

Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure

<https://www.2passeasy.com/dumps/DP-100/>



NEW QUESTION 1

- (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 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_output = PipelineData("processed_data", datastore=datastore)
process_step = PythonScriptStep(script_name="process.py",
    arguments=["--data_for_train", data_output],
    outputs=[data_output], compute_target=aml_compute,
    source_directory=process_directory)
pipeline = Pipeline(workspace=ws, steps=[process_step])
```

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

train_step is missing. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinedata?view=azu>

NEW QUESTION 2

- (Exam Topic 3)

You train a machine learning model.

You must deploy the model as a real-time inference service for testing. The service requires low CPU utilization and less than 48 MB of RAM. The compute target for the deployed service must initialize automatically while minimizing cost and administrative overhead.

Which compute target should you use?

- A. Azure Kubernetes Service (AKS) inference cluster
- B. Azure Machine Learning compute cluster
- C. Azure Container Instance (ACI)
- D. attached Azure Databricks cluster

Answer: C

Explanation:

Azure Container Instances (ACI) are suitable only for small models less than 1 GB in size. Use it for low-scale CPU-based workloads that require less than 48 GB of RAM.

Note: Microsoft recommends using single-node Azure Kubernetes Service (AKS) clusters for dev-test of larger models.

Reference:

<https://docs.microsoft.com/id-id/azure/machine-learning/how-to-deploy-and-where>

NEW QUESTION 3

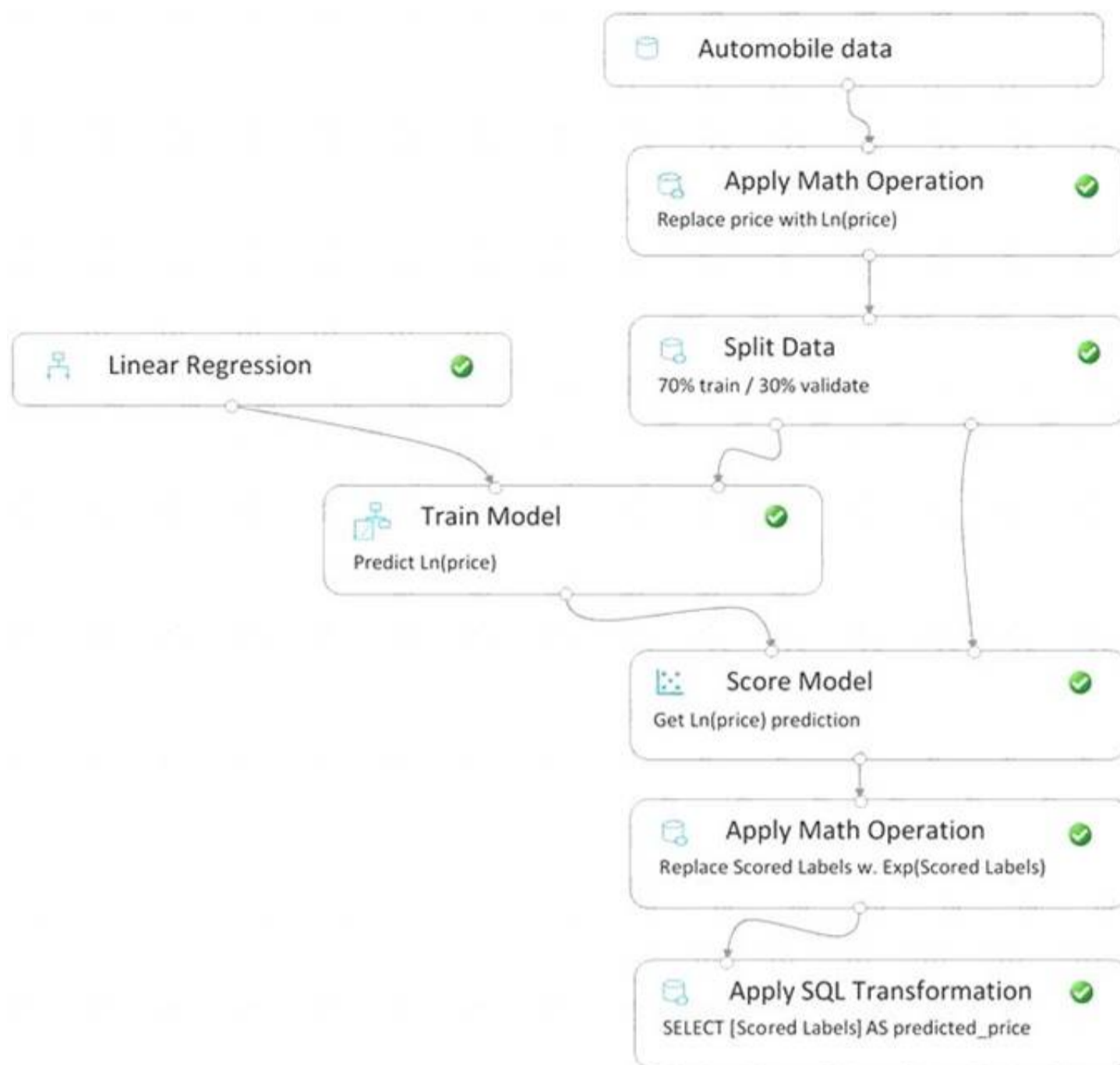
- (Exam Topic 3)

You create a pipeline in designer to train a model that predicts automobile prices.

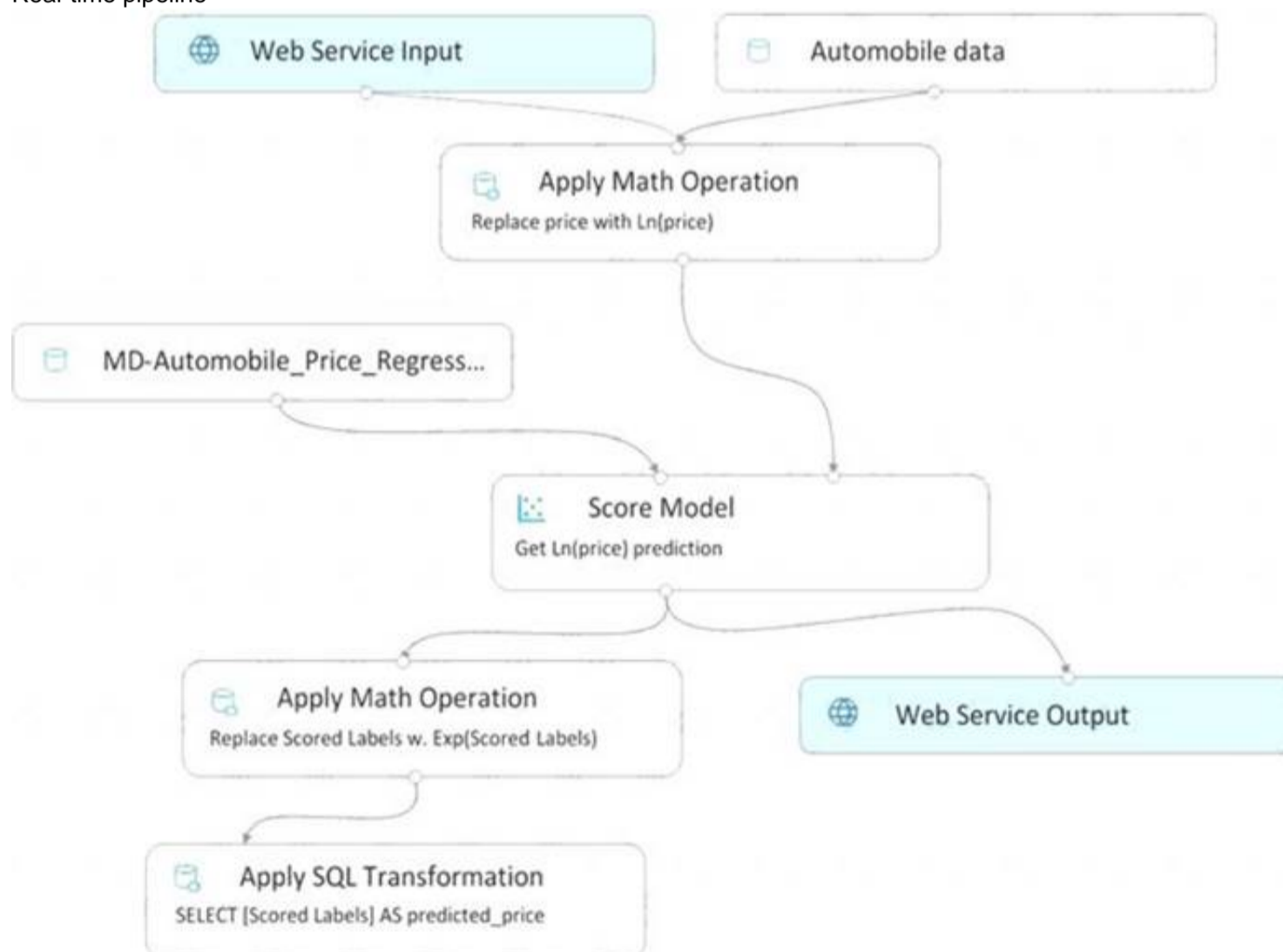
Because of non-linear relationships in the data, the pipeline calculates the natural log (Ln) of the prices in the training data, trains a model to predict this natural log of price value, and then calculates the exponential of the scored label to get the predicted price.

The training pipeline is shown in the exhibit. (Click the Training pipeline tab.)

Training pipeline



You create a real-time inference pipeline from the training pipeline, as shown in the exhibit. (Click the Real-time pipeline tab.)
 Real-time pipeline



You need to modify the inference pipeline to ensure that the web service returns the exponential of the scored label as the predicted automobile price and that client applications are not required to include a price value in the input values.

Which three modifications must you make to the inference pipeline? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Connect the output of the Apply SQL Transformation to the Web Service Output module.
- B. Replace the Web Service Input module with a data input that does not include the price column.
- C. Add a Select Columns module before the Score Model module to select all columns other than price.
- D. Replace the training dataset module with a data input that does not include the price column.
- E. Remove the Apply Math Operation module that replaces price with its natural log from the data flow.

F. Remove the Apply SQL Transformation module from the data flow.

Answer: ACE

NEW QUESTION 4

- (Exam Topic 3)

You are preparing to build a deep learning convolutional neural network model for image classification. You create a script to train the model using CUDA devices. You must submit an experiment that runs this script in the Azure Machine Learning workspace. The following compute resources are available:

- > a Microsoft Surface device on which Microsoft Office has been installed. Corporate IT policies prevent the installation of additional software
- > a Compute Instance named ds-workstation in the workspace with 2 CPUs and 8 GB of memory
- > an Azure Machine Learning compute target named cpu-cluster with eight CPU-based nodes
- > an Azure Machine Learning compute target named gpu-cluster with four CPU and GPU-based nodes

You need to specify the compute resources to be used for running the code to submit the experiment, and for running the script in order to minimize model training time.

Which resources should the data scientist use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Resource type	Option
Run code to submit the experiment	<div><div></div><div><div>the Microsoft Surface device</div><div>the ds-workstation notebook VM</div><div>the cpu-cluster compute target</div><div>the gpu-cluster compute target</div></div></div>
Run the training script	<div><div></div><div><div>the ds-workstation notebook VM</div><div>the cpu-compute target</div><div>the gpu-compute target</div><div>the Microsoft Surface device</div></div></div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Resource type	Option
Run code to submit the experiment	<div><div></div><div><div>the Microsoft Surface device</div><div>the ds-workstation notebook VM</div><div>the cpu-cluster compute target</div><div>the gpu-cluster compute target</div></div></div>
Run the training script	<div><div></div><div><div>the ds-workstation notebook VM</div><div>the cpu-compute target</div><div>the gpu-compute target</div><div>the Microsoft Surface device</div></div></div>

NEW QUESTION 5

- (Exam Topic 3)

You are running a training experiment on remote compute in Azure Machine Learning.

The experiment is configured to use a conda environment that includes the mlflow and azureml-contrib-run packages.

You must use MLflow as the logging package for tracking metrics generated in the experiment. You need to complete the script for the experiment.

How should you complete the code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

```
import numpy as np
# Import library to log metrics
```

```
from azureml.core import Run
import mlflow
import logging
```

```
# Start logging for this run
```

```
run = Run.get_context()
mlflow.start_run()
logger = logging.getLogger('Run')
reg_rate = 0.01
# Log the reg_rate metric
```

```
run.log('reg_rate', np.float(reg_rate))
mlflow.log_metric('reg_rate', np.float(reg_rate))
logger.info(np.float(reg_rate))
```

```
# Stop logging for this run
```

```
run.complete()
mlflow.end_run()
logger.setLevel(logging.INFO)
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: import mlflow

Import the mlflow and Workspace classes to access MLflow's tracking URI and configure your workspace. Box 2: mlflow.start_run()

Set the MLflow experiment name with set_experiment() and start your training run with start_run(). Box 3: mlflow.log_metric('..')

Use log_metric() to activate the MLflow logging API and begin logging your training run metrics. Box 4: mlflow.end_run()

Close the run: run.endRun() Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-mlflow>

NEW QUESTION 6

- (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 7

- (Exam Topic 3)

You are preparing to use the Azure ML SDK to run an experiment and need to create compute. You run the following code:

```
from azureml.core.compute import ComputeTarget, AmlCompute
from azureml.core.compute_target import ComputeTargetException
ws = Workspace.from_config()
cluster_name = 'aml-cluster'
try:
    training_compute = ComputeTarget(workspace=ws, name=cluster_name)
except ComputeTargetException:
    compute_config = AmlCompute.provisioning_configuration(vm_size='STANDARD_D2_V2', vm_priority='lowpriority',
max_nodes=4)
    training_compute = ComputeTarget.create(ws, cluster_name, compute_config)
    training_compute.wait_for_completion(show_output=True)
```

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
If a training cluster named aml-cluster already exists in the workspace, it will be deleted and replaced.	<input type="radio"/>	<input type="radio"/>
The wait_for_completion() method will not return until the aml-cluster compute has four active nodes.	<input type="radio"/>	<input type="radio"/>
If the code creates a new aml-cluster compute target, it may be preempted due to capacity constraints.	<input type="radio"/>	<input type="radio"/>
The aml-cluster compute target is deleted from the workspace after the training experiment completes.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: No

If a training cluster already exists it will be used. Box 2: Yes

The wait_for_completion method waits for the current provisioning operation to finish on the cluster. Box 3: Yes

Low Priority VMs use Azure's excess capacity and are thus cheaper but risk your run being pre-empted.

Box 4: No

Need to use training_compute.delete() to deprovision and delete the AmlCompute target. Reference:

<https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/training/train-on> <https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.computetarget>

NEW QUESTION 8

- (Exam Topic 3)

You use the following code to run a script as an experiment in Azure Machine Learning:

```
from azureml.core import Workspace, Experiment, Run
from azureml.core import RunConfig, ScriptRunConfig
ws = Workspace.from_config()
run_config = RunConfiguration()
run_config.target='local'
script_config = ScriptRunConfig(source_directory='./script', script='experiment.py', run_config=run_config)
experiment = Experiment(workspace=ws, name='script experiment')
run = experiment.submit(config=script_config)
run.wait_for_completion()
```

You must identify the output files that are generated by the experiment run. You need to add code to retrieve the output file names.
 Which code segment should you add to the script?

- A. files = run.get_properties()
- B. files= run.get_file_names()
- C. files = run.get_details_with_logs()
- D. files = run.get_metrics()
- E. files = run.get_details()

Answer: B

Explanation:

You can list all of the files that are associated with this run record by called run.get_file_names() Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-track-experiments>

NEW QUESTION 9

- (Exam Topic 3)

You plan to create a speech recognition deep learning model. The model must support the latest version of Python.

You need to recommend a deep learning framework for speech recognition to include in the Data Science Virtual Machine (DSVM).

What should you recommend?

- A. Apache Drill
- B. Tensorflow
- C. Rattle
- D. Weka

Answer: B

Explanation:

TensorFlow is an open source library for numerical computation and large-scale machine learning. It uses Python to provide a convenient front-end API for building applications with the framework

TensorFlow can train and run deep neural networks for handwritten digit classification, image recognition, word embeddings, recurrent neural networks, sequence-to-sequence models for machine translation, natural language processing, and PDE (partial differential equation) based simulations.

References:

<https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html>

NEW QUESTION 10

- (Exam Topic 3)

You previously deployed a model that was trained using a tabular dataset named training-dataset, which is based on a folder of CSV files.

Over time, you have collected the features and predicted labels generated by the model in a folder containing a CSV file for each month. You have created two tabular datasets based on the folder containing the inference data: one named predictions-dataset with a schema that matches the training data exactly, including the predicted label; and another named features-dataset with a schema containing all of the feature columns and a timestamp column based on the filename, which includes the day, month, and year.

You need to create a data drift monitor to identify any changing trends in the feature data since the model was trained. To accomplish this, you must define the required datasets for the data drift monitor.

Which datasets should you use to configure the data drift monitor? To answer, drag the appropriate datasets to the correct data drift monitor options. Each source 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.

Target datasets	Answer Area
training-dataset	Baseline dataset
predictions-dataset	Target dataset
features-dataset	Target dataset

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Text Description automatically generated with medium confidence

Box 1: training-dataset

Baseline dataset - usually the training dataset for a model. Box 2: predictions-dataset

Target dataset - usually model input data - is compared over time to your baseline dataset. This comparison means that your target dataset must have a timestamp column specified.

The monitor will compare the baseline and target datasets. Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-monitor-datasets>

NEW QUESTION 10

- (Exam Topic 3)

You are creating an experiment by using Azure Machine Learning Studio.

You must divide the data into four subsets for evaluation. There is a high degree of missing values in the data. You must prepare the data for analysis.

You need to select appropriate methods for producing the experiment.

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

NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.

Actions

Answer Area

Build Counting Transform

Missing Values Scrubber

Feature Hashing

Clean Missing Data

Replace Discrete Values

Import Data

Latent Dirichlet Transformation

Partition and Sample



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

The Clean Missing Data module in Azure Machine Learning Studio, to remove, replace, or infer missing values.

NEW QUESTION 13

- (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 have a Python script named train.py in a local folder named scripts. The script trains a regression model by using scikit-learn. The script includes code to load a training data file which is also located in the scripts folder.

You must run the script as an Azure ML experiment on a compute cluster named aml-compute.

You need to configure the run to ensure that the environment includes the required packages for model training. You have instantiated a variable named aml-compute that references the target compute cluster.

Solution: Run the following code:

```
from azureml.train.estimator import Estimator
sk_est = Estimator(source_directory='./scripts',
    compute_target=aml-compute,
    entry_script='train.py',
    conda_packages=['scikit-learn'])
```

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

The scikit-learn estimator provides a simple way of launching a scikit-learn training job on a compute target. It is implemented through the SKLearn class, which can be used to support single-node CPU training.

Example:

```
from azureml.train.sklearn import SKLearn
}
estimator = SKLearn(source_directory=project_folder, compute_target=compute_target, entry_script='train_iris.py'
)
```

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-scikit-learn>

NEW QUESTION 14

- (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 Microsot 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 15

- (Exam Topic 3)

You need to select a pre built development environment for a series of data science experiments. You must use the R language for the experiments. Which three environments can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. MI.NET Library on a local environment
- B. Azure Machine Learning Studio
- C. Data Science Virtual Machine (OSVM)
- D. Azure Data bricks
- E. Azure Cognitive Services

Answer: ABD

NEW QUESTION 19

- (Exam Topic 3)

You create a multi-class image classification deep learning model that uses the PyTorch deep learning framework.

You must configure Azure Machine Learning Hyperdrive to optimize the hyperparameters for the classification model.

You need to define a primary metric to determine the hyperparameter values that result in the model with the best accuracy score.

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

NOTE: Each correct selection is worth one point.

- A. Set the primary_metric_goal of the estimator used to run the bird_classifier_train.py script to maximize.
- B. Add code to the bird_classifier_train.py script to calculate the validation loss of the model and log it as a float value with the key loss.
- C. Set the primary_metric_goal of the estimator used to run the bird_classifier_train.py script to minimize.
- D. Set the primary_metric_name of the estimator used to run the bird_classifier_train.py script to accuracy.
- E. Set the primary_metric_name of the estimator used to run the bird_classifier_train.py script to loss.
- F. Add code to the bird_classifier_train.py script to calculate the validation accuracy of the model and log it as a float value with the key accuracy.

Answer: ADF

Explanation:

AD:

primary_metric_name="accuracy", primary_metric_goal=PrimaryMetricGoal.MAXIMIZE

Optimize the runs to maximize "accuracy". Make sure to log this value in your training script. Note:

primary_metric_name: The name of the primary metric to optimize. The name of the primary metric needs to exactly match the name of the metric logged by the training script.

primary_metric_goal: It can be either PrimaryMetricGoal.MAXIMIZE or PrimaryMetricGoal.MINIMIZE and determines whether the primary metric will be maximized or minimized when evaluating the runs.

F: The training script calculates the val_accuracy and logs it as "accuracy", which is used as the primary metric.

NEW QUESTION 24

- (Exam Topic 3)

You train and register a machine learning model. You create a batch inference pipeline that uses the model to generate predictions from multiple data files.

You must publish the batch inference pipeline as a service that can be scheduled to run every night. You need to select an appropriate compute target for the inference service.

Which compute target should you use?

- A. Azure Machine Learning compute instance
- B. Azure Machine Learning compute cluster
- C. Azure Kubernetes Service (AKS)-based inference cluster
- D. Azure Container Instance (ACI) compute target

Answer: B

Explanation:

Azure Machine Learning compute clusters is used for Batch inference. Run batch scoring on serverless compute. Supports normal and low-priority VMs. No support for real-time inference.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target>

NEW QUESTION 29

- (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, Relative Squared Error, and the Coefficient of Determination.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

The following metrics are reported for evaluating regression models. When you compare models, they are ranked by the metric you select for evaluation.

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

Root mean squared error (RMSE) creates a single value that summarizes the error in the model. By squaring the difference, the metric disregards the difference between over-prediction and under-prediction.

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

Relative squared error (RSE) similarly normalizes the total squared error of the predicted values by dividing by the total squared error of the actual values.

Mean Zero One Error (MZOE) indicates whether the prediction was correct or not. In other words: ZeroOneLoss(x,y) = 1 when x!=y; otherwise 0.

Coefficient of determination, often referred to as R2, represents the predictive power of the model as a value between 0 and 1. Zero means the model is random (explains nothing); 1 means there is a perfect fit. However, caution should be used in interpreting R2 values, as low values can be entirely normal and high values can be suspect.

AUC.

References:

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

NEW QUESTION 30

- (Exam Topic 3)

You create an Azure Databricks workspace and a linked Azure Machine Learning workspace. You have the following Python code segment in the Azure Machine Learning workspace:

```
import mlflow
import mlflow.azureml import azureml.mlflow import azureml.core
from azureml.core import Workspace subscription_id = 'subscription_id' resource_group = 'resource_group_name' workspace_name = 'workspace_name'
ws = Workspace.get(name=workspace_name, subscription_id=subscription_id, resource_group=resource_group)
experimentName = "/Users/{user_name}/{experiment_folder}/{experiment_name}" mlflow.set_experiment(experimentName)
uri = ws.get_mlflow_tracking_uri() mlflow.set_tracking_uri(uri)
```

Instructions: 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
A resource group and Azure Machine Learning workspace will be created.	<input type="radio"/>	<input type="radio"/>
An Azure Databricks experiment will be tracked only in the Azure Machine Learning workspace.	<input type="radio"/>	<input type="radio"/>
The epoch loss metric is set to be tracked.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
 B. Not Mastered

Answer: A

Explanation:

A screenshot of a computer Description automatically generated with medium confidence

Box 1: No

The Workspace.get method loads an existing workspace without using configuration files. ws = Workspace.get(name="myworkspace", subscription_id='<azure-subscription-id>', resource_group='myresourcegroup')

Box 2: Yes

MLflow Tracking with Azure Machine Learning lets you store the logged metrics and artifacts from your local runs into your Azure Machine Learning workspace. The get_mlflow_tracking_uri() method assigns a unique tracking URI address to the workspace, ws, and set_tracking_uri() points the MLflow tracking URI to that address.

Box 3: Yes

Note: In Deep Learning, epoch means the total dataset is passed forward and backward in a neural network once.

Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-mlflow>

NEW QUESTION 33

- (Exam Topic 3)

You plan to explore demographic data for home ownership in various cities. The data is in a CSV file with the following format:

age,city,income,home_owner 21,Chicago,50000,0 35,Seattle,120000,1 23,Seattle,65000,0 45,Seattle,130000,1 18,Chicago,48000,0

You need to run an experiment in your Azure Machine Learning workspace to explore the data and log the results. The experiment must log the following information:

- > the number of observations in the dataset
- > a box plot of income by home_owner
- > a dictionary containing the city names and the average income for each city

You need to use the appropriate logging methods of the experiment's run object to log the required information.

How should you complete the code? To answer, drag the appropriate code segments to the correct locations. Each code segment 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.

Code segments

- log
- log_list
- log_row
- log_table
- log_image

Answer Area

```
from azureml.core import Experiment, Run
import pandas as pd
import matplotlib.pyplot as plt
# Create an Azure ML experiment in workspace
experiment = Experiment(workspace = ws, name = "demo-experiment")
# Start logging data from the experiment
run = experiment.start_logging()
# load the dataset
data = pd.read_csv('research/demographics.csv')
# Log the number of observations
row_count = (len(data))
run. Segment ("observations", row_count)
# Log box plot for income by home_owner
fig = plt.figure(figsize=(9, 6))
ax = fig.gca()
data.boxplot(column = 'income', by = "home_owner", ax = ax)
ax.set_title('income by home_owner')
ax.set_ylabel('income')
run. Segment (name = 'income_by_home_owner', plot = fig)
# Create a dataframe of mean income per city
mean_inc_df = data.groupby('city')['income'].agg(np.mean).to_frame().reset_index()
# Convert to a dictionary
mean_inc_dict = mean_inc_df.to_dict('dict')
# Log city names and average income dictionary
run. Segment (name="mean_income_by_city", value= mean_inc_dict)
# Complete tracking and get link to details
run.complete()
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: log

The number of observations in the dataset. `run.log(name, value, description=)`

Scalar values: Log a numerical or string value to the run with the given name. Logging a metric to a run causes that metric to be stored in the run record in the experiment. You can log the same metric multiple times within a run, the result being considered a vector of that metric.

Example: `run.log("accuracy", 0.95)`

Box 2: log_image

A box plot of income by home_owner.

log_image Log an image to the run record. Use log_image to log a .PNG image file or a matplotlib plot to the run. These images will be visible and comparable in the run record.

Example: `run.log_image("ROC", plot=plt)` Box 3: log_table

A dictionary containing the city names and the average income for each city. log_table: Log a dictionary object to the run with the given name.

NEW QUESTION 36

- (Exam Topic 3)

You have a Python script that executes a pipeline. The script includes the following code:

```
from azureml.core import Experiment
```

```
pipeline_run = Experiment(ws, 'pipeline_test').submit(pipeline)
```

 You want to test the pipeline before deploying the script.

You need to display the pipeline run details written to the STDOUT output when the pipeline completes. Which code segment should you add to the test script?

- A. `pipeline_run.get.metrics()`
- B. `pipeline_run.wait_for_completion(show_output=True)`
- C. `pipeline_param = PipelineParameter(name="stdout", default_value="console")`
- D. `pipeline_run.get_status()`

Answer: B

Explanation:

wait_for_completion: Wait for the completion of this run. Returns the status object after the wait. Syntax: `wait_for_completion(show_output=False, wait_post_processing=False, raise_on_error=True)` Parameter: show_output

Indicates whether to show the run output on sys.stdout.

NEW QUESTION 40

- (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 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

Answer: A

Explanation:

Python printing/logging example: `logging.info(message)`

Destination: Driver logs, Azure Machine Learning designer

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines>

NEW QUESTION 45

- (Exam Topic 3)

You use the following code to define the steps for a pipeline: `from azureml.core import Workspace, Experiment, Run` from `azureml.pipeline.core` import Pipeline
from `azureml.pipeline.steps` import PythonScriptStep `ws = Workspace.from_config()`

...
 step1 = PythonScriptStep(name="step1", ...) step2 = PythonScriptsStep(name="step2", ...) pipeline_steps = [step1, step2]
 You need to add code to run the steps.
 Which two code segments can you use to achieve this goal? Each correct answer presents a complete solution.
 NOTE: Each correct selection is worth one point.

- A. experiment = Experiment(workspace=ws, name='pipeline-experiment')run = experiment.submit(config=pipeline_steps)
- B. run = Run(pipeline_steps)
- C. pipeline = Pipeline(workspace=ws, steps=pipeline_steps) experiment = Experiment(workspace=ws, name='pipeline-experiment')run = experiment.submit(pipeline)
- D. pipeline = Pipeline(workspace=ws, steps=pipeline_steps)run = pipeline.submit(experiment_name='pipeline-experiment')

Answer: CD

Explanation:

After you define your steps, you build the pipeline by using some or all of those steps.
 # Build the pipeline. Example:
 pipeline1 = Pipeline(workspace=ws, steps=[compare_models])
 # Submit the pipeline to be run
 pipeline_run1 = Experiment(ws, 'Compare_Models_Exp').submit(pipeline1) Reference:
<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-machine-learning-pipelines>

NEW QUESTION 46

- (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.
 An IT department creates the following Azure resource groups and resources:

Resource group	Resources
ml_resources	<ul style="list-style-type: none"> an Azure Machine Learning workspace named amlworkspace an Azure Storage account named amlworkspace12345 an Application Insights instance named amlworkspace54321 an Azure Key Vault named amlworkspace67890 an Azure Container Registry named amlworkspace09876
general_compute	A virtual machine named mlvm with the following configuration: <ul style="list-style-type: none"> Operating system: Ubuntu Linux Software installed: Python 3.6 and Jupyter Notebooks Size: NC6 (6 vCPUs, 1 vGPU, 56 Gb RAM)

The IT department creates an Azure Kubernetes Service (AKS)-based inference compute target named aks-cluster in the Azure Machine Learning workspace.
 You have a Microsoft Surface Book computer with a GPU. Python 3.6 and Visual Studio Code are installed. You need to run a script that trains a deep neural network (DNN) model and logs the loss and accuracy metrics.
 Solution: Attach the mlvm virtual machine as a compute target in the Azure Machine Learning workspace. Install the Azure ML SDK on the Surface Book and run Python code to connect to the workspace. Run the training script as an experiment on the mlvm remote compute resource.

- A. Yes
- B. No

Answer: A

Explanation:

Use the VM as a compute target.
 Note: A compute target is a designated compute resource/environment where you run your training script or host your service deployment. This location may be your local machine or a cloud-based compute resource.
 Reference:
<https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target>

NEW QUESTION 49

- (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 the 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 Scale and Reduce sampling mode.
 Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

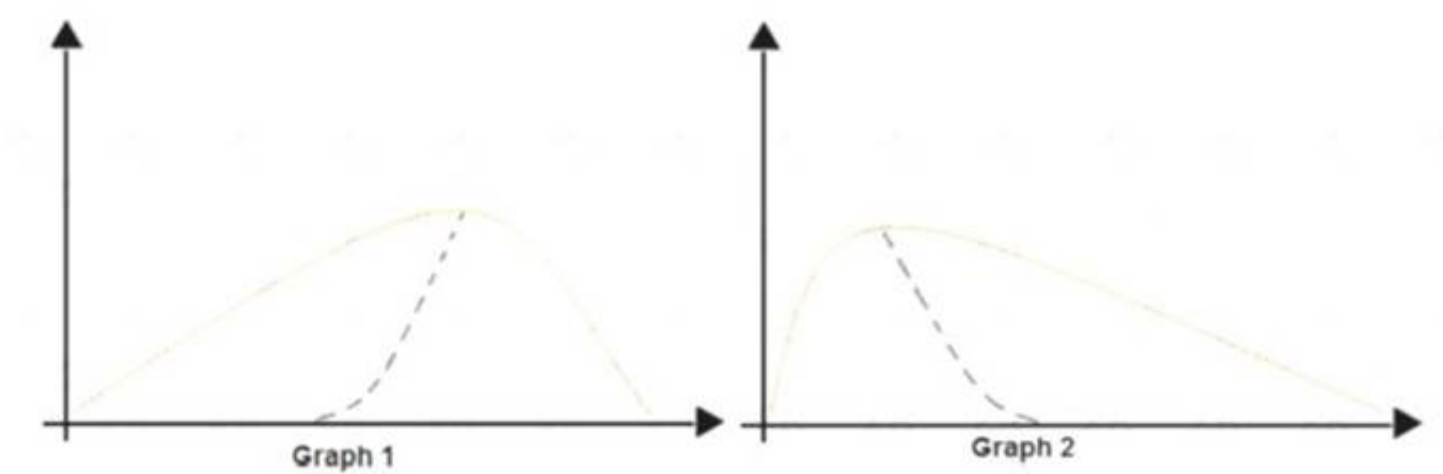
Instead use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode.
 Note: SMOTE is used to increase the number of underrepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.
 References:
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote>

NEW QUESTION 50

- (Exam Topic 3)

You are analyzing the asymmetry in a statistical distribution.

The following image contains two density curves that show the probability distribution of two datasets.



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 type of distribution is shown for the dataset density curve of Graph 1?	<div><div></div><div>Negative skew</div><div>Positive skew</div><div>Normal distribution</div><div>Bimodal distribution</div></div>
Which type of distribution is shown for the dataset density curve of Graph 2?	<div><div></div><div>Negative skew</div><div>Positive skew</div><div>Normal distribution</div><div>Bimodal distribution</div></div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Positive skew

Positive skew values means the distribution is skewed to the right. Box 2: Negative skew

Negative skewness values mean the distribution is skewed to the left. References:

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

NEW QUESTION 53

- (Exam Topic 3)

You arc 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
- B. Replace using Probabilistic PAC
- C. Replace using MICE
- D. Normalization

Answer: B

NEW QUESTION 55

- (Exam Topic 3)

You are performing sentiment analysis using a CSV file that includes 12,000 customer reviews written in a short sentence format. You add the CSV file to Azure Machine Learning Studio and configure it as the starting point dataset of an experiment. You add the Extract N-Gram Features from Text module to the experiment to extract key phrases from the customer review column in the dataset.

You must create a new n-gram dictionary from the customer review text and set the maximum n-gram size to trigrams.

What should you select? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Properties

Project

Extract N-Gram Features from Text

Text column

Selected columns

Column type: String Feature

Launch column selector

Vocabulary mode

▼

Create

ReadOnly

Update

Merge

N-Grams size

▼

3

4

4,000

12,000

0

Weighting function

▼

Minimum word length

3

Maximum word length

25

Minimum n-gram document absolu...

5

Maximum n-gram document ratio

1

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Vocabulary mode: Create
For Vocabulary mode, select Create to indicate that you are creating a new list of n-gram features. N-Grams size: 3
For N-Grams size, type a number that indicates the maximum size of the n-grams to extract and store. For example, if you type 3, unigrams, bigrams, and trigrams will be created.
Weighting function: Leave blank
The option, Weighting function, is required only if you merge or update vocabularies. It specifies how terms in the two vocabularies and their scores should be weighted against each other.
References:
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/extract-n-gram-features-from>

NEW QUESTION 58

- (Exam Topic 3)
You train and register a model in your Azure Machine Learning workspace.
You must publish a pipeline that enables client applications to use the model for batch inferencing. You must use a pipeline with a single ParallelRunStep step that runs a Python inferencing script to get predictions from the input data.
You need to create the inferencing script for the ParallelRunStep pipeline step.
Which two functions should you include? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

- A. run(mini_batch) D
- B. main()

- C. batch()
- D. init()
- E. score(mini_batch)

Answer: AD

Explanation:

Reference:

<https://github.com/Azure/MachineLearningNotebooks/tree/master/how-to-use-azureml/machine-learningpipelin>

NEW QUESTION 60

- (Exam Topic 3)

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=1
- B. k=10
- C. k=0.5
- D. k=0.9

Answer: B

Explanation:

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.

NEW QUESTION 62

- (Exam Topic 3)

You plan to use the Hyperdrive feature of Azure Machine Learning to determine the optimal hyperparameter values when training a model.

You must use Hyperdrive to try combinations of the following hyperparameter values:

- learning_rate: any value between 0.001 and 0.1
- batch_size: 16, 32, or 64

You need to configure the search space for the Hyperdrive experiment.

Which two parameter expressions should you use? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. a choice expression for learning_rate
- B. a uniform expression for learning_rate
- C. a normal expression for batch_size
- D. a choice expression for batch_size
- E. a uniform expression for batch_size

Answer: BD

Explanation:

B: Continuous hyperparameters are specified as a distribution over a continuous range of values. Supported distributions include:

➤ uniform(low, high) - Returns a value uniformly distributed between low and high

D: Discrete hyperparameters are specified as a choice among discrete values. choice can be:

➤ one or more comma-separated values

➤ a range object

➤ any arbitrary list object Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters>

NEW QUESTION 66

- (Exam Topic 3)

You are building an experiment using the Azure Machine Learning designer.

You split a dataset into training and testing sets. You select the Two-Class Boosted Decision Tree as the algorithm.

You need to determine the Area Under the Curve (AUC) of the model.

Which three modules should you use 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

Export Data

Tune Model Hyperparameters

Cross Validate Model

Evaluate Model

Score Model

Train Model

Answer Area

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Train Model

Two-Class Boosted Decision Tree

First, set up the boosted decision tree model.

* 1. Find the Two-Class Boosted Decision Tree module in the module palette and drag it onto the canvas.

* 2. Find the Train Model module, drag it onto the canvas, and then connect the output of the Two-Class Boosted Decision Tree module to the left input port of the Train Model module.

The Two-Class Boosted Decision Tree module initializes the generic model, and Train Model uses training data to train the model.

* 3. Connect the left output of the left Execute R Script module to the right input port of the Train Model

module (in this tutorial you used the data coming from the left side of the Split Data module for training). This portion of the experiment now looks something like this:



Step 2: Score Model

Score and evaluate the models

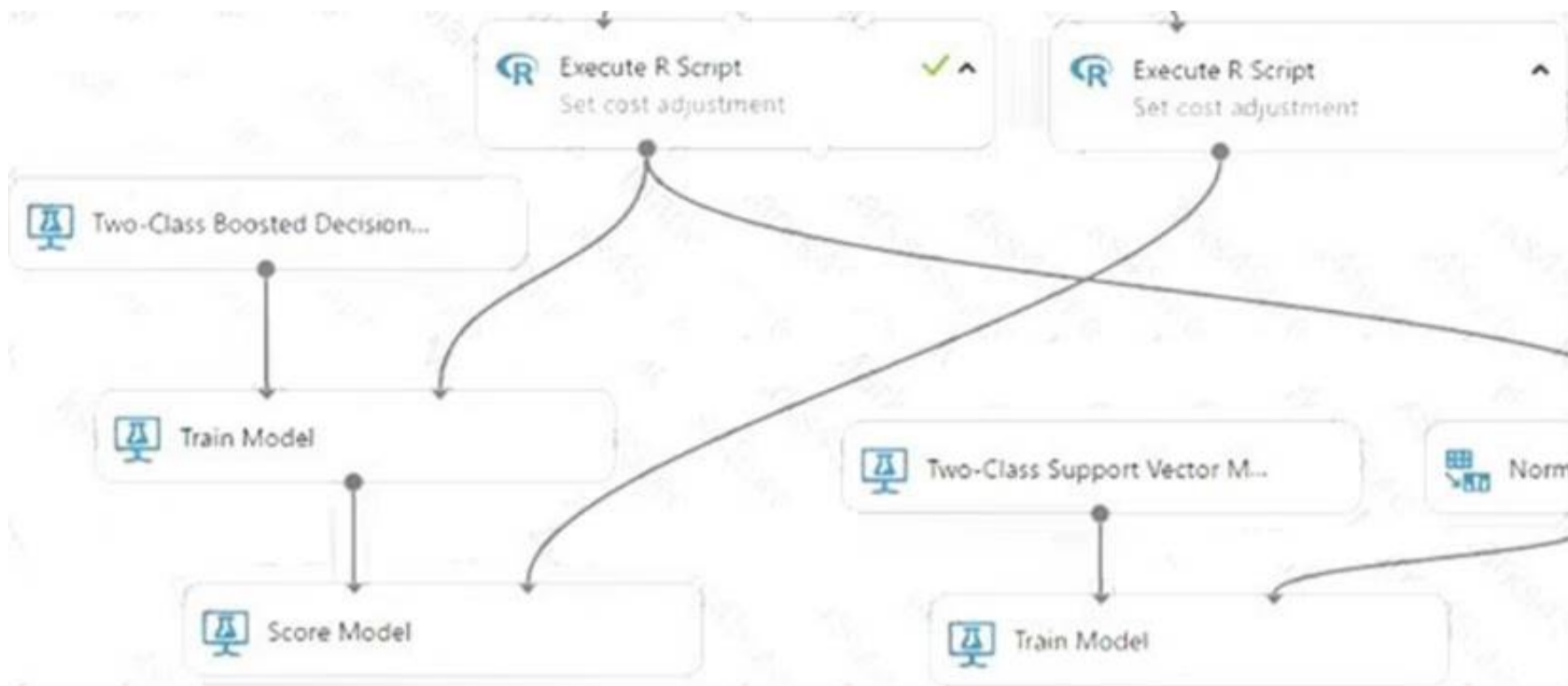
You use the testing data that was separated out by the Split Data module to score our trained models. You can then compare the results of the two models to see which generated better results.

Add the Score Model modules

* 1. Find the Score Model module and drag it onto the canvas.

* 2. Connect the Train Model module that's connected to the Two-Class Boosted Decision Tree module to the left input port of the Score Model module.

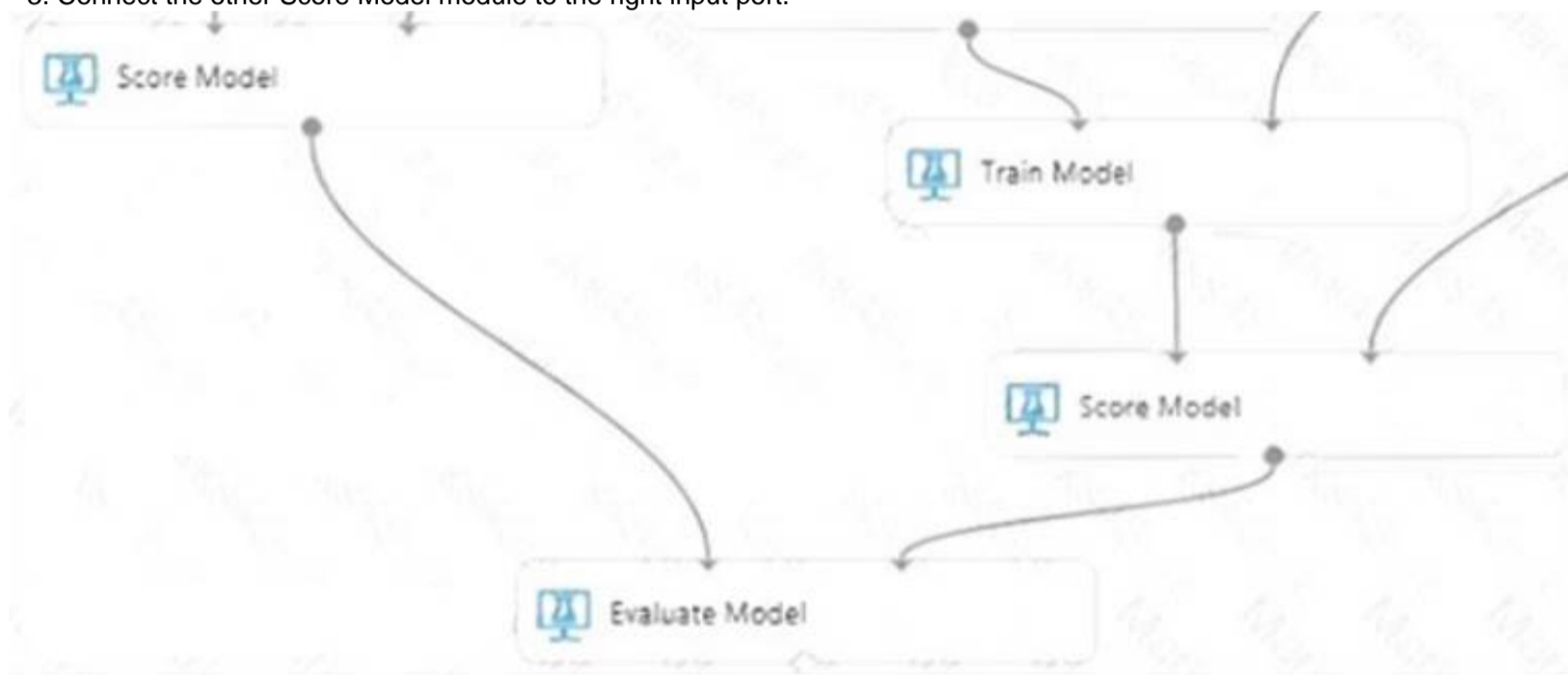
* 3. Connect the right Execute R Script module (our testing data) to the right input port of the Score Model module.



Step 3: Evaluate Model

To evaluate the two scoring results and compare them, you use an Evaluate Model module.

- * 1. Find the Evaluate Model module and drag it onto the canvas.
- * 2. Connect the output port of the Score Model module associated with the boosted decision tree model to the left input port of the Evaluate Model module.
- * 3. Connect the other Score Model module to the right input port.



NEW QUESTION 70

- (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 71

- (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 75

- (Exam Topic 3)

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:

References:

<https://machinelearningmastery.com/how-to-reduce-overfitting-in-deep-learning-with-weight-regularization/>

https://en.wikipedia.org/wiki/Convolutional_neural_network

NEW QUESTION 80

- (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 84

- (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 85

- (Exam Topic 3)

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 a confusion 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.

Answer: D

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.

Reference:

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

NEW QUESTION 87

- (Exam Topic 3)

You have a dataset that contains 2,000 rows. You are building a machine learning classification model by using Azure Learning Studio. You add a Partition and Sample module to the experiment.

You need to configure the module. You must meet the following requirements:

- > Divide the data into subsets
- > Assign the rows into folds using a round-robin method
- > Allow rows in the dataset to be reused

How should you configure the module? To answer, select the appropriate options in the dialog box 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

☐ Use replacement in the partitioning

☐ Randomized split

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Use the Split data into partitions option when you want to divide the dataset into subsets of the data. This option is also useful when you want to create a custom number of folds for cross-validation, or to split rows into several groups.

- > For Partition or sample mode, select Assign to Folds.
- > Use replacement in the partitioning: Select this option if you want the sampled row to be put back into the pool of rows for potential reuse. As a result, the same row might be assigned to several folds.
- > If you do not use replacement (the default option), the sampled row is not put back into the pool of rows for potential reuse. As a result, each row can be assigned to only one fold.
- > Randomized split: Select this option if you want rows to be randomly assigned to folds. If you do not select this option, rows are assigned to folds using the round-robin method. References:
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample>

NEW QUESTION 92

- (Exam Topic 3)

You create an Azure Machine Learning workspace and set up a development environment. You plan to train a deep neural network (DNN) by using the Tensorflow framework and by using estimators to submit training scripts.

You must optimize computation speed for training runs.

You need to choose the appropriate estimator to use as well as the appropriate training compute target configuration.

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

Answer Area

Parameter	Value
Estimator	<div>Estimator</div> <div>SkLearn</div> <div>PyTorch</div> <div>Tensorflow</div> <div>Chainer</div>
Training compute	<div>12 vCPU, 48 GB memory, 96 GB SSD</div> <div>12 vCPU, 112 GB memory, 680 GB SSD, 2 GPU, 24 GB GPU memory</div> <div>16 vCPU, 128 GB memory, 160 GB HDD, 80 GB NVME disk (4000 MBps)</div> <div>44 vCPU, 352 GB memory, 3.4 GHz CPU frequency all cores</div>

Use the selections for training compute

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Tensorflow

TensorFlow represents an estimator for training in TensorFlow experiments. Box 2: 12 vCPU, 112 GB memory...,2 GPU,...

Use GPUs for the deep neural network. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.dnn>

NEW QUESTION 94

- (Exam Topic 3)

You create an Azure Machine Learning workspace named workspace1. You assign a custom role to a user of workspace1. The custom role has the following JSON definition:

```
{
  "Name": "MyRole",
  "IsCustom": true,
  "Description": "New custom role description.",
  "Actions": ["*"],
  "NotActions": [
    "Microsoft.MachineLearningServices/workspaces/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "AssignableScopes": [
    "/subscriptions/<subscription_id>/resourceGroups/resourcegroup1/providers/
    Microsoft.MachineLearningServices/workspaces/workspace1"
  ]
}
```

Instructions: For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

Statements	Yes	No
The user can perform all actions in the workspace	<input type="radio"/>	<input type="radio"/>
The user can delete a compute resource in the workspace	<input type="radio"/>	<input type="radio"/>
The user can write metrics to the workspace	<input type="radio"/>	<input type="radio"/>

A.

Answer:

Explanation:

Graphical user interface, text, application, email Description automatically generated

Box 1: No

The actions listed in NotActions are prohibited.

If the roles include Actions that have a wildcard (*), the effective permissions are computed by subtracting the NotActions from the allowed Actions.

Box 2: No

Deleting compute resources in the workspace is in the NotActions list. Box 3: Yes

Writing metrics is not listed in NotActions. Reference:

<https://docs.microsoft.com/en-us/azure/role-based-access-control/overview#how-azure-rbac-determines-if-a-use>

NEW QUESTION 98

- (Exam Topic 3)

You use Azure Machine Learning designer to create a real-time service endpoint. You have a single Azure Machine Learning service compute resource. You train the model and prepare the real-time pipeline for deployment You need to publish the inference pipeline as a web service. Which compute type should you use?

- A. HDInsight
- B. Azure Databricks
- C. Azure Kubernetes Services
- D. the existing Machine Learning Compute resource
- E. a new Machine Learning Compute resource

Answer: C

Explanation:

Azure Kubernetes Service (AKS) can be used real-time inference. Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target>

NEW QUESTION 100

- (Exam Topic 3)

You have a comma-separated values (CSV) file containing data from which you want to train a classification model.

You are using the Automated Machine Learning interface in Azure Machine Learning studio to train the classification model. You set the task type to Classification.

You need to ensure that the Automated Machine Learning process evaluates only linear models. What should you do?

- A. Add all algorithms other than linear ones to the blocked algorithms list.
- B. Set the Exit criterion option to a metric score threshold.
- C. Clear the option to perform automatic featurization.
- D. Clear the option to enable deep learning.
- E. Set the task type to Regression.

Answer: C

Explanation:

Automatic featurization can fit non-linear models. Reference: <https://econml.azurewebsites.net/spec/estimation/dml.html>
<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-automated-ml-for-ml-models>

NEW QUESTION 104

- (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 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 PFIE explainer. Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

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.

Reference:

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

NEW QUESTION 105

- (Exam Topic 3)

You are performing a filter based feature selection for a dataset 10 build a multi class classifiers by using Azure Machine Learning Studio.

The dataset contains categorical features that are highly correlated to the output label column.

You need to select the appropriate feature scoring statistical method to identify the key predictors. Which method should you use?

- A. Chi-squared
- B. Spearman correlation
- C. Kendall correlation
- D. Person correlation

Answer: D

Explanation:

Pearson's correlation statistic, or Pearson's correlation coefficient, is also known in statistical models as the r value. For any two variables, it returns a value that indicates the strength of the correlation

Pearson's correlation coefficient is the test statistics that measures the statistical relationship, or association, between two continuous variables. It is known as the best method of measuring the association between variables of interest because it is based on the method of covariance. It gives information about the magnitude of the association, or correlation, as well as the direction of the relationship.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/filter-based-feature-selection> <https://www.statisticssolutions.com/pearsons-correlation-coefficient/>

NEW QUESTION 108

- (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 plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:

```
from azureml.core import Run
import pandas as pd
run = Run.get_context()
data = pd.read_csv('data.csv')
label_vals = data['label'].unique()
# Add code to record metrics here
run.complete()
```

The experiment must record the unique labels in the data as metrics for the run that can be reviewed later.

You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.

Solution: Replace the comment with the following code: `run.log_table('Label Values', label_vals)`

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Instead use the run_log function to log the contents in label_vals: for label_val in label_vals:

run.log('Label Values', label_val) Reference:

<https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai>

NEW QUESTION 110

- (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 create an Azure Machine Learning service datastore in a workspace. The datastore contains the following files:

- /data/2018/Q1 .csv
- /data/2018/Q2.csv
- /data/2018/Q3.csv
- /data/2018/Q4.csv
- /data/2019/Q1.csv

All files store data in the following format:

- id,f1,f2,l
- 1,1,2,0
- 2,1,1,1
- 3.2.1.0

You run the following code:

```
data_store = Datastore.register_azure_blob_container(workspace=ws,  
datastore_name='data_store',  
container_name='quarterly_data',  
account_name='companydata',  
account_key='NRPxk8duxbM3...'  
create_if_not_exists=False)
```

You need to create a dataset named training_data and load the data from all files into a single data frame by using the following code:

```
data_frame = training_data.to_pandas_dataframe()
```

Solution: Run the following code:

```
from azureml.core import Dataset  
paths = [(data_store,'data/2018/*.csv'),(data_store,'data/2019/*.csv')]  
training_data = Dataset.Tabular.from_delimited_files(paths)
```

Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

Use two file paths.

Use Dataset.Tabular_from_delimited as the data isn't cleansed. Note:

A TabularDataset represents data in a tabular format by parsing the provided file or list of files. This provides you with the ability to materialize the data into a pandas or Spark DataFrame so you can work with familiar data preparation and training libraries without having to leave your notebook. You can create a TabularDataset object from .csv, .tsv, .parquet, .jsonl files, and from SQL query results.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-register-datasets>

NEW QUESTION 115

- (Exam Topic 3)

You are using an Azure Machine Learning workspace. You set up an environment for model testing and an environment for production.

The compute target for testing must minimize cost and deployment efforts. The compute target for production must provide fast response time, autoscaling of the deployed service, and support real-time inferencing.

You need to configure compute targets for model testing and production.

Which compute targets should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Environment

Compute target

Testing

	▼
Local web service	
Azure Kubernetes Services (AKS)	
Azure Container Instances	
Azure Machine Learning compute clusters	

Production

	▼
Local web service	
Azure Kubernetes Services (AKS)	
Azure Container Instances	
Azure Machine Learning compute clusters	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Text, application Description automatically generated

Box 1: Local web service

The Local web service compute target is used for testing/debugging. Use it for limited testing and troubleshooting. Hardware acceleration depends on use of libraries in the local system.

Box 2: Azure Kubernetes Service (AKS)

Azure Kubernetes Service (AKS) is used for Real-time inference. Recommended for production workloads.

Use it for high-scale production deployments. Provides fast response time and autoscaling of the deployed service

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target>

NEW QUESTION 120

- (Exam Topic 3)

You create a script that trains a convolutional neural network model over multiple epochs and logs the validation loss after each epoch. The script includes arguments for batch size and learning rate.

You identify a set of batch size and learning rate values that you want to try.

You need to use Azure Machine Learning to find the combination of batch size and learning rate that results in the model with the lowest validation loss.

What should you do?

- A. Run the script in an experiment based on an AutoMLConfig object
- B. Create a PythonScriptStep object for the script and run it in a pipeline
- C. Use the Automated Machine Learning interface in Azure Machine Learning studio
- D. Run the script in an experiment based on a ScriptRunConfig object
- E. Run the script in an experiment based on a HyperDriveConfig object

Answer: E

Explanation:

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters>

NEW QUESTION 122

- (Exam Topic 3)

You use Azure Machine Learning to train a model based on a dataset named dataset1. You define a dataset monitor and create a dataset named dataset2 that contains new data.

You need to compare dataset1 and dataset2 by using the Azure Machine Learning SDK for Python. Which method of the DataDriftDetector class should you use?

- A. run
- B. get
- C. backfill
- D. update

Answer: C

Explanation:

A backfill run is used to see how data changes over time. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-datadrift/azureml.datadrift.datadriftdetector.datadriftdetect>

NEW QUESTION 123

- (Exam Topic 3)

You create the following config.json file.

```
{
  "workspace_name" : "ml-workspace"
}
```

You must use the Azure Machine Learning SDK to interact with data and experiments in the workspace. You need to configure the config.json file to connect to the workspace from the Python environment. Which two additional parameters must you add to the config.json file in order to connect to the workspace?
Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

- A. subscription_id
- B. Key
- C. resource_group
- D. region
- E. Login

Answer: AC

Explanation:

To use the same workspace in multiple environments, create a JSON configuration file. The configuration file saves your subscription (subscription_id), resource (resource_group), and workspace name so that it can be easily loaded.
The following sample shows how to create a workspace. from azureml.core import Workspace
ws = Workspace.create(name='myworkspace', subscription_id='<azure-subscription-id>', resource_group='myresourcegroup', create_resource_group=True, location='eastus2')
Reference:
<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace>

NEW QUESTION 125

- (Exam Topic 3)

You register a model that you plan to use in a batch inference pipeline.
The batch inference pipeline must use a ParallelRunStep step to process files in a file dataset. The script has the ParallelRunStep step runs must process six input files each time the inferencing function is called.
You need to configure the pipeline.
Which configuration setting should you specify in the ParallelRunConfig object for the PrallelRunStep step?

- A. process_count_per_node= "6"
- B. node_count= "6"
- C. mini_batch_size= "6"
- D. error_threshold= "6"

Answer: B

Explanation:

node_count is the number of nodes in the compute target used for running the ParallelRunStep. Reference:
<https://docs.microsoft.com/en-us/python/api/azureml-contrib-pipeline-steps/azureml.contrib.pipeline.steps.parall>

NEW QUESTION 126

- (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

Score Matchbox Recommender

Apply Transformation

Evaluate Recommender

Evaluate Model

Train Model

Sweep Clustering

Score Model

Load Trained Model

Answer Area

<

>

^

v

- 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 129

- (Exam Topic 2)

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

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

Answer: C

Explanation:

In statistics, the Kendall rank correlation coefficient, commonly referred to as Kendall's tau coefficient (after the Greek letter τ), is a statistic used to measure the ordinal association between two measured quantities.

It is a supported method of the Azure Machine Learning Feature selection.

Scenario: When you train a Linear Regression module using a property dataset that shows data for property prices for a large city, you need to determine the best features to use in a model. You can choose standard metrics provided to measure performance before and after the feature importance process completes. You must ensure that the distribution of the features across multiple training models is consistent.

References:





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

NEW QUESTION 130

- (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 134

- (Exam Topic 2)

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

- A. Spearman correlation
- B. Mutual information
- C. Mann-Whitney test
- D. Pearson's correlation

Answer: A

Explanation:

Spearman's rank correlation coefficient assesses how well the relationship between two variables can be described using a monotonic function.

Note: Both Spearman's and Kendall's can be formulated as special cases of a more general correlation coefficient, and they are both appropriate in this scenario.

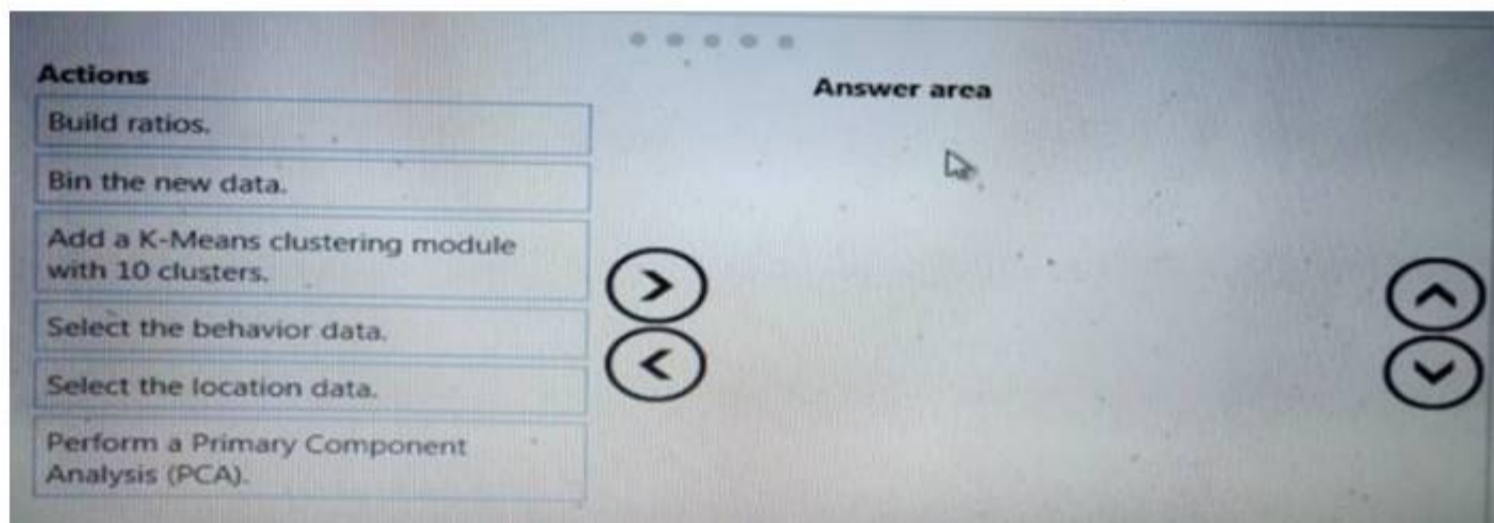
Scenario: The MedianValue and AvgRoomsInHouse columns both hold data in numeric format. You need to select a feature selection algorithm to analyze the relationship between the two columns in more detail. References:

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

NEW QUESTION 138

- (Exam Topic 1)

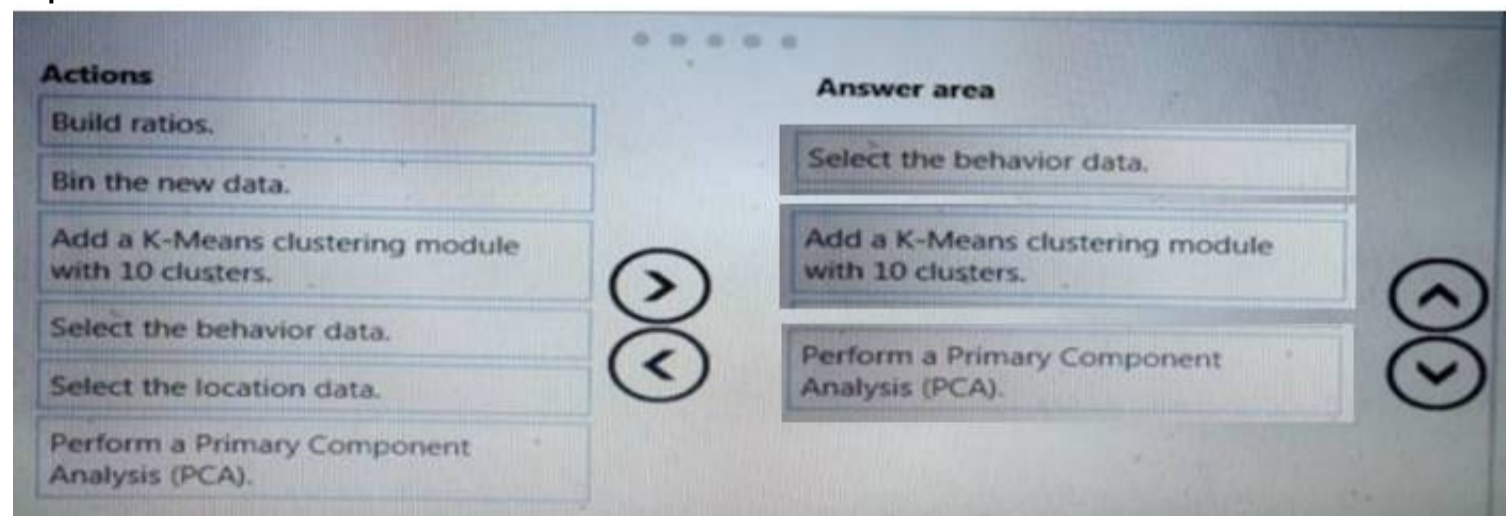
You need to modify the inputs for the global penalty event model to address the bias and variance issue. Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 142

- (Exam Topic 1)

You need to implement a new cost factor scenario for the ad response models as illustrated in the performance curve exhibit. Which technique should you use?

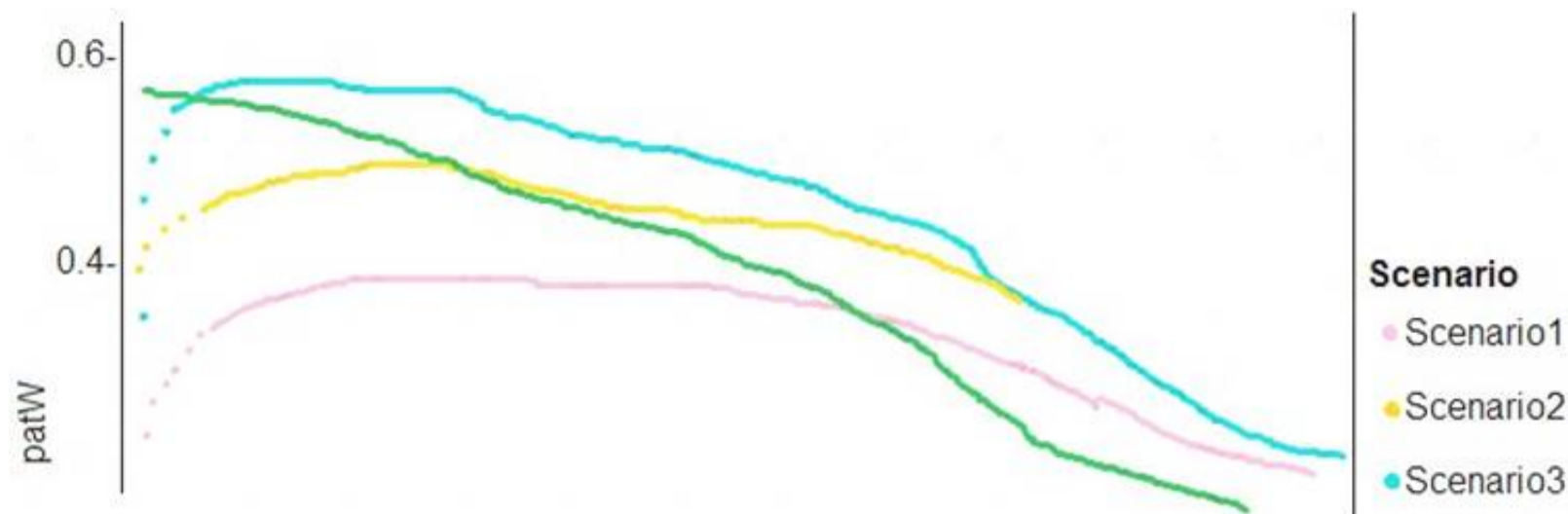
- A. Set the threshold to 0.5 and retrain if weighted Kappa deviates +/- 5% from 0.45.
- B. Set the threshold to 0.05 and retrain if weighted Kappa deviates +/- 5% from 0.5.
- C. Set the threshold to 0.2 and retrain if weighted Kappa deviates +/- 5% from 0.6.
- D. Set the threshold to 0.75 and retrain if weighted Kappa deviates +/- 5% from 0.15.

Answer: A

Explanation:

Scenario:

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



The ad propensity model uses a cut threshold is 0.45 and retrains occur if weighted Kappa deviated from 0.1 +/- 5%.

NEW QUESTION 146

- (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 torch 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 torch 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 150

.....

THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual DP-100 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the DP-100 Product From:

<https://www.2passeasy.com/dumps/DP-100/>

Money Back Guarantee

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