono and stereo
Formats.
Advertisements
* Ad response models must be trained at the beginning of each event and applied during the sporting event.
* Market segmentation nxxlels must optimize for similar ad resporr.r history.
* Sampling must guarantee mutual and collective exclusivity local and global segmentation models that share the same features.
* Local market segmentation models will be applied before determining a user's propensity to respond to an advertisement.
* Data scientists must be able to detect model degradation and decay.
* Ad response models must support non linear boundaries features.
* The ad propensity model uses a cut threshold is 0.45 and retrains occur if weighted Kappa deviates from 0.1 +/-5%.
* The ad propensity model uses cost factors shown in the following diagram:

		Actual	
		1	0
Predicted	0	1	2
	1	2	1

The ad propensity model uses proposed cost factors shown in the following diagram:

		Actual	
		1	0
Predicted	0	1	5
	1	5	1

Performance curves of current and proposed cost factor scenarios are shown in the following diagram:



Penalty detection and sentiment

Findings

*Data scientists must build an intelligent solution by using multiple machine learning models for penalty event detection.

*Data scientists must build notebooks in a local environment using automatic feature engineering and model building in machine learning pipelines.

*Notebooks must be deployed to retrain by using Spark instances with dynamic worker allocation

*Notebooks must execute with the same code on new Spark instances to recode only the source of the data.

*Global penalty detection models must be trained by using dynamic runtime graph computation during training.

*Local penalty detection models must be written by using BrainScript.

* Experiments for local crowd sentiment models must combine local penalty detection data.

* Crowd sentiment models must identify known sounds such as cheers and known catch phrases. Individual crowd sentiment models will detect similar sounds.

* All shared features for local models are continuous variables.

* Shared features must use double precision. Subsequent layers must have aggregate running mean and standard deviation metrics Available.

segments

During the initial weeks in production, the following was observed:

*Ad response rates declined.

*Drops were not consistent across ad styles.

*The distribution of features across training and production data are not consistent.

Analysis shows that of the 100 numeric features on user location and behavior, the 47 features that come from location sources are being used as raw features. A suggested experiment to remedy the bias and variance issue is to engineer 10 linearly uncorrected features.

Penalty detection and sentiment

*Initial data discovery shows a wide range of densities of target states in training data used for crowd sentiment models.
*All penalty detection models show inference phases using a Stochastic Gradient Descent (SGD) are running too stow.
*Audio samples show that the length of a catch phrase varies between 25%-47%, depending on region.
*The performance of the global penalty detection models show lower variance but higher bias when comparing training and validation sets. Before implementing any feature changes, you must confirm the bias and variance using all training and validation cases.

Question No. 1

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).

Correct Answer: A

QUESTION 2

You need to identify the methods for dividing the data according to the testing requirements. Which properties should you select? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Properties Project

Partition and Sample

Assign to Folds
Sampling
Head

Partition or sample mode

☐ Use replacement in the partitioning

☑ Randomized split

Random seed

0

True
False
Partition evenly
Partition with custom partitions

Specify the partitioner method

Partition evenly

Specify number of folds to split evenly into

3

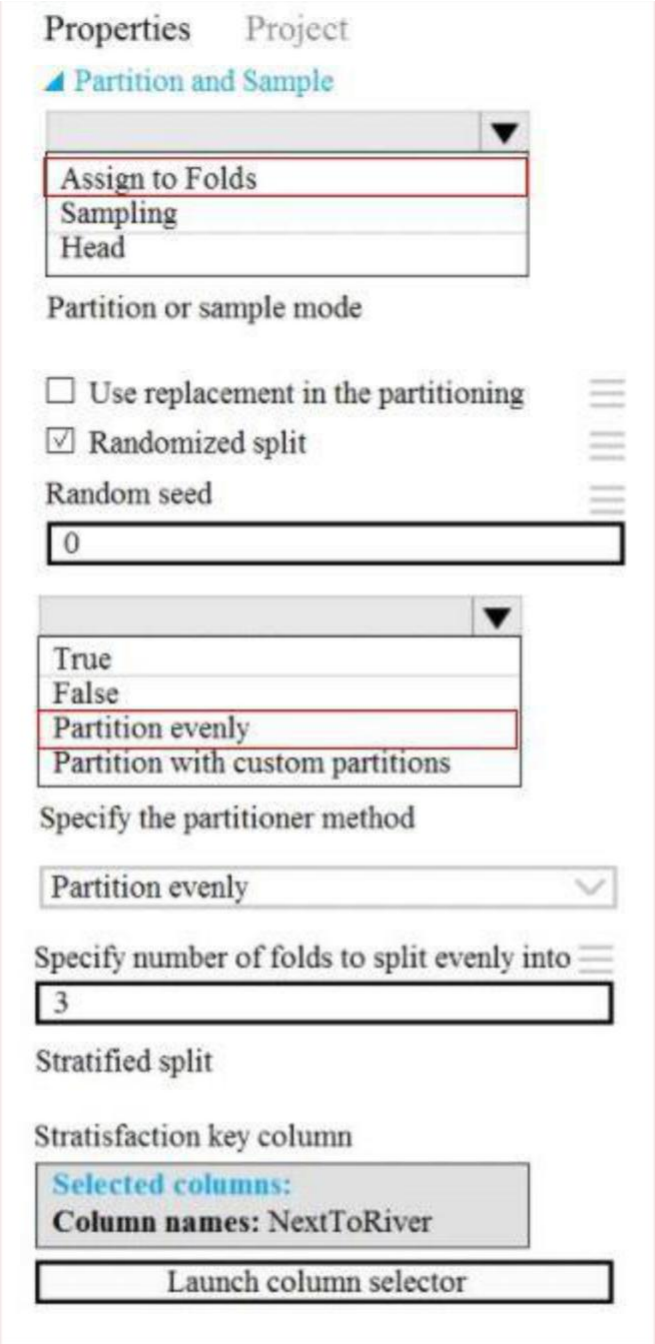
Stratified split

Stratisfaction key column

Selected columns:
Column names: NextToRiver

Launch column selector

Correct Answer:



QUESTION 3

You are building a regression model tot estimating the number of calls during an event. You need to determine whether the feature values achieve the conditions to build a Poisson regression model.
Which two conditions must the feature set contain? I ach correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. The label data must be a negative value.
B. The label data can be positive or negative,
C. The label data must be a positive value
D. The label data must be non discrete. E.
The data must be whole numbers.

Correct Answer: C,E

Explanation/Reference:

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

QUESTION 4

You are solving a classification task.
You must evaluate your model on a limited data sample by using k-fold cross validation. You start by configuring a k parameter as the number of splits.
You need to configure the k parameter for the cross-validation.
Which value should you use?

A. k=0.5
B. k=0
C. k=5
D. k=1

Correct Answer: C

Explanation/Reference:

Leave One Out (LOO) cross-validation
Setting K = n (the number of observations) yields n-fold and is called leave-one out cross-validation (LOO), a special case of the K-fold approach.
LOO CV is sometimes useful but typically doesn't shake up the data enough. The estimates from each fold are highly correlated and hence their average can have high variance.
This is why the usual choice is K=5 or 10. It provides a good compromise for the bias-variance tradeoff.

QUESTION 5

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

Correct Answer: A

Explanation/Reference:

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

QUESTION 6

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 using Azure Machine Learning to run an experiment that trains a classification model.
You want to use Hyperdrive to find parameters that optimize the AUC metric for the model. You configure a HyperDriveConfig for the experiment by running the following code:

```
hyperdrive = HyperDriveConfig(estimator=your_estimator,
 hyperparameter_sampling=your_params,
 policy=policy,
 primary_metric_name='AUC',
 primary_metric_goal=PrimaryMetricGoal.MAXIMIZE,
 max_total_runs=6,
 max_concurrent_runs=4)
```

variable named y_test variable, and the predicted probabilities from the model are stored in a variable named y_predicted. You need to add logging to the script to allow Hyperdrive to optimize hyperparameters for the AUC metric. Solution: Run the following code:

```
from sklearn.metrics import roc_auc_score
import logging
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
logging.info("AUC: " + str(auc))
```

Does the solution meet the goal?

A. Yes
B. No

Correct Answer: A

Explanation/Reference:

Python printing/logging example:
logging.info(message)
Destination: Driver logs, Azure Machine Learning designer
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines

QUESTION 7

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 train a classification model by using a logistic regression algorithm.
You must be able to explain the model's predictions by calculating the importance of each feature, both as an overall global relative importance value and as a measure of local importance for a specific set of predictions.
You need to create an explainer that you can use to retrieve the required global and local feature importance values.
Solution: Create a PFIExplainer.
Does the solution meet the goal?

A. Yes
B. No

Correct Answer: A

Explanation/Reference:

Permutation Feature Importance Explainer (PFI): Permutation Feature Importance is a technique used to explain classification and regression models. At a high level, the way it works is by randomly shuffling data one feature at a time for the entire dataset and calculating how much the performance metric of interest changes. The larger the change, the more important that feature is. PFI can explain the overall behavior of any underlying model but does not explain individual predictions.
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability

QUESTION 8

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

Correct Answer: C

Explanation/Reference:

Replace using Probabilistic PCA: Compared to other options, such as Multiple Imputation using Chained Equations (MICE), this option has the advantage of not requiring the application of predictors for each column. Instead, it approximates the covariance for the full dataset. Therefore, it might offer better performance for datasets that have missing values in many columns. References:

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

QUESTION 9

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

Correct Answer: B

Explanation/Reference:

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

QUESTION 10

You write code to retrieve an experiment that is run from your Azure Machine Learning workspace. The run used the model interpretation support in Azure Machine Learning to generate and upload a model explanation.
Business managers in your organization want to see the importance of the features in the model. You need to print out the model features and their relative importance in an output that looks similar to the following.

```
Feature          Importance
0                1.5627435610083558
2                0.6077689312583112
4                0.5574002432900718
3                0.42858759955671777
1                0.3501361539771977
```

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

```
# Assume required modules are imported

ws = Workspace.from_config()
feature_importances = explanation.[ from_run | list_model_explanations | from_run_id | download_model_explanation ]( workspace = ws,
experiment_name='train_and_explain',
run_id='train_and_explain_12345')

explanation = client.[ upload_model_explanation | list_model_explanations | run | download_model_explanation ]()

feature_importances = explanation.[ explanation | explanation_client | get_feature_important_dict | download_model_explanation ]()

for key, value in feature_importances.items():
            print(key, "\t", value)
```

Correct Answer:

```
# Assume required modules are imported

ws = Workspace.from_config()
feature_importances = explanation.[from_run | list_model_explanations | from_run_id | download_model_explanation]( workspace = ws,
experiment_name='train_and_explain',
run_id='train_and_explain_12345')

explanation = client.[upload_model_explanation | list_model_explanations | run | download_model_explanation]()

feature_importances = explanation.[explanation | explanation_client | get_feature_important_dict | download_model_explanation]()

for key, value in feature_importances.items():
            print(key, "\t", value)
```

Explanation/Reference:

Explanation:
https://docs.microsoft.com/en-us/python/api/azureml-contrib-interpret/azureml.contrib.interpret.explanation.explanation_client.explanationclient?view=azure-ml-p y

QUESTION 11

You are creating a new Azure Machine Learning pipeline using the designer.
The pipeline must train a model using data in a comma-separated values (CSV) file that is published on a
website. You have not created a dataset for this file.
You need to ingest the data from the CSV file into the designer pipeline using the minimal administrative effort.
Which module should you add to the pipeline in Designer?

A. Convert to CSV
B. Enter Data Manually
C. Import Data
D. Dataset

Correct Answer: D

QUESTION 12

You register the following versions of a model.

Model name	Model version	Tags	Properties
healthcare_model	3	'Training context':'CPU Compute'	value:87.43
healthcare_model	2	'Training context':'CPU Compute'	value:54.98
healthcare_model	1	'Training context':'CPU Compute'	value:23.56

You use the Azure ML Python SDK to run a training experiment. You use a variable named run to reference the experiment run.
After the run has been submitted and completed, you run the following code:

```
run.register_model(model_path='outputs/model.pkl',
 model_name='healthcare_model',
 tags={'Training context':'CPU Compute'} )
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.

	Yes	No
The code will cause a previous version of the saved model to be overwritten.	○	○
The version number will now be 4.	○	○
The latest version of the stored model will have a property of value: 87.43.	○	○

Correct Answer:

	Yes	No
The code will cause a previous version of the saved model to be overwritten.	○	○
The version number will now be 4.	○	○
The latest version of the stored model will have a property of value: 87.43.	○	○

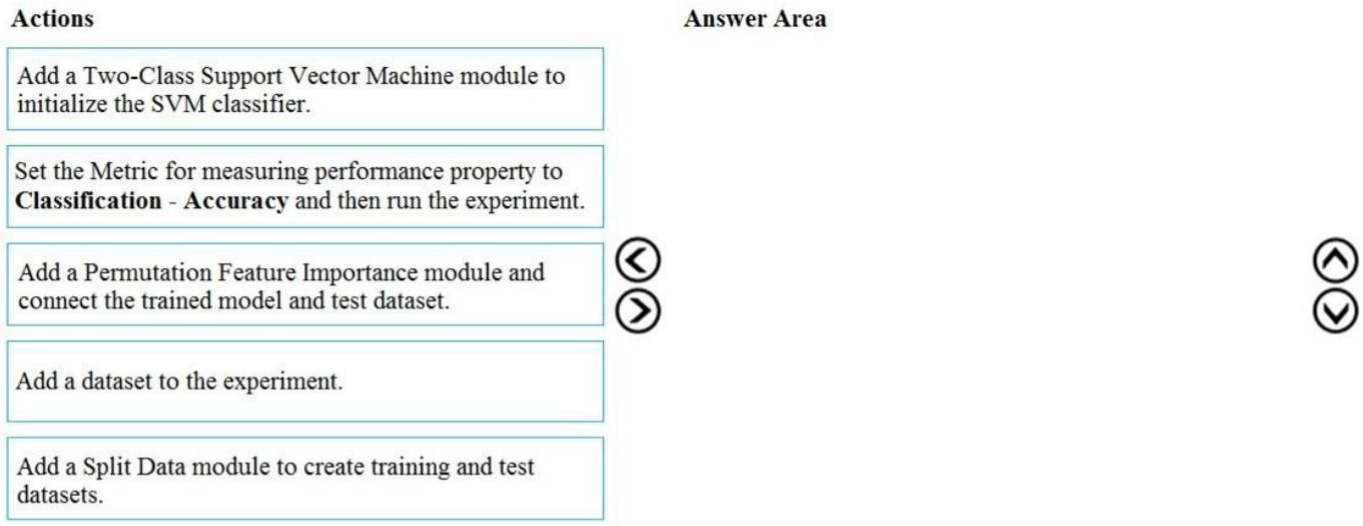
Explanation/Reference:

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

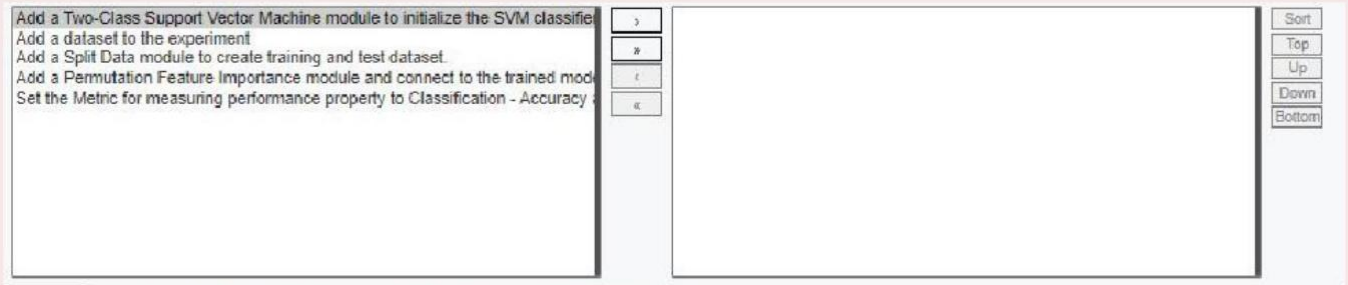
QUESTION 13

You have a dataset that contains over 150 features. You use the dataset to train a Support Vector Machine (SVM) binary classifier.
You need to use the Permutation Feature Importance module in Azure Machine Learning Studio to

compute a set of feature importance scores for the dataset.
In which order should you perform the actions? To answer, move all actions from the list of actions to the answer area and arrange them in the correct order.



Correct Answer:



Add a Two-Class Support Vector Machine module to initialize the SVM classifier.
Add a dataset to the experiment
Add a Split Data module to create training and test dataset.
Add a Permutation Feature Importance module and connect to the trained model and test dataset.
Set the Metric for measuring performance property to Classification - Accuracy and then run the experiment.

https:HYPERLINK 'https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-support-vector-machine'//docs.microsoft.com/en-us/aHYPERLINK 'https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-support-vector-machine'zure/machine-learning/studio-module-reference/two-class-support-vector-machine

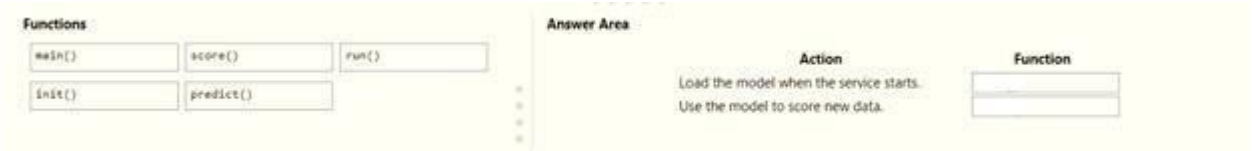
https:HYPERLINK 'https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-support-vector-machine'//docs.microsoft.com/en-us/aHYPERLINK 'https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-support-vector-machine'zure/machine-learning/studio-module-reference/two-class-support-vector-machine https://docs.microsoHYPERLINK 'https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/permutation-feature-importance'ft.com/en-us/azure/machine-learnHYPERLINK 'https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/permutation-feature-importance'ing/studio-module-reference/permutation-feature-importance

https:HYPERLINK 'https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-support-vector-machine'//docs.microsoft.com/en-us/aHYPERLINK 'https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-support-vector-machine'zure/machine-learning/studio-module-reference/two-class-support-vector-machine https://docs.microsoHYPERLINK 'https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/permutation-feature-importance'ft.com/en-us/azure/machine-learnHYPERLINK 'https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/permutation-feature-importance'ing/studio-module-reference/permutation-feature-importance

https://docs.microsoHYPERLINK 'https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/permutation-feature-importance'ft.com/en-us/azure/machine-learnHYPERLINK 'https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/permutation-feature-importance'ing/studio-module-reference/permutation-feature-importance

QUESTION 14

You use Azure Machine Learning to deploy a model as a real-time web service.
You need to create an entry script for the service that ensures that the model is loaded when the service starts and is used to score new data as it is received.
Which functions should you include in the script? To answer, drag the appropriate functions to the correct actions. Each function 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.



Correct Answer:

Explanation/Reference:

Explanation:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-existing-model

QUESTION 15

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 overcloking
E. High Random Access Memory (RAM) configuration

Correct Answer: C

Explanation/Reference:

A Deep Learning Virtual Machine is a pre-configured environment for deep learning using GPU instances.
References:
https://azuremarketplace.microsoft.com/en-au/marketplace/apps/microsoft-ads.dsvm-deep-learning

QUESTION 16

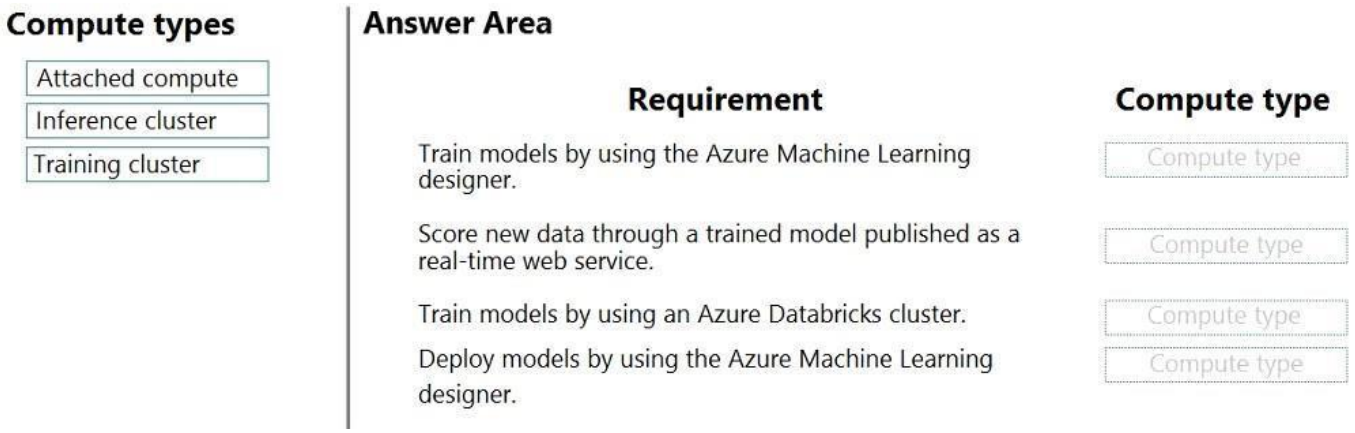
You are building a recurrent neural network to perform a binary classification. You review the training loss, validation loss, training accuracy, and validation accuracy for each training epoch. You need to analyze model performance.
Which observation indicates that the classification model is over fitted?

A. The training loss .stays constant and the validation loss stays on a constant value and close to the training loss value when training the model.
B. The training loss increases while the validation loss decreases when training the model.
C. The training loss decreases while the validation loss increases when training the model.
D. The training loss stays constant and the validation loss decreases when training the model.

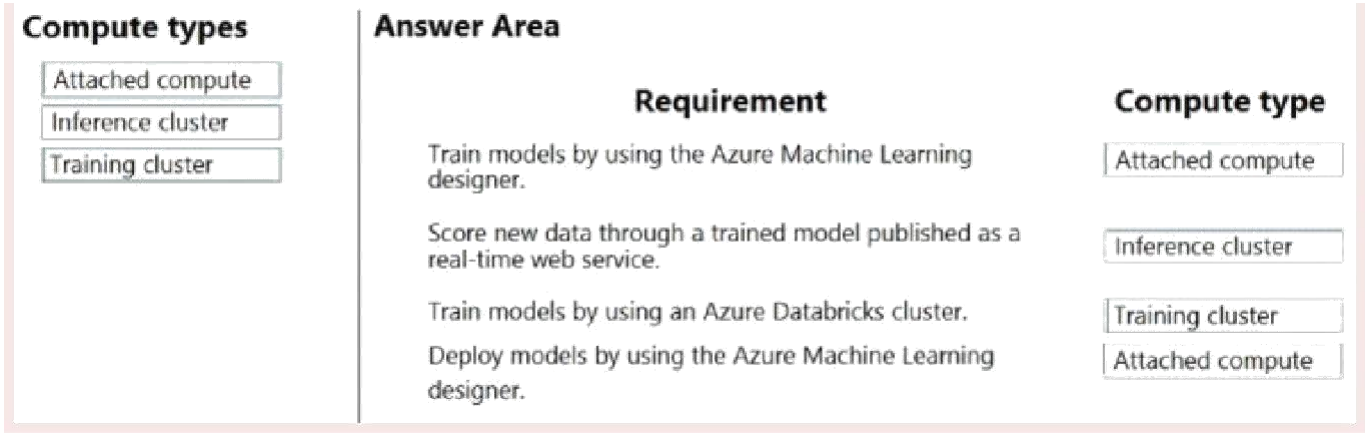
Correct Answer: B

QUESTION 17

You create machine learning models by using Azure Machine Learning.
You plan to train and score models by using a variety of compute contexts. You also plan to create a new compute resource in Azure Machine Learning studio. You need to select the appropriate compute types.
Which compute types should you select? To answer, drag the appropriate compute types to the correct requirements. Each compute type 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.



Correct Answer:



QUESTION 18

You are a data scientist working for a bank and have used Azure ML to train and register a machine learning model that predicts whether a customer is likely to repay a loan.
You want to understand how your model is making selections and must be sure that the model does not violate government regulations such as denying loans based on where an applicant lives.
You need to determine the extent to which each feature in the customer data is influencing predictions.
What should you do?

A. Enable data drift monitoring for the model and its training dataset.
B. Score the model against some test data with known label values and use the results to calculate aconfusion matrix.
C. Use the Hyperdrive library to test the model with multiple hyperparameter values.
D. Use the interpretability package to generate an explainer for the model.
E. Add tags to the model registration indicating the names of the features in the training dataset.

Correct Answer: D

Explanation/Reference:

When you compute model explanations and visualize them, you're not limited to an existing model explanation for an automated ML model. You can also get an explanation for your model with different test data. The steps in this section show you how to compute and visualize engineered feature importance based on your test data.
Incorrect Answers:
A: In the context of machine learning, data drift is the change in model input data that leads to model performance degradation. It is one of the top reasons where model accuracy degrades over time, thus monitoring data drift helps detect model performance issues.
B: A confusion matrix is used to describe the performance of a classification model. Each row displays the instances of the true, or actual class in your dataset, and each column represents the instances of the class that was predicted by the model.
C: Hyperparameters are adjustable parameters you choose for model training that guide the training process. The HyperDrive package helps you automate choosing these parameters.

https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability-automl

QUESTION 19

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You create a model to forecast weather conditions based on historical data.
You need to create a pipeline that runs a processing script to load data from a datastore and pass the processed data to a machine learning model training script. Solution: Run the following code:

```
datastore = ws.get_default_datastore()
data_input = PipelineData("raw_data", datastore=rawdatastore)
data_output = PipelineData("processed_data", datastore=datastore)
process_step = PythonScriptStep(script_name="process.py",
 arguments=["--data_for_train", data_input],
 outputs=[data_output], compute_target=aml_compute,
 source_directory=process_directory)
train_step = PythonScriptStep(script_name="train.py",
 arguments=["--data_for_train", data_input], inputs=[data_output],
 compute_target=aml_compute, source_directory=train_directory)
pipeline = Pipeline(workspace=ws, steps=[process_step, train_step])
```

Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B

Explanation/Reference:

Note: Data used in pipeline can be produced by one step and consumed in another step by providing a PipelineData object as an output of one step and an input of one or more subsequent steps.
Compare with this example, the pipeline train step depends on the process_step_output output of the pipeline process step:
from azureml.pipeline.core import Pipeline, PipelineData
from azureml.pipeline.steps import PythonScriptStep
datastore = ws.get_default_datastore()
process_step_output = PipelineData('processed_data', datastore=datastore)
process_step = PythonScriptStep(script_name='process.py', arguments=['--data_for_train', process_step_output], outputs=[process_step_output],

compute_target=aml_compute,
source_directory=process_directory)
train_step = PythonScriptStep(script_name='train.py',
arguments=['--data_for_train', process_step_output],
inputs=[process_step_output],
compute_target=aml_compute,
source_directory=train_directory)
pipeline = Pipeline(workspace=ws, steps=[process_step, train_step])
https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinedata?view=azure-ml-py

QUESTION 20

An organization uses Azure Machine Learning service and wants to expand their use of machine learning.
You have the following compute environments. The organization does not want to create another compute environment.

Environment name	Compute type
nb_server	Compute Instance
aks_cluster	Azure Kubernetes Service
mlc_cluster	Machine Learning Compute

You need to determine which compute environment to use for the following scenarios.
Which compute types should you use? To answer, drag the appropriate compute environments to the correct scenarios. Each compute environment 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.

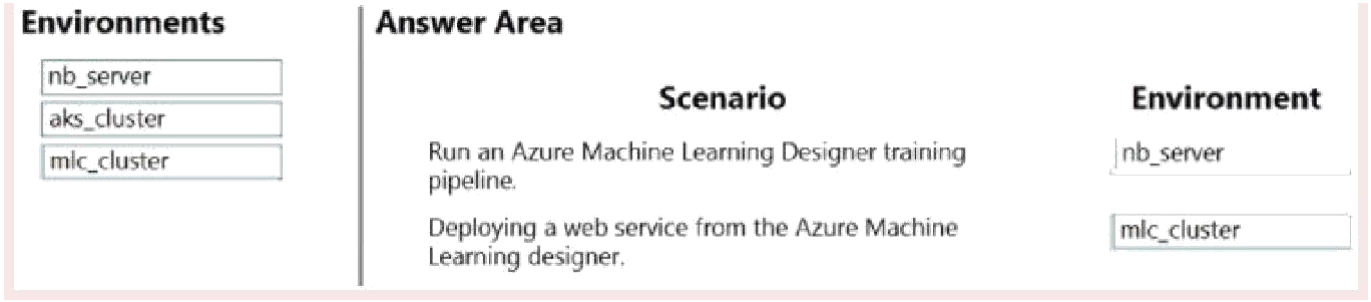
Environments

- nb_server
- aks_cluster
- mlc_cluster

Answer Area

Scenario	Environment
Run an Azure Machine Learning Designer training pipeline.	Environment
Deploying a web service from the Azure Machine Learning designer.	Environment

Correct Answer:



Explanation/Reference:

Explanation:
https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-set-up-training-targets

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 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

Correct Answer: B

Explanation/Reference:

Instead use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode.
Note: SMOTE is used to increase the number of underepresented 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.
Incorrect Answers:
The Principal Component Analysis module in Azure Machine Learning Studio (classic) is used to reduce the dimensionality of your training data. The module analyzes your data and creates a reduced feature set that captures all the information contained in the dataset, but in a smaller number of features.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/principal-

component-analysis