

Microsoft

Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure



NEW QUESTION 1

- (Exam Topic 3)

You have an Azure Machine Learning workspace that contains a training cluster and an inference cluster. You plan to create a classification model by using the Azure Machine Learning designer.

You need to ensure that client applications can submit data as HTTP requests and receive predictions as responses.

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 2

- (Exam Topic 3)

HOTSPOT

You create a script for training a machine learning model in Azure Machine Learning service. You create an estimator by running the following code:

```
from azureml.core import Workspace, Datastore
from azureml.core.compute import ComputeTarget
from azureml.train.estimator import Estimator
work_space = Workspace.from_config()
data_source = work_space.get_default_datastore()
train_cluster = ComputeTarget(workspace=work_space, name= 'train-cluster')
estimator = Estimator(source_directory =
    'training-experiment',
    script_params = { ' --data-folder' : data_source.as_mount(), ' --regularization':0.8},
    compute_target = train_cluster,
    entry_script = 'train.py',
    conda_packages = ['scikit-learn'])
```

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 estimator will look for the files it needs to run an experiment in the training-experiment directory of the local compute environment. Yes No
- The estimator will mount the local data-folder folder and make it available to the script through a parameter. Yes No
- The train.py script file will be created if it does not exist. Yes No
- The estimator can run Scikit-learn experiments. Yes No

- A. Mastered
 B. Not Mastered

Answer: A

Explanation:

Box 1: Yes
 Parameter source_directory is a local directory containing experiment configuration and code files needed for a training job.
 Box 2: Yes
 script_params is a dictionary of command-line arguments to pass to the training script specified in entry_script.
 Box 3: No
 Box 4: Yes
 The conda_packages parameter is a list of strings representing conda packages to be added to the Python environment for the experiment.

NEW QUESTION 3

- (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 4

- (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 5

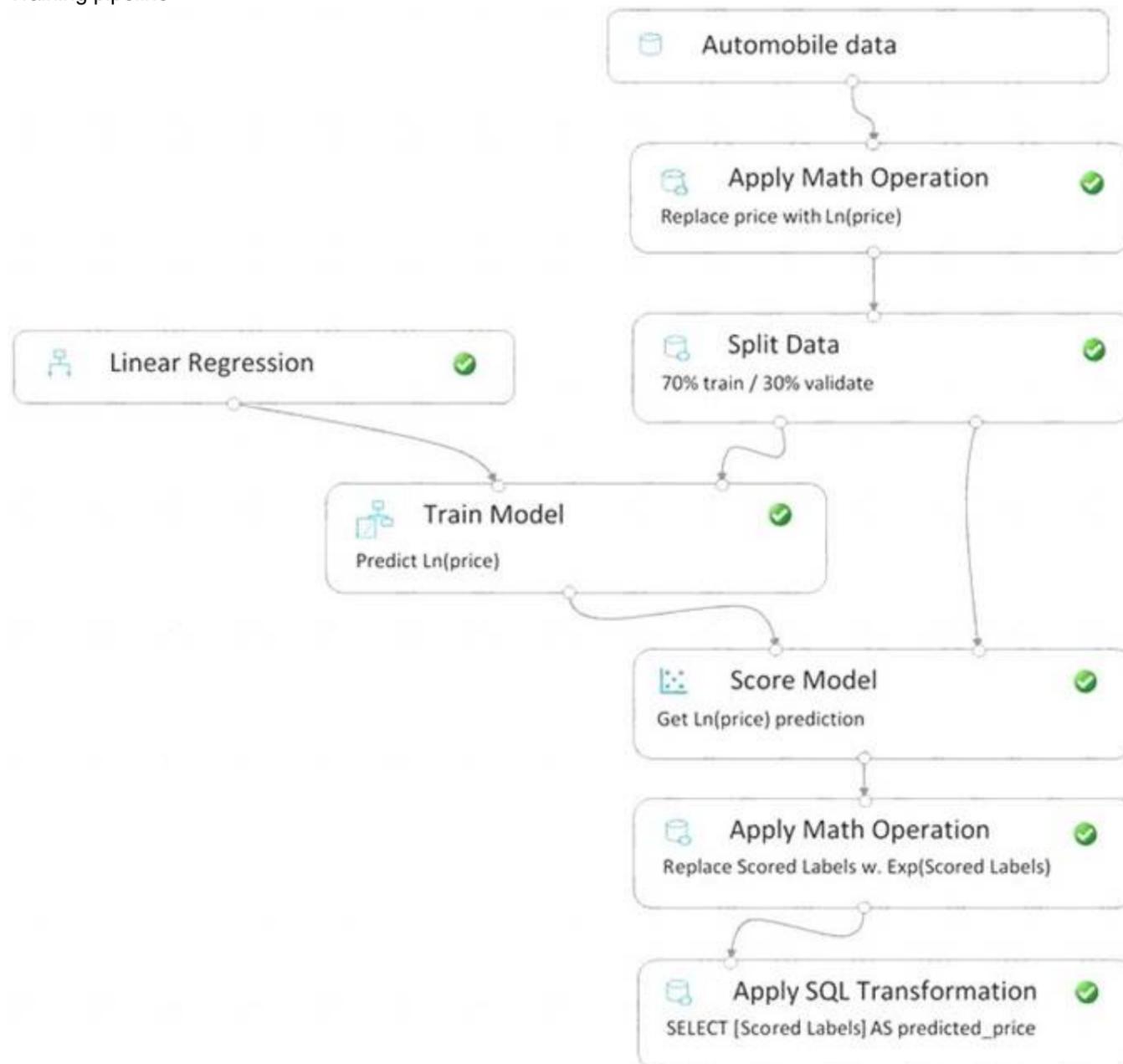
- (Exam Topic 3)

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.

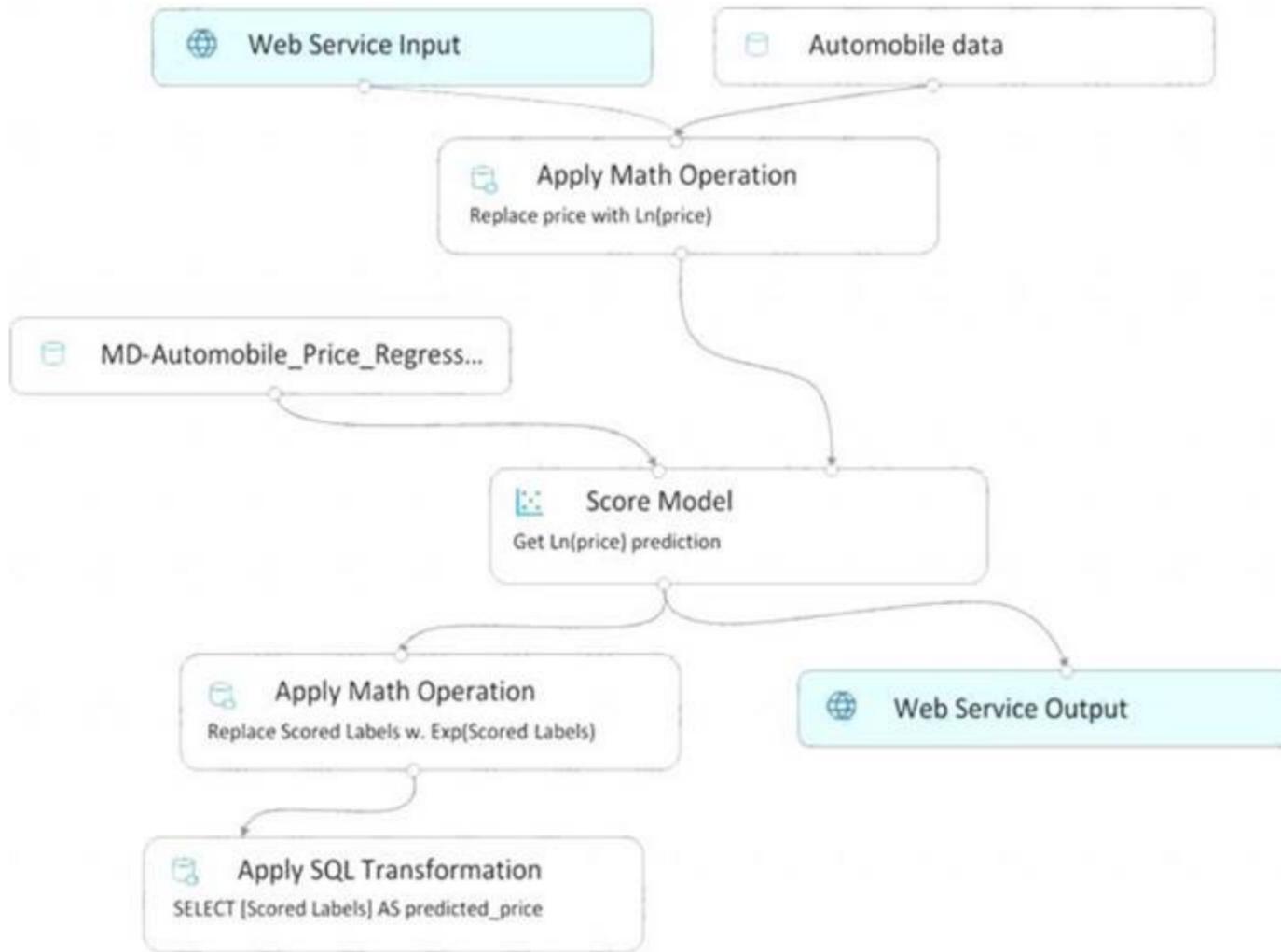
The training pipeline is shown in the exhibit. (Click the Training pipeline tab.)

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 6

- (Exam Topic 3)

You create a training pipeline using the Azure Machine Learning designer. You upload a CSV file that contains the data from which you want to train your model. You need to use the designer to create a pipeline that includes steps to perform the following tasks:

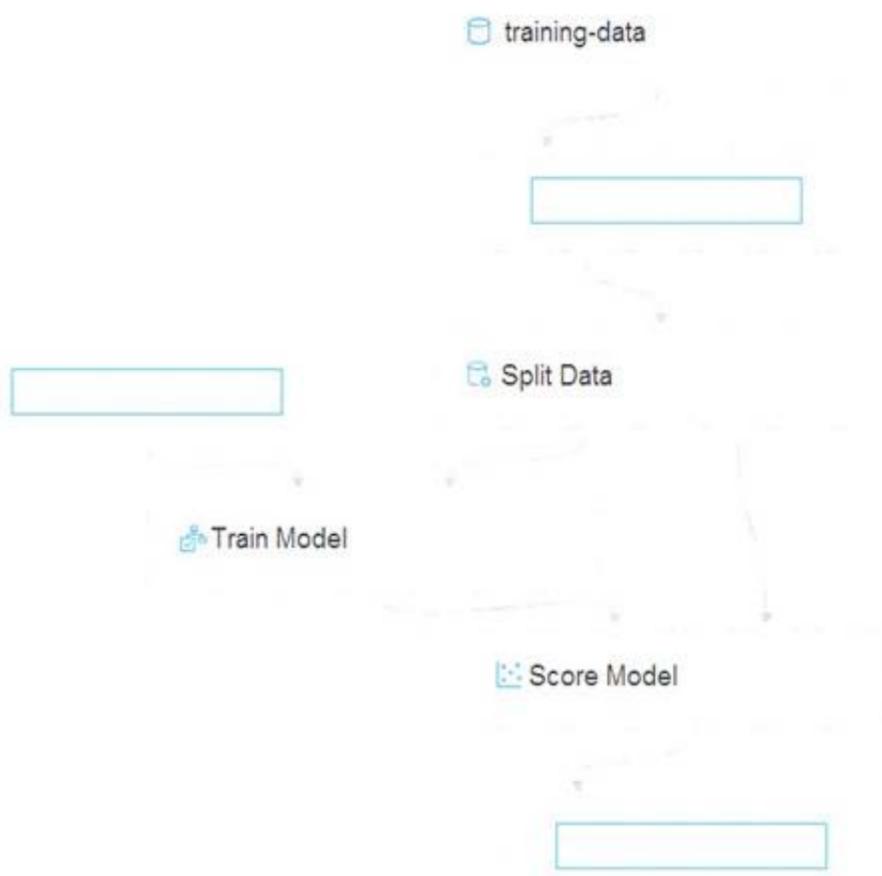
- > Select the training features using the pandas filter method.
- > Train a model based on the naive_bayes.GaussianNB algorithm.
- > Return only the Scored Labels column by using the query `SELECT [Scored Labels] FROM t1`; Which modules should you use? To answer, drag the appropriate modules to the appropriate locations. Each module name 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.

Modules

Answer Area

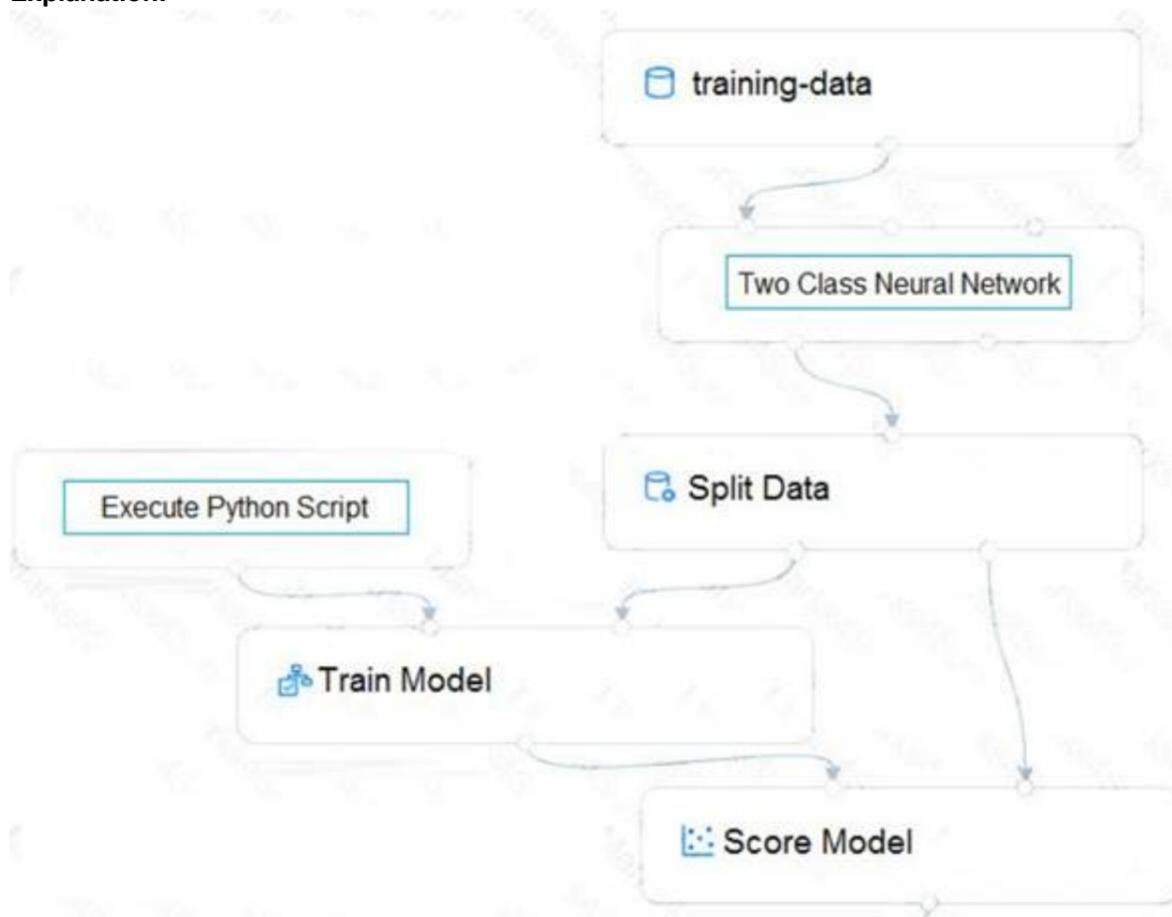
- Create Python Model
- Train Model
- Two Class Neural Network
- Execute Python Script
- Apply SQL Transformation
- Select Columns in Dataset



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 7

- (Exam Topic 3)

You are creating a classification model for a banking company to identify possible instances of credit card fraud. You plan to create the model in Azure Machine Learning by using automated machine learning.

The training dataset that you are using is highly unbalanced. You need to evaluate the classification model.

Which primary metric should you use?

- A. normalized_mean_absolute_error
- B. [spearman_correlation
- C. AUC.weighted
- D. accuracy
- E. normalized_root_mean_squared_error

Answer: C

Explanation:

AUC_weighted is a Classification metric.

Note: AUC is the Area under the Receiver Operating Characteristic Curve. Weighted is the arithmetic mean of the score for each class, weighted by the number of true instances in each class.

Reference:

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

NEW QUESTION 8

- (Exam Topic 3)

You are developing a deep learning model by using TensorFlow. You plan to run the model training workload on an Azure Machine Learning Compute Instance.

You must use CUDA-based model training. You need to provision the Compute Instance.

Which two virtual machines sizes can you use? To answer, select the appropriate virtual machine sizes in the answer area.

NOTE: Each correct selection is worth one point.

Virtual machine size

Name ↑	vCPUs	GPUs	RAM	Resource disk
BASIC_A0	1		0.75 GB	20 GB
STANDARD_D3_V2	4		14 GB	200 GB
STANDARD_E64_V3	64		432 GB	1,600 GB
STANDARD_M64LS	64		512 GB	2,000 GB
STANDARD_NC12	12	2	112 GB	680 GB
STANDARD_NC24	24	4	224 GB	1,440 GB

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

CUDA is a parallel computing platform and programming model developed by Nvidia for general computing on its own GPUs (graphics processing units). CUDA enables developers to speed up compute-intensive applications by harnessing the power of GPUs for the parallelizable part of the computation.

Reference:

<https://www.infoworld.com/article/3299703/what-is-cuda-parallel-programming-for-gpus.html>

NEW QUESTION 9

- (Exam Topic 3)

You create an Azure Machine Learning workspace.

You need to detect data drift between a baseline dataset and a subsequent target dataset by using the DataDriftDetector class.

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.

```

from azureml.core import Workspace, Dataset
from datetime import datetime

ws = Workspace.from_config()
dset = Dataset.get_by_name(ws, 'target')
baseline = target.time_before(datetime(2021, 2, 1))
features = ['windAngle', 'windSpeed', 'temperature', 'stationName']

monitor = DataDriftDetector.      (ws, 'drift-monitor', baseline,
                                backfill
                                create_from_datasets
                                create_from_model

target, compute_target='cpu-cluster', frequency='Week', feature_list=None,
drift_threshold=.6, latency=24)

monitor = DataDriftDetector.get_by_name(ws, 'drift-monitor')
monitor = monitor.update(feature_list=features)
complete = monitor.      (datetime(2021, 1, 1), datetime.today())
                                backfill
                                list
                                update

```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

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

Box 1: create_from_datasets

The create_from_datasets method creates a new DataDriftDetector object from a baseline tabular dataset and a target time series dataset.

Box 2: backfill

The backfill method runs a backfill job over a given specified start and end date.

Syntax: backfill(start_date, end_date, compute_target=None, create_compute_target=False) Reference:

[https://docs.microsoft.com/en-us/python/api/azureml-datadrift/azureml.datadrift.datadriftdetector\(class\)](https://docs.microsoft.com/en-us/python/api/azureml-datadrift/azureml.datadrift.datadriftdetector(class))

NEW QUESTION 10

- (Exam Topic 3)

Your Azure Machine Learning workspace has a dataset named real_estate_data. A sample of the data in the dataset follows.

postal_code	num_bedrooms	sq_feet	garage	price
12345	3	1300	0	23,9000
54321	1	950	0	11,0000
12346	2	1200	1	15,0000

You want to use automated machine learning to find the best regression model for predicting the price column. You need to configure an automated machine learning experiment using the Azure Machine Learning SDK. How should you complete the code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

```
from azureml.core import Workspace
from azureml.core.compute import ComputeTarget
from azureml.core.runconfig import RunConfiguration
from azureml.train.automl import AutoMLConfig

ws = Workspace.from_config()
training_cluster = ComputeTarget(workspace=ws, name='aml-cluster1')
real_estate_ds = ws.datasets.get('real_estate_data')
split1_ds, split2_ds = real_estate_ds.random_split(percentage=0.7, seed=123)
automl_run_config = RunConfiguration(framework="python")
automl_config = AutoMLConfig(
    task='regression',
    compute_target=training_cluster,
    run_configuration=automl_run_config,
    primary_metric='r2_score',
    =split1_ds,
```

X
Y
X_valid
Y_valid
training_data

X
Y
X_valid
Y_valid
validation_data
training_data

y
y_valid
y_max
label_column_name
exclude_nan_labels

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: training_data

The training data to be used within the experiment. It should contain both training features and a label column (optionally a sample weights column). If training_data is specified, then the label_column_name parameter must also be specified.

Box 2: validation_data

Provide validation data: In this case, you can either start with a single data file and split it into training and validation sets or you can provide a separate data file for the validation set. Either way, the validation_data parameter in your AutoMLConfig object assigns which data to use as your validation set.

Example, the following code example explicitly defines which portion of the provided data in dataset to use for training and validation.

```
dataset = Dataset.Tabular.from_delimited_files(data)
```

```
training_data, validation_data = dataset.random_split(percentage=0.8, seed=1) automl_config = AutoMLConfig(compute_target = aml_remote_compute, task = 'classification',
```

```
primary_metric = 'AUC_weighted', training_data = training_data,
validation_data = validation_data, label_column_name = 'Class'
)
```

Box 3: label_column_name label_column_name:

The name of the label column. If the input data is from a pandas.DataFrame which doesn't have column names, column indices can be used instead, expressed as integers.

This parameter is applicable to training_data and validation_data parameters. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train.automl.automlconfig.auto>

NEW QUESTION 10

- (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

	▼
the Microsoft Surface device	
the ds-workstation notebook VM	
the cpu-cluster compute target	
the gpu-cluster compute target	

Run the training script

	▼
the ds-workstation notebook VM	
the cpu-compute target	
the gpu-compute target	
the Microsoft Surface device	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Resource type

Option

Run code to submit the experiment

	▼
the Microsoft Surface device	
the ds-workstation notebook VM	
the cpu-cluster compute target	
the gpu-cluster compute target	

Run the training script

	▼
the ds-workstation notebook VM	
the cpu-compute target	
the gpu-compute target	
the Microsoft Surface device	

NEW QUESTION 15

- (Exam Topic 3)

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.

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Add a Two-Class Support Vector Machine module to initialize the SVM classifier.

Step 2: Add a dataset to the experiment

Step 3: Add a Split Data module to create training and test dataset.

To generate a set of feature scores requires that you have an already trained model, as well as a test dataset. Step 4: Add a Permutation Feature Importance module and connect to the trained model and test dataset. Step 5: Set the Metric for measuring performance property to Classification - Accuracy and then run the experiment.

Reference:

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

NEW QUESTION 17

- (Exam Topic 3)

You use the designer to create a training pipeline for a classification model. The pipeline uses a dataset that includes the features and labels required for model training.

You create a real-time inference pipeline from the training pipeline. You observe that the schema for the generated web service input is based on the dataset and includes the label column that the model predicts. Client applications that use the service must not be required to submit this value.

You need to modify the inference pipeline to meet the requirement. What should you do?

- A. Add a Select Columns in Dataset module to the inference pipeline after the dataset and use it to select all columns other than the label.
- B. Delete the dataset from the training pipeline and recreate the real-time inference pipeline.
- C. Delete the Web Service Input module from the inference pipeline.
- D. Replace the dataset in the inference pipeline with an Enter Data Manually module that includes data for the feature columns but not the label column.

Answer: A

Explanation:

By default, the Web Service Input will expect the same data schema as the module output data which connects to the same downstream port as it. You can remove the target variable column in the inference pipeline using Select Columns in Dataset module. Make sure that the output of Select Columns in Dataset removing target variable column is connected to the same port as the output of the Web Service Input module.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-deploy>

NEW QUESTION 20

- (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.dnn import TensorFlow
sk_est = TensorFlow(source_directory='./scripts',
compute_target=aml-compute,
entry_script='train.py')
```

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'
)
```

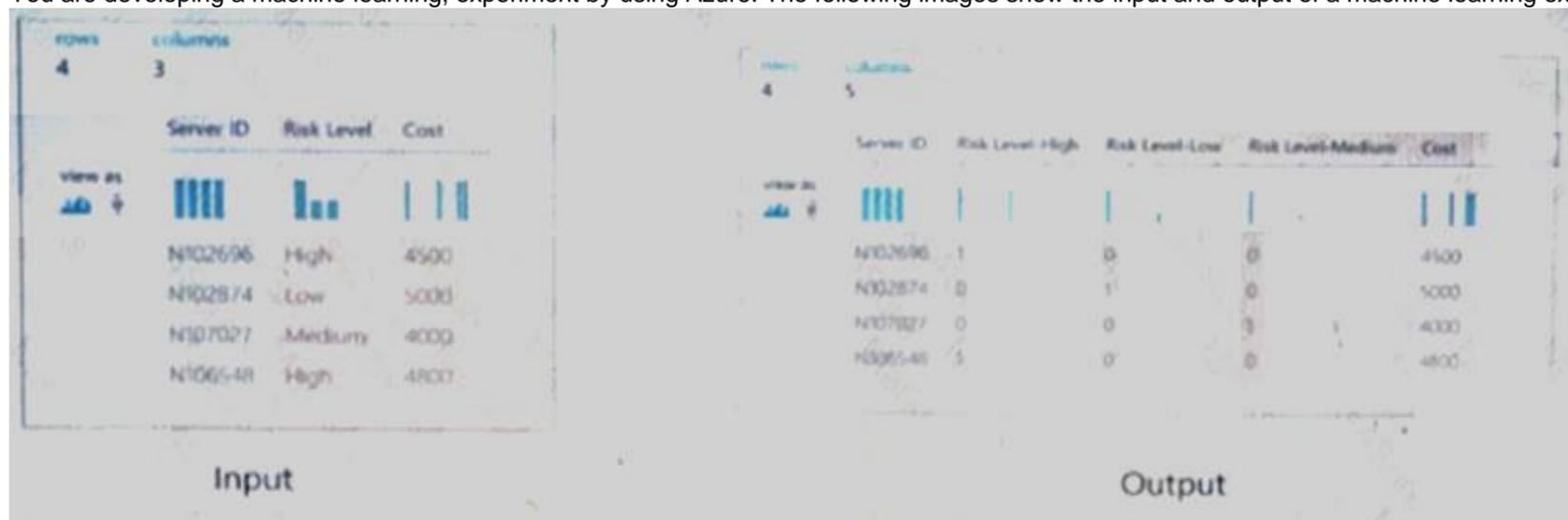
Reference:

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

NEW QUESTION 25

- (Exam Topic 3)

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



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

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

What is the expected input column type for this transformation?

Apply Filter

Build Counting Transform

Convert to Indicator Values

Categorical

Numerical

String

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

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

What is the expected input column type for this transformation?

Apply Filter

Build Counting Transform

Convert to Indicator Values

Categorical

Numerical

String

NEW QUESTION 27

- (Exam Topic 3)

A set of CSV files contains sales records. All the CSV files have the same data schema.

Each CSV file contains the sales record for a particular month and has the filename sales.csv. Each file is stored in a folder that indicates the month and year when the data was recorded. The folders are in an Azure blob container for which a datastore has been defined in an Azure Machine Learning workspace. The folders are organized in a parent folder named sales to create the following hierarchical structure:

```
/sales
/01-2019
  /sales.csv
/02-2019
  /sales.csv
/03-2019
  /sales.csv
...
```

At the end of each month, a new folder with that month's sales file is added to the sales folder.

You plan to use the sales data to train a machine learning model based on the following requirements:

- You must define a dataset that loads all of the sales data to date into a structure that can be easily converted to a dataframe.
- You must be able to create experiments that use only data that was created before a specific previous month, ignoring any data that was added after that month.
- You must register the minimum number of datasets possible.

You need to register the sales data as a dataset in Azure Machine Learning service workspace. What should you do?

- A. Create a tabular dataset that references the datastore and explicitly specifies each 'sales/mm-yyyy/ sales.csv' file every month
- B. Register the dataset with the name sales_dataset each month, replacing the existing dataset and specifying a tag named month indicating the month and year it was registered
- C. Use this dataset for all experiments.
- D. Create a tabular dataset that references the datastore and specifies the path 'sales/*/sales.csv', register the dataset with the name sales_dataset and a tag named month indicating the month and year it was registered, and use this dataset for all experiments.
- E. Create a new tabular dataset that references the datastore and explicitly specifies each 'sales/mm-yyyy/ sales.csv' file every month
- F. Register the dataset with the name sales_dataset_MM-YYYY each month with appropriate MM and YYYY values for the month and year
- G. Use the appropriate month-specific dataset for experiments.
- H. Create a tabular dataset that references the datastore and explicitly specifies each 'sales/mm-yyyy/ sales.csv' file
- I. Register the dataset with the name sales_dataset each month as a new version and with a tag named month indicating the month and year it was registered
- J. Use this dataset for all experiments, identifying the version to be used based on the month tag as necessary.

Answer: B

Explanation:

Specify the path. Example:

The following code gets the workspace existing workspace and the desired datastore by name. And then passes the datastore and file locations to the path parameter to create a new TabularDataset, weather_ds.

```
from azureml.core import Workspace, Datastore, Dataset
datastore_name = 'your datastore name'
# get existing workspace
workspace = Workspace.from_config()
# retrieve an existing datastore in the workspace by name
datastore = Datastore.get(workspace, datastore_name)
# create a TabularDataset from 3 file paths in datastore
datastore_paths = [(datastore, 'weather/2018/11.csv'), (datastore, 'weather/2018/12.csv'), (datastore, 'weather/2019/*.csv')]
weather_ds = Dataset.Tabular.from_delimited_files(path=datastore_paths)
```

NEW QUESTION 32

- (Exam Topic 3)

You retrain an existing model.

You need to register the new version of a model while keeping the current version of the model in the registry.

What should you do?

- A. Register a model with a different name from the existing model and a custom property named version with the value 2.
- B. Register the model with the same name as the existing model.
- C. Save the new model in the default datastore with the same name as the existing model
- D. Do not register the new model.
- E. Delete the existing model and register the new one with the same name.

Answer: B

Explanation:

Model version: A version of a registered model. When a new model is added to the Model Registry, it is added as Version 1. Each model registered to the same model name increments the version number.

Reference:

<https://docs.microsoft.com/en-us/azure/databricks/applications/mlflow/model-registry>

NEW QUESTION 37

- (Exam Topic 3)

You use the Azure Machine Learning designer to create and run a training pipeline.

The pipeline must be run every night to inference predictions from a large volume of files. The folder where the files will be stored is defined as a dataset.

You need to publish the pipeline as a REST service that can be used for the nightly inferencing run. What should you do?

- A. Create a batch inference pipeline
- B. Set the compute target for the pipeline to an inference cluster
- C. Create a real-time inference pipeline
- D. Clone the pipeline

Answer: A

Explanation:

Azure Machine Learning Batch Inference targets large inference jobs that are not time-sensitive. Batch Inference provides cost-effective inference compute scaling, with unparalleled throughput for asynchronous applications. It is optimized for high-throughput, fire-and-forget inference over large collections of data. You can submit a batch inference job by pipeline_run, or through REST calls with a published pipeline. Reference: <https://github.com/Azure/MachineLearningNotebooks/blob/master/how-to-use-azureml/machine-learning-pipeli>

NEW QUESTION 39

- (Exam Topic 3)

You are training machine learning models in Azure Machine Learning. You use Hyperdrive to tune the hyperparameters. In previous model training and tuning runs, many models showed similar performance. You need to select an early termination policy that meets the following requirements:

- accounts for the performance of all previous runs when evaluating the current run
- avoids comparing the current run with only the best performing run to date

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

- A. Bandit
- B. Median stopping
- C. Default
- D. Truncation selection

Answer: BC

Explanation:

The Median Stopping policy computes running averages across all runs and cancels runs whose best performance is worse than the median of the running averages.

If no policy is specified, the hyperparameter tuning service will let all training runs execute to completion. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.medianstoppingpolicy> <https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.truncationselectionpoli> <https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.banditpolicy>

NEW QUESTION 40

- (Exam Topic 3)

You run an experiment that uses an AutoMLConfig class to define an automated machine learning task with a maximum of ten model training iterations. The task will attempt to find the best performing model based on a metric named accuracy.

You submit the experiment with the following code:

You need to create Python code that returns the best model that is generated by the automated machine learning task. Which code segment should you use?

A)

```
best_model = automl_run.get_details()
```

B)

```
best_model = automl_run.get_output()[1]
```

C)

```
best_model = automl_run.get_file_names()[1]
```

D)

```
best_model = automl_run.get_metrics()
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D

Answer: B

Explanation:

The get_output method returns the best run and the fitted model. Reference:

<https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/automated-mach>

NEW QUESTION 43

- (Exam Topic 3)

You plan to use a Deep Learning Virtual Machine (DLVM) to train deep learning models using Compute Unified Device Architecture (CUDA) computations.

You need to configure the DLVM to support CUDA. What should you implement?

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

Answer: C

Explanation:

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>

NEW QUESTION 48

- (Exam Topic 3)

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

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

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

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

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

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

- A. Yes
- B. No

Answer: B

Explanation:

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

NEW QUESTION 53

- (Exam Topic 3)

You create an Azure Machine Learning workspace.

You must create a custom role named DataScientist that meets the following requirements:

- Role members must not be able to delete the workspace.
- Role members must not be able to create, update, or delete compute resource in the workspace.
- Role members must not be able to add new users to the workspace.

You need to create a JSON file for the DataScientist role in the Azure Machine Learning workspace. The custom role must enforce the restrictions specified by the IT Operations team.

Which JSON code segment should you use?

A)

```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": ["*"],
  "NotActions": [
    "Microsoft.MachineLearningServices/workspaces/*/delete",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft/MachineLearningServices/workspaces/ml-ws"
  ]
}
```

B)

```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": ["*"],
  "NotActions": [],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
  ]
}
```

C)

```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": [
    "Microsoft.MachineLearningServices/workspaces/*/delete",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "NotActions": [],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
  ]
}
```

D)

```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": [],
  "NotActions": ["*"],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
  ]
}
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D

Answer: A

Explanation:

The following custom role can do everything in the workspace except for the following actions:

- > It can't create or update a compute resource.
- > It can't delete a compute resource.
- > It can't add, delete, or alter role assignments.
- > It can't delete the workspace.

To create a custom role, first construct a role definition JSON file that specifies the permission and scope for the role. The following example defines a custom role named "Data Scientist Custom" scoped at a specific workspace level:

data_scientist_custom_role.json :

```
{
  "Name": "Data Scientist Custom", "IsCustom": true,
  "Description": "Can run experiment but can't create or delete compute.", "Actions": ["*"],
  "NotActions": [
    "Microsoft.MachineLearningServices/workspaces/*/delete", "Microsoft.MachineLearningServices/workspaces/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write", "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "AssignableScopes": [ "/subscriptions/<subscription_id>/resourceGroups/<resource_group_name>/providers/Microsoft.MachineLearnin"
]
}
```

Reference:

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

NEW QUESTION 57

- (Exam Topic 3)

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

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

Answer: D

NEW QUESTION 60

- (Exam Topic 3)

You are using the Hyperdrive feature in Azure Machine Learning to train a model. You configure the Hyperdrive experiment by running the following code:

```
from azureml.train.hyperdrive import RandomParameterSampling
param_sampling = RandomParameterSampling( {
  "learning_rate": normal(10, 3),
  "keep_probability": uniform(0.05, 0.1),
  "batch_size": choice(16, 32, 64, 128)
  "number_of_hidden_layers": choice(range(3,5))
})
```

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
By defining sampling in this manner, every possible combination of the parameters will be tested.	<input type="radio"/>	<input type="radio"/>
Random values of the learning_rate parameter will be selected from a normal distribution with a mean of 10 and a standard deviation of 3.	<input type="radio"/>	<input type="radio"/>
The keep_probability parameter value will always be either 0.05 or 0.1 .	<input type="radio"/>	<input type="radio"/>
Random values for the number_of_hidden_layers parameter will be selected from a normal distribution with a mean of 3 and a standard deviation of 5.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Yes

In random sampling, hyperparameter values are randomly selected from the defined search space. Random sampling allows the search space to include both discrete and continuous hyperparameters.

Box 2: Yes

learning_rate has a normal distribution with mean value 10 and a standard deviation of 3.

Box 3: No

keep_probability has a uniform distribution with a minimum value of 0.05 and a maximum value of 0.1.

Box 4: No

number_of_hidden_layers takes on one of the values [3, 4, 5].

Reference:

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

NEW QUESTION 62

- (Exam Topic 3)

You create a multi-class image classification deep learning model that uses a set of labeled images. You create a script file named train.py that uses the PyTorch 1.3 framework to train the model.

You must run the script by using an estimator. The code must not require any additional Python libraries to be installed in the environment for the estimator. The time required for model training must be minimized.

You need to define the estimator that will be used to run the script. Which estimator type should you use?

- A. TensorFlow
- B. PyTorch
- C. SKLearn
- D. Estimator

Answer: B

Explanation:

For PyTorch, TensorFlow and Chainer tasks, Azure Machine Learning provides respective PyTorch, TensorFlow, and Chainer estimators to simplify using these frameworks.

Reference:

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

NEW QUESTION 65

- (Exam Topic 3)

You create a batch inference pipeline by using the Azure ML SDK. You run the pipeline by using the following code:

```
from azureml.pipeline.core import Pipeline
from azureml.core.experiment import Experiment
```

```
pipeline = Pipeline(workspace=ws, steps=[parallelrun_step]) pipeline_run = Experiment(ws, 'batch_pipeline').submit(pipeline) You need to monitor the progress of the pipeline execution.
```

What are two possible ways to achieve this goal? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. Run the following code in a notebook:
- ```
from azureml.contrib.interpret.explanation.explanation_client import ExplanationClient
client = ExplanationClient.from_run(pipeline_run)
explanation = client.download_model_explanation()
explanation = client.download_model_explanation(top_k=4)
global_importance_values = explanation.get_ranked_global_values()
global_importance_names = explanation.get_ranked_global_names()
print('global importance values: {}'.format(global_importance_values))
print('global importance names: {}'.format(global_importance_names))
```
- B. Use the Inference Clusters tab in Machine Learning Studio.
- C. Use the Activity log in the Azure portal for the Machine Learning workspace.
- D. Run the following code in a notebook:
- ```
from azureml.widgets import RunDetails
RunDetails(pipeline_run).show()
```
- E. Run the following code and monitor the console output from the PipelineRun object:
- ```
pipeline_run.wait_for_completion(show_output=True)
```

- A. Option A  
 B. Option B  
 C. Option C  
 D. Option D  
 E. Option E

**Answer:** DE

**Explanation:**

A batch inference job can take a long time to finish. This example monitors progress by using a Jupyter widget. You can also manage the job's progress by using:

- Azure Machine Learning Studio.
- Console output from the PipelineRun object. `from azureml.widgets import RunDetails RunDetails(pipeline_run).show()`

`pipeline_run.wait_for_completion(show_output=True)` Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-parallel-run-step#monitor-the-parallel-run>

**NEW QUESTION 69**

- (Exam Topic 3)

You use Azure Machine Learning designer to create a training pipeline for a regression model.

You need to prepare the pipeline for deployment as an endpoint that generates predictions asynchronously for a dataset of input data values.

What should you do?

- A. Clone the training pipeline.  
 B. Create a batch inference pipeline from the training pipeline.  
 C. Create a real-time inference pipeline from the training pipeline.  
 D. Replace the dataset in the training pipeline with an Enter Data Manually module.

**Answer:** C

**Explanation:**

You must first convert the training pipeline into a real-time inference pipeline. This process removes training modules and adds web service inputs and outputs to handle requests.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-deploy> <https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/enter-data-manually>

**NEW QUESTION 71**

- (Exam Topic 3)

You create an Azure Machine Learning workspace and a new Azure DevOps organization. You register a model in the workspace and deploy the model to the target environment.

All new versions of the model registered in the workspace must automatically be deployed to the target environment.

You need to configure Azure Pipelines to deploy the model.

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**

- Create a service connection
- Create a release pipeline
- Create a build pipeline
- Create an Azure DevOps project
- Install the Machine Learning extension for Azure Pipelines



- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Graphical user interface, text, application, email Description automatically generated  
 Step 1: Create an Azure DevOps project  
 Step 2: Create a release pipeline  
 > Sign in to your Azure DevOps organization and navigate to your project.  
 > Go to Pipelines, and then select New pipeline.  
 Step 3: Install the Machine Learning extension for Azure Pipelines  
 You must install and configure the Azure CLI and ML extension.  
 Step 4: Create a service connection  
 How to set up your service connection  
 Graphical user interface, text, application, email Description automatically generated

**Project Settings**

**General**

Overview

Teams

Security

Notifications

Service hooks

Dashboards

**Boards**

Project configuration

Team configuration

GitHub connections

**Pipelines**

Service connections

Agent pools

Retention and parallel jobs

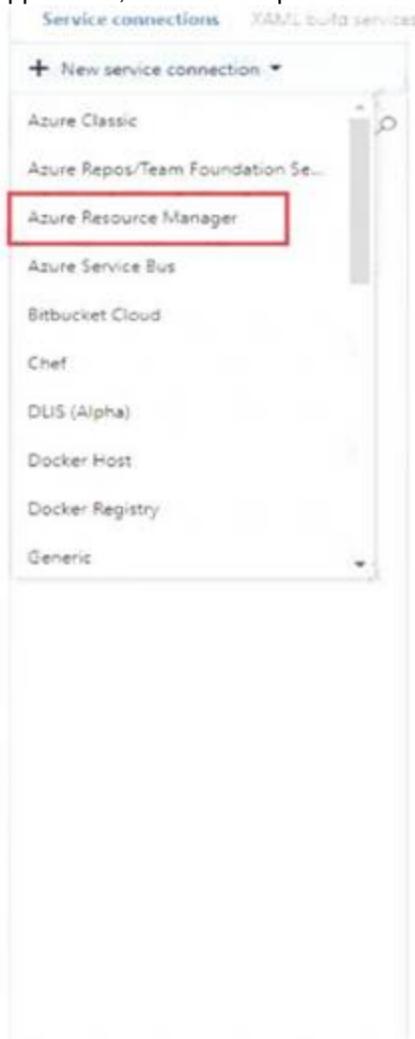
Release retention

**Repos**

Repositories

Policies

**Test**



**Service connection: demo**

Details Roles Request history Policies

**INFORMATION**

**ACTIONS**

List of actions that can be performed on this service connection:

- Update service connection
- Manage service connection roles
- Manage Service Principal
- Disconnect

Select AzureMLWorkspace for the scope level, then fill in the following subsequent parameters. Graphical user interface, text, application Description automatically generated

Note: How to enable model triggering in a release pipeline

- > Go to your release pipeline and add a new artifact. Click on AzureML Model artifact then select the appropriate AzureML service connection and select from the available models in your workspace.
- > Enable the deployment trigger on your model artifact as shown here. Every time a new version of that model is registered, a release pipeline will be triggered.

Reference:

<https://marketplace.visualstudio.com/items?itemName=ms-air-aiagility.vss-services-azureml> <https://docs.microsoft.com/en-us/azure/devops/pipelines/targets/azure-machine-learning>

**NEW QUESTION 73**

- (Exam Topic 3)

You are building a binary classification model by using a supplied training set. The training set is imbalanced between two classes. You need to resolve the data imbalance.

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

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

**Answer:** ABD

**Explanation:**

References:  
<https://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/>

**NEW QUESTION 74**

- (Exam Topic 3)

You create a Python script that runs a training experiment in Azure Machine Learning. The script uses the Azure Machine Learning SDK for Python. You must add a statement that retrieves the names of the logs and outputs generated by the script. You need to reference a Python class object from the SDK for the statement.

Which class object should you use?

- A. Run
- B. ScriptRunConfig
- C. Workspace
- D. Experiment

**Answer:** A

**Explanation:**

A run represents a single trial of an experiment. Runs are used to monitor the asynchronous execution of a trial, log metrics and store output of the trial, and to analyze results and access artifacts generated by the trial.

The run Class get\_all\_logs method downloads all logs for the run to a directory. Reference:  
[https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run\(class\)](https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run(class))

**NEW QUESTION 79**

- (Exam Topic 3)

You are the owner of an Azure Machine Learning workspace.

You must prevent the creation or deletion of compute resources by using a custom role. You must allow all other operations inside the workspace.

You need to configure the custom role.

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

NOTE: Each correct selection is worth one point.

**Answer Area**

```
{
 "Name": "Data Scientist Custom",
 "IsCustom": true
 "Description": "Description"
 "Actions": [
```

▼

Microsoft.MachineLearningServices/workspaces/\*/read  
 Microsoft.MachineLearningServices/workspaces/computes/\*/write  
 Microsoft.MachineLearningServices/workspaces/delete

▼

Microsoft.MachineLearningServices/workspaces/\*/write  
 Microsoft.MachineLearningServices/workspaces/computes/\*/write  
 Microsoft.MachineLearningServices/workspaces/delete

```
],
 "NotActions": [
```

▼

Microsoft.MachineLearningServices/workspaces/\*/read  
 Microsoft.MachineLearningServices/workspaces/\*/write  
 Microsoft.MachineLearningServices/workspaces/computes/\*/delete

▼

Microsoft.MachineLearningServices/workspaces/\*/read  
 Microsoft.MachineLearningServices/workspaces/\*/write  
 Microsoft.MachineLearningServices/workspaces/computes/\*/write

```
],
 "AssignableScopes": [
 "/subscriptions/<subscription_id>"
]
}
```

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

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

Box 1: Microsoft.MachineLearningServices/workspaces/\*/read

Reader role: Read-only actions in the workspace. Readers can list and view assets, including datastore credentials, in a workspace. Readers can't create or update these assets.

Box 2: Microsoft.MachineLearningServices/workspaces/\*/write

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

Box 3: Box 2: Microsoft.MachineLearningServices/workspaces/computes/\*/delete

Box 4: Microsoft.MachineLearningServices/workspaces/computes/\*/write Reference:

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

**NEW QUESTION 82**

- (Exam Topic 3)

You are using a Git repository to track work in an Azure Machine Learning workspace. You need to authenticate a Git account by using SSH.

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 |
|---------------------------------------------------------|-------------|
| Generate a public/private key pair                      |             |
| Add the private key to the Git account                  |             |
| Clone the Git repository by using an SSH repository URL | ➤           |
| Add the public key to the Git account                   | ➤           |
| Create a new Azure Key Vault resource                   |             |

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Graphical user interface, text, application, chat or text message Description automatically generated

Authenticate your Git Account with SSH: Step 1: Generating a public/private key pair Generate a new SSH key

\* 1. Open the terminal window in the Azure Machine Learning Notebook Tab.

\* 2. Paste the text below, substituting in your email address. `ssh-keygen -t rsa -b 4096 -C "your_email@example.com"`

This creates a new ssh key, using the provided email as a label.

> Generating public/private rsa key pair.

Step 2: Add the public key to the Git Account

In your terminal window, copy the contents of your public key file. Step 3: Clone the Git repository by using an SSH repository URL

\* 1. Copy the SSH Git clone URL from the Git repo.

\* 2. Paste the url into the git clone command below, to use your SSH Git repo URL. This will look something like:

`git clone git@example.com:GitUser/azureml-example.git Cloning into 'azureml-example'.`

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-train-model-git-integration>

**NEW QUESTION 83**

- (Exam Topic 3)

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

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

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

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

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

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

- A. Yes
- B. NO

**Answer:** A

**Explanation:**

Replace using MICE: For each missing value, this option assigns a new value, which is calculated by using a method described in the statistical literature as

"Multivariate Imputation using Chained Equations" or "Multiple Imputation by Chained Equations". With a multiple imputation method, each variable with missing data is modeled conditionally using the other variables in the data before filling in the missing values.

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

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

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

**NEW QUESTION 86**

- (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.

- 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 88**

- (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 92**

- (Exam Topic 3)

You register a file dataset named csvfolder that references a folder. The folder includes multiple comma-separated values (CSV) files in an Azure storage blob container. You plan to use the following code to run a script that loads data from the file dataset. You create and instantiate the following variables:

| Variable       | Description                                           |
|----------------|-------------------------------------------------------|
| remote_cluster | References the Azure Machine Learning compute cluster |
| ws             | References the Azure Machine Learning workspace       |

You have the following code:

```
from azureml.train.estimator import Estimator
file_dataset = ws.datasets.get('csv_folder')
estimator = Estimator(source_directory=script_folder,

 compute_target = remote_cluster,
 entry_script='script.py')
run = experiment.submit(config=estimator)
run.wait_for_completion(show_output=True)
```

You need to pass the dataset to ensure that the script can read the files it references. Which code segment should you insert to replace the code comment?

A)

```
inputs=[file_dataset.as_named_input('training_files').to_pandas_dataframe()],
```

B)

```
inputs=[file_dataset.as_named_input('training_files').as_mount()],
```

C)

```
script_params={'--training_files': file_dataset},
```

D)

```
inputs=[file_dataset.as_named_input('training_files')],
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D

**Answer: D**

**Explanation:**

Example:

```
from azureml.train.estimator import Estimator
script_params = {
to mount files referenced by mnist dataset
'--data-folder': mnist_file_dataset.as_named_input('mnist_opendataset').as_mount(),
'--regularization': 0.5
}
est = Estimator(source_directory=script_folder, script_params=script_params, compute_target=compute_target, environment_definition=env,
entry_script='train.py')
```

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-train-models-with-aml>

**NEW QUESTION 94**

- (Exam Topic 3)

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

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

**Answer: D**

**Explanation:**

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

Reference:

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

**NEW QUESTION 96**

- (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 101**

- (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.sklearn import SKLearn
sk_est = SKLearn(source_directory='./scripts',
 compute_target=aml-compute,
 entry_script='train.py')
```

Does the solution meet the goal?

- A. Yes
- B. No

**Answer:** A

**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 103**

- (Exam Topic 3)

You are implementing a machine learning model to predict stock prices. The model uses a PostgreSQL database and requires GPU processing. You need to create a virtual machine that is pre-configured with the required tools. What should you do?

- A. Create a Data Science Virtual Machine (DSVM) Windows edition.
- B. Create a Geo AI Data Science Virtual Machine (Geo-DSVM) Windows edition.
- C. Create a Deep Learning Virtual Machine (DLVM) Linux edition.
- D. Create a Deep Learning Virtual Machine (DLVM) Windows edition.
- E. Create a Data Science Virtual Machine (DSVM) Linux edition.

**Answer:** E

**NEW QUESTION 108**

- (Exam Topic 3)

The finance team asks you to train a model using data in an Azure Storage blob container named `finance-data`. You need to register the container as a datastore in an Azure Machine Learning workspace and ensure that an error will be raised if the container does not exist.

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

```

datastore = Datastore. (workspace = ws,
register_azure_blob_container
register_azure_file_share
register_azure_data_lake
register_azure_sql_database

datastore_name = 'finance_datastore',
container_name = 'finance-data',
account_name = 'fintrainingdatastorage',
account_key = 'FWUYORRv3XoyNe...',

create_if_not_exists = True
create_if_not_exists = False
overwrite = True
overwrite = False

```

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: register\_azure\_blob\_container  
 Register an Azure Blob Container to the datastore.  
 Box 2: create\_if\_not\_exists = False  
 Create the file share if it does not exists, defaults to False. Reference:  
<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.datastore.datastore>

**NEW QUESTION 110**

- (Exam Topic 3)

You create a machine learning model by using the Azure Machine Learning designer. You publish the model as a real-time service on an Azure Kubernetes Service (AKS) inference compute cluster. You make no changes to the deployed endpoint configuration. You need to provide application developers with the information they need to consume the endpoint. Which two values should you provide to application developers? Each correct answer presents part of the solution.  
 NOTE: Each correct selection is worth one point.

- A. The name of the AKS cluster where the endpoint is hosted.
- B. The name of the inference pipeline for the endpoint.
- C. The URL of the endpoint.
- D. The run ID of the inference pipeline experiment for the endpoint.
- E. The key for the endpoint.

**Answer:** CE

**Explanation:**

Deploying an Azure Machine Learning model as a web service creates a REST API endpoint. You can send data to this endpoint and receive the prediction returned by the model.  
 You create a web service when you deploy a model to your local environment, Azure Container Instances, Azure Kubernetes Service, or field-programmable gate arrays (FPGA). You retrieve the URI used to access the web service by using the Azure Machine Learning SDK. If authentication is enabled, you can also use the SDK to get the authentication keys or tokens.  
 Example:  
 # URL for the web service  
 scoring\_uri = '<your web service URI>'  
 # If the service is authenticated, set the key or token key = '<your key or token>'  
 Reference:  
<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-consume-web-service>

**NEW QUESTION 115**

- (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:  $\text{ZeroOneLoss}(x,y) = 1$  when  $x \neq y$ ; 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 117**

- (Exam Topic 3)

You use the Azure Machine Learning SDK in a notebook to run an experiment using a script file in an experiment folder. The experiment fails. You need to troubleshoot the failed experiment. What are two possible ways to achieve this goal? Each correct answer presents a complete solution.

- A. Use the `get_metrics()` method of the run object to retrieve the experiment run logs.
- B. Use the `get_details_with_logs()` method of the run object to display the experiment run logs.
- C. View the log files for the experiment run in the experiment folder.
- D. View the logs for the experiment run in Azure Machine Learning studio.
- E. Use the `get_output()` method of the run object to retrieve the experiment run logs.

**Answer:** BD

**Explanation:**

Use `get_details_with_logs()` to fetch the run details and logs created by the run. You can monitor Azure Machine Learning runs and view their logs with the Azure Machine Learning studio. Reference: <https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.steprun> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-monitor-view-training-logs>

**NEW QUESTION 121**

- (Exam Topic 3)

You train and register a model by using the Azure Machine Learning SDK on a local workstation. Python 3.6 and Visual Studio Code are installed on the workstation. When you try to deploy the model into production as an Azure Kubernetes Service (AKS)-based web service, you experience an error in the scoring script that causes deployment to fail. You need to debug the service on the local workstation before deploying the service to production. 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                                                                         |
|-------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| Create an AksWebservice deployment configuration for the service and deploy the model to it                                   |  |
| Install Docker on the workstation                                                                                             |                                                                                     |
| Create a LocalWebservice deployment configuration for the service and deploy the model to it                                  |                                                                                     |
| Debug and modify the scoring script as necessary. Use the <code>reload()</code> method of the service after each modification |                                                                                     |
| Create an AciWebservice deployment configuration for the service and deploy the model to it                                   |                                                                                     |

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

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

Step 1: Install Docker on the workstation

Prerequisites include having a working Docker installation on your local system. Build or download the dockerfile to the compute node.

Step 2: Create an AksWebservice deployment configuration and deploy the model to it

To deploy a model to Azure Kubernetes Service, create a deployment configuration that describes the compute resources needed.

# If deploying to a cluster configured for dev/test, ensure that it was created with enough

# cores and memory to handle this deployment configuration. Note that memory is also used by

# things such as dependencies and AML components.

```
deployment_config = AksWebservice.deploy_configuration(cpu_cores = 1, memory_gb = 1)
```

```
service = Model.deploy(ws, "myservice", [model], inference_config, deployment_config, aks_target) service.wait_for_deployment(show_output = True)
```

```
print(service.state) print(service.get_logs())
```

Step 3: Create a LocalWebservice deployment configuration for the service and deploy the model to it

To deploy locally, modify your code to use LocalWebservice.deploy\_configuration() to create a deployment configuration. Then use Model.deploy() to deploy the service.

Step 4: Debug and modify the scoring script as necessary. Use the reload() method of the service after each modification.

During local testing, you may need to update the score.py file to add logging or attempt to resolve any problems that you've discovered. To reload changes to the score.py file, use reload(). For example, the following code reloads the script for the service, and then sends data to it.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-azure-kubernetes-service> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment-local>

**NEW QUESTION 124**

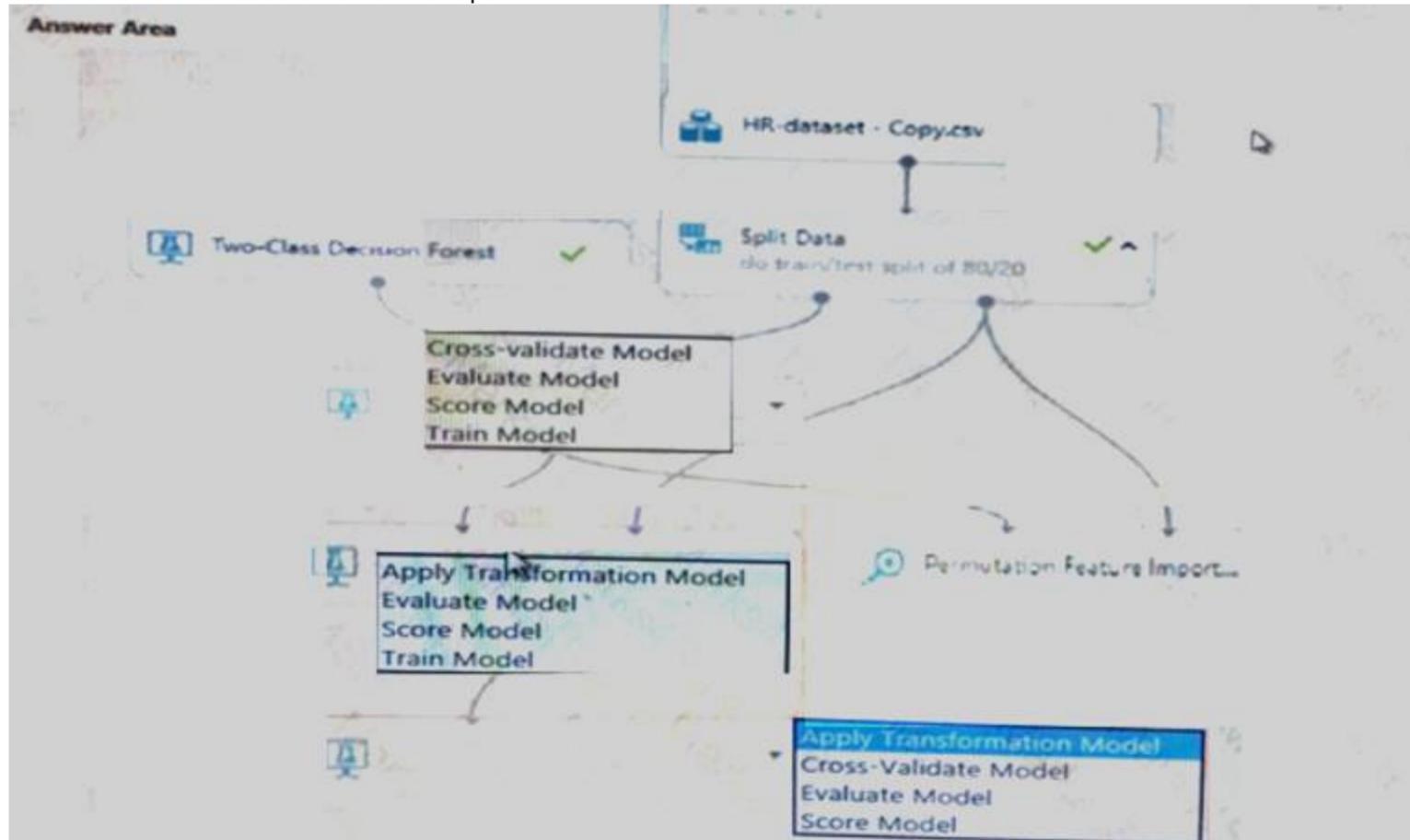
- (Exam Topic 3)

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

You must use a Receiver Operating Characteristic (ROC) curve and an F1 score to evaluate the model. You need to create the required business metrics.

How should you complete the experiment? To answer, select the appropriate options in the dialog box in the answer area.

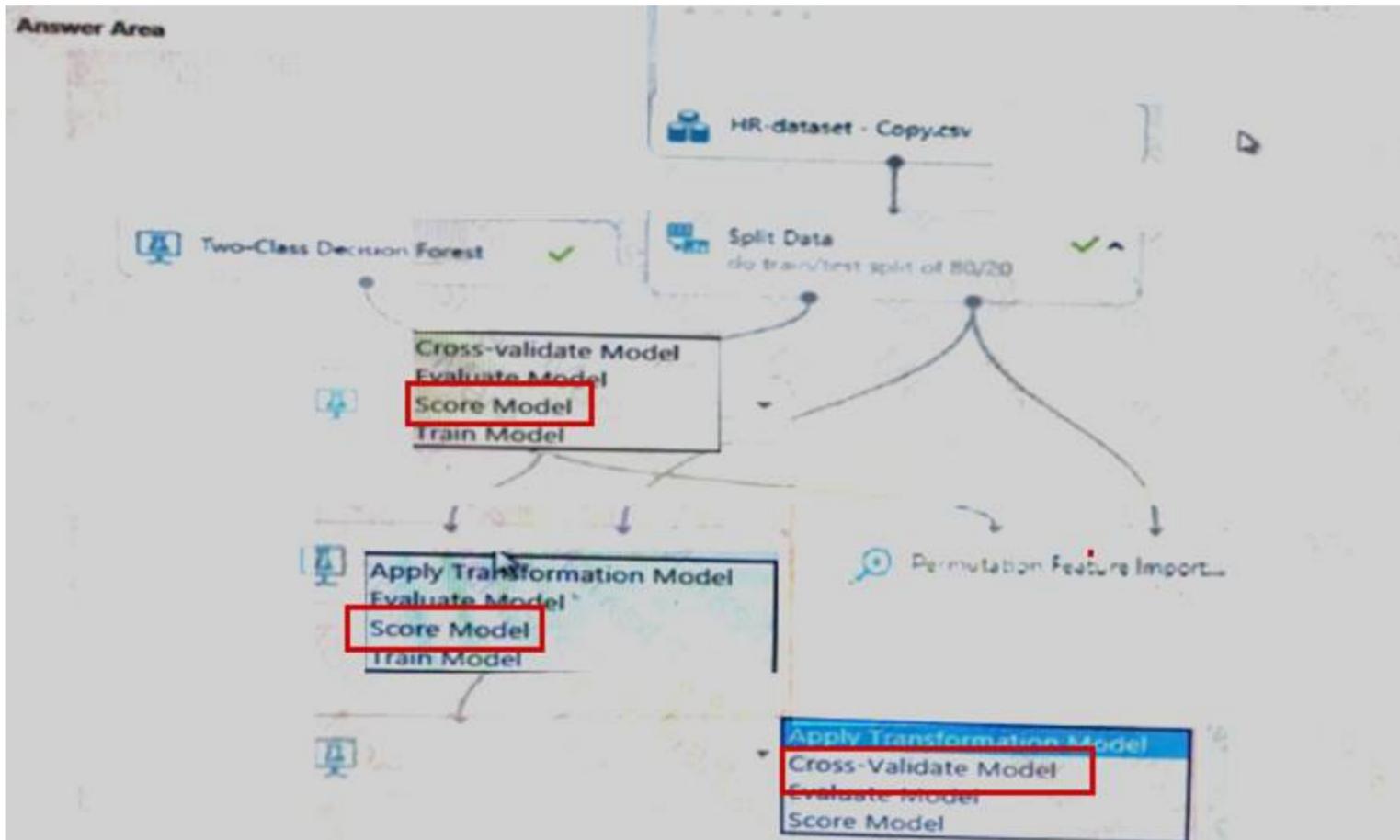
NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**



**NEW QUESTION 128**

- (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 129**

- (Exam Topic 3)

You deploy a model as an Azure Machine Learning real-time web service using the following code.

```
ws, model, inference_config, and deployment_config defined previously
service = Model.deploy(ws, 'classification-service', [model], inference_config, deployment_config)
service.wait_for_deployment(True)
```

The deployment fails.

You need to troubleshoot the deployment failure by determining the actions that were performed during deployment and identifying the specific action that failed. Which code segment should you run?

- A. `service.get_logs()`
- B. `service.state`
- C. `service.serialize()`
- D. `service.update_deployment_state()`

**Answer: A**

**Explanation:**

You can print out detailed Docker engine log messages from the service object. You can view the log for ACI, AKS, and Local deployments. The following example demonstrates how to print the logs.

# if you already have the service object handy `print(service.get_logs())`

# if you only know the name of the service (note there might be multiple services with the same name but different version number)

`print(ws.webservices['mysvc'].get_logs())` Reference:

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

**NEW QUESTION 132**

- (Exam Topic 3)

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

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

NOTE: Each correct selection is worth one point.

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

**Answer: CE**

**Explanation:**

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

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



E: To run Docker, your machine must have a 64-bit operating system running Windows 7 or higher. References:  
[https://docs.docker.com/toolbox/toolbox\\_install\\_windows/](https://docs.docker.com/toolbox/toolbox_install_windows/) <https://blogs.technet.microsoft.com/canitpro/2015/09/08/step-by-step-enabling-hyper-v-for-use-on-windows-10/>

**NEW QUESTION 136**

- (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 140**

- (Exam Topic 3)

You are planning to register a trained model in an Azure Machine Learning workspace.

You must store additional metadata about the model in a key-value format. You must be able to add new metadata and modify or delete metadata after creation.

You need to register the model. Which parameter should you use?

- A. description
- B. model\_framework
- C. cags
- D. properties

**Answer: D**

**Explanation:**

azureml.core.Model.properties:

Dictionary of key value properties for the Model. These properties cannot be changed after registration, however new key value pairs can be added.

Reference:

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

**NEW QUESTION 145**

- (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 148**

- (Exam Topic 3)

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.

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: init()

The entry script has only two required functions, init() and run(data). These functions are used to initialize the service at startup and run the model using request data passed in by a client. The rest of the script handles loading and running the model(s).

Box 2: run() Reference:

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

**NEW QUESTION 151**

- (Exam Topic 3)

You use the Azure Machine Learning SDK to run a training experiment that trains a classification model and calculates its accuracy metric.

The model will be retrained each month as new data is available. You must register the model for use in a batch inference pipeline.

You need to register the model and ensure that the models created by subsequent retraining experiments are registered only if their accuracy is higher than the currently registered model.

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

- A. Specify a different name for the model each time you register it.
- B. Register the model with the same name each time regardless of accuracy, and always use the latest version of the model in the batch inferencing pipeline.
- C. Specify the model framework version when registering the model, and only register subsequent models if this value is higher.
- D. Specify a property named accuracy with the accuracy metric as a value when registering the model, and only register subsequent models if their accuracy is higher than the accuracy property value of the currently registered model.
- E. Specify a tag named accuracy with the accuracy metric as a value when registering the model, and only register subsequent models if their accuracy is higher than the accuracy tag value of the currently registered mode

**Answer:** CE

**Explanation:**

E: Using tags, you can track useful information such as the name and version of the machine learning library used to train the model. Note that tags must be alphanumeric.

Reference:

<https://notebooks.azure.com/xavierheriat/projects/azureml-getting-started/html/how-to-use-azureml/deployment/>

**NEW QUESTION 152**

- (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 154**

- (Exam Topic 3)

You are analyzing a dataset by using Azure Machine Learning Studio.

YOU need to generate a statistical summary that contains the p value and the unique value count for each feature column.  
 Which two modules can you users? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

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

**Answer:** BE

**Explanation:**

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

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

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

References:

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

**NEW QUESTION 159**

- (Exam Topic 3)

You are building recurrent neural network to perform a binary classification.

The training loss, validation loss, training accuracy, and validation accuracy of each training epoch has been provided. You need to identify whether the classification model is over fitted.

Which of the following is correct?

- A. The training loss increases while the validation loss decreases when training the model.
- B. The training loss decreases while the validation loss increases when training the model.
- C. The training loss stays constant and the validation loss decreases when training the model.
- D. 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.

**Answer:** B

**Explanation:**

An overfit model is one where performance on the train set is good and continues to improve, whereas performance on the validation set improves to a point and then begins to degrade.

References:

<https://machinelearningmastery.com/diagnose-overfitting-underfitting-lstm-models/>

**NEW QUESTION 161**

- (Exam Topic 3)

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

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

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

**Question**

**Answer Choise**

Which H value should you select based on the data?

▼

1

2

3

4

5

What H value displays the poorest training result?

▼

1

2

3

4

5

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: 4  
Choose the one which has lower training and validation error and also the closest match. Minimize variance (difference between validation error and train error).  
Box 2: 5

Minimize variance (difference between validation error and train error). Reference:  
<https://medium.com/comet-ml/organizing-machine-learning-projects-project-management-guidelines-2d2b8565>

**NEW QUESTION 166**

- (Exam Topic 3)

Your team is building a data engineering and data science development environment. The environment must support the following requirements:

- > support Python and Scala
- > compose data storage, movement, and processing services into automated data pipelines
- > the same tool should be used for the orchestration of both data engineering and data science
- > support workload isolation and interactive workloads
- > enable scaling across a cluster of machines You need to create the environment.

What should you do?

- A. Build the environment in Apache Hive for HDInsight and use Azure Data Factory for orchestration.
- B. Build the environment in Azure Databricks and use Azure Data Factory for orchestration.
- C. Build the environment in Apache Spark for HDInsight and use Azure Container Instances for orchestration.
- D. Build the environment in Azure Databricks and use Azure Container Instances for orchestration.

**Answer:** B

**Explanation:**

In Azure Databricks, we can create two different types of clusters.

- > Standard, these are the default clusters and can be used with Python, R, Scala and SQL
- > High-concurrency

Azure Databricks is fully integrated with Azure Data Factory.

**NEW QUESTION 168**

- (Exam Topic 3)

You plan to run a Python script as an Azure Machine Learning experiment. The script contains the following code:

```
import os, argparse, glob
from azureml.core import Run
parser = argparse.ArgumentParser()
parser.add_argument('--input-data',
 type=str, dest='data_folder')
args = parser.parse_args()
data_path = args.data_folder
file_paths = glob.glob(data_path + "/*.jpg")
```

You must specify a file dataset as an input to the script. The dataset consists of multiple large image files and must be streamed directly from its source.

You need to write code to define a ScriptRunConfig object for the experiment and pass the ds dataset as an argument.

Which code segment should you use?

- A. arguments = ['--input-data', ds.to\_pandas\_dataframe()]
- B. arguments = ['--input-data', ds.as\_mount()]
- C. arguments = ['--data-data', ds]
- D. arguments = ['--input-data', ds.as\_download()]

**Answer:** A

**Explanation:**

If you have structured data not yet registered as a dataset, create a TabularDataset and use it directly in your training script for your local or remote experiment.

To load the TabularDataset to pandas DataFrame df = dataset.to\_pandas\_dataframe()

Note: TabularDataset represents data in a tabular format created by parsing the provided file or list of files. Reference:

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

**NEW QUESTION 170**

- (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 171**

- (Exam Topic 3)

You configure a Deep Learning Virtual Machine for Windows.

You need to recommend tools and frameworks to perform the following:

- > Build deep neural network (DNN) models
- > Perform interactive data exploration and visualization

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

NOTE: Each correct selection is worth one point.

| Tools                       | Task                                                  | Tool |
|-----------------------------|-------------------------------------------------------|------|
| Vowpal Wabbit               | Build DNN models                                      | Tool |
| PowerBI Desktop             | Enable interactive data exploration and visualization | Tool |
| Azure Data Factory          |                                                       |      |
| Microsoft Cognitive Toolkit |                                                       |      |

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Vowpal Wabbit

Use the Train Vowpal Wabbit Version 8 module in Azure Machine Learning Studio (classic), to create a machine learning model by using Vowpal Wabbit.

Box 2: PowerBI Desktop

Power BI Desktop is a powerful visual data exploration and interactive reporting tool

BI is a name given to a modern approach to business decision making in which users are empowered to find, explore, and share insights from data across the enterprise.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/train-vowpal-wabbit-version-> <https://docs.microsoft.com/en-us/azure/architecture/data-guide/scenarios/interactive-data-exploration>

**NEW QUESTION 176**

- (Exam Topic 3)

```
train_cluster = ComputeTarget(workspace=work_space, name='train-cluster')
estimator = Estimator(source_directory =
 'training-experiment',
 script_params = {'--data-folder': data_source.as_mount(), '--regularization': 0.8},
 compute_target = train_cluster,
 entry_script = 'train.py',
 conda_packages = ['scikit-learn'])
```

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

| Statement                                                                                                                                  | Yes                              | No                    |
|--------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------|-----------------------|
| The estimator will look for the files it needs to run an experiment in the training-experiment directory of the local compute environment. | <input checked="" type="radio"/> | <input type="radio"/> |
| The estimator will mount the local data-folder folder and make it available to the script through a parameter.                             | <input type="radio"/>            | <input type="radio"/> |
| The train.py script file will be created if it does not exist.                                                                             | <input type="radio"/>            | <input type="radio"/> |

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Answer Area

The estimator will look for the files it needs to run an experiment in the training-experiment directory of the local compute environment.

The estimator will mount the local data-folder folder and make it available to the script through a parameter.

The train.py script file will be created if it does not exist.

| Yes                                 | No                                  |
|-------------------------------------|-------------------------------------|
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> |

**NEW QUESTION 181**

- (Exam Topic 3)

```
source_directory=scripts_folder,
entry_script="batch_pipeline.py",
mini_batch_size="5",
error_threshold=10,
output_action="append_row",
environment=batch_env,
compute_target=compute_target,
logging_level="DEBUG",
node_count=4)
```

You need to obtain the output from the pipeline execution. Where will you find the output?

- A. the Activity Log in the Azure portal for the Machine Learning workspace
- B. a file named parallel\_run\_step.txt located in the output folder
- C. the digitidentification.py script
- D. the Inference Clusters tab in Machine Learning studio
- E. the debug log

**Answer: B**

**Explanation:**

output\_action (str): How the output is to be organized. Currently supported values are 'append\_row' and 'summary\_only'.

> 'append\_row' – All values output by run() method invocations will be aggregated into one unique file named parallel\_run\_step.txt that is created in the output location.

> 'summary\_only' Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-contrib-pipeline-steps/azureml.contrib.pipeline.steps.parall>

**NEW QUESTION 185**

- (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 186**

- (Exam Topic 3)

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

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

You are a data scientist using Azure Machine Learning Studio.

You need to normalize values to produce an output column into bins to predict a target column. Solution: Apply an Equal Width with Custom Start and Stop binning mode.

Does the solution meet the goal?

- A. Yes
- B. No

**Answer: B**

**Explanation:**

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

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

**NEW QUESTION 189**

- (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:

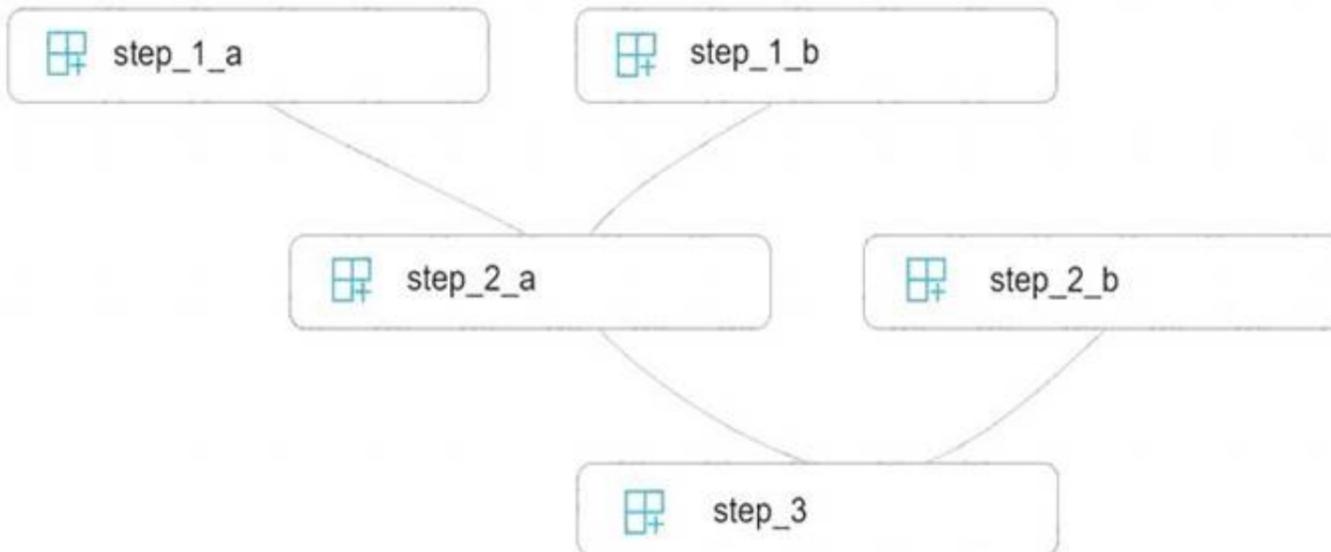
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins>

**NEW QUESTION 193**

- (Exam Topic 3)

You write five Python scripts that must be processed in the order specified in Exhibit A – which allows the same modules to run in parallel, but will wait for modules with dependencies.

You must create an Azure Machine Learning pipeline using the Python SDK, because you want to script to create the pipeline to be tracked in your version control system. You have created five PythonScriptSteps and have named the variables to match the module names.



You need to create the pipeline shown. Assume all relevant imports have been done. Which Python code segment should you use?

- A. `p = Pipeline(ws, steps=[[[[step_1_a, step_1_b], step_2_a], step_2_b], step_3])`
- B. 

```

pipeline_steps = {
 "Pipeline": {
 "run": step_3,
 "run_after": [{
 {"run": step_2_a,
 "run_after": [
 {"run": step_1_a},
 {"run": step_1_b}
]
 },
 {"run": step_2_b}
]
 }
}
p = Pipeline(ws, steps=pipeline_steps)

```
- C. 

```

step_2_a.run_after(step_1_b)
step_2_a.run_after(step_1_a)
step_3.run_after(step_2_b)
step_3.run_after(step_2_a)
p = Pipeline(ws, steps=[step_3])

```
- D. `p = Pipeline(ws, steps=[step_1_a, step_1_b, step_2_a, step_2_b, step_3])`

- A. Option A
- B. Option B
- C. Option C
- D. Option D

**Answer: A**

**Explanation:**

The steps parameter is an array of steps. To build pipelines that have multiple steps, place the steps in order in this array.

Reference:

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

**NEW QUESTION 198**

- (Exam Topic 3)

You train a classification model by using a decision tree algorithm.

You create an estimator by running the following Python code. The variable `feature_names` is a list of all feature names, and `class_names` is a list of all class names.

```
from interpret.ext.blackbox import TabularExplainer
```

```
explainer = TabularExplainer(model,
 x_train,
 features=feature_names,
 classes=class_names)
```

You need to explain the predictions made by the model for all classes by determining the importance of all features.

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 SHAP TreeExplainer will be used to interpret the model.                                                                   | <input type="radio"/> | <input type="radio"/> |
| If you omit the features and classes parameters in the TabularExplainer instantiation, the explainer still works as expected. | <input type="radio"/> | <input type="radio"/> |
| You could interpret the model by using a MimicExplainer instead of a TabularExplainer.                                        | <input type="radio"/> | <input type="radio"/> |

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

and visualizations more informative, you can choose to pass in feature names and output class names if doing classification.

Box 3: No

TabularExplainer automatically selects the most appropriate one for your use case, but you can call each of its three underlying explainers underneath (TreeExplainer, DeepExplainer, or KernelExplainer) directly.

Reference:

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

**NEW QUESTION 203**

- (Exam Topic 3)

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.

| Compute types                                                                                                                                                                                                                                                          | Answer Area                                                                  |                                                                                     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| <div style="border: 1px solid gray; padding: 2px; margin-bottom: 2px;">Attached compute</div> <div style="border: 1px solid gray; padding: 2px; margin-bottom: 2px;">Inference cluster</div> <div style="border: 1px solid gray; padding: 2px;">Training cluster</div> | Requirement                                                                  | Compute type                                                                        |
|                                                                                                                                                                                                                                                                        | Train models by using the Azure Machine Learning designer.                   | <div style="border: 1px solid gray; padding: 2px; width: 100px;">Compute type</div> |
|                                                                                                                                                                                                                                                                        | Score new data through a trained model published as a real-time web service. | <div style="border: 1px solid gray; padding: 2px; width: 100px;">Compute type</div> |
|                                                                                                                                                                                                                                                                        | Train models by using an Azure Databricks cluster.                           | <div style="border: 1px solid gray; padding: 2px; width: 100px;">Compute type</div> |
|                                                                                                                                                                                                                                                                        | Deploy models by using the Azure Machine Learning designer.                  | <div style="border: 1px solid gray; padding: 2px; width: 100px;">Compute type</div> |

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Attached compute

| Training targets                               | Automated ML                | ML pipelines | Azure Machine Learning designer |
|------------------------------------------------|-----------------------------|--------------|---------------------------------|
| Local computer                                 | yes                         |              |                                 |
| Azure Machine Learning compute cluster         | yes & hyperparameter tuning | yes          | yes                             |
| <b>Azure Machine Learning compute instance</b> | yes & hyperparameter tuning | yes          | <b>yes</b>                      |

Box 2: Inference cluster Box 3: Training cluster Box 4: Attached compute

**NEW QUESTION 205**

- (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 207**

- (Exam Topic 3)

You are performing feature engineering on a dataset.

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

You need to add the new feature to the dataset.

Which Azure Machine Learning Studio module should you use?

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

**Answer: A**

**Explanation:**

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

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

**NEW QUESTION 212**

- (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. Split Data
- B. Load Trained Model
- C. Assign Data to Clusters
- D. Group Data into Bins

**Answer: D**

**Explanation:**

The Group Data into Bins module supports multiple options for binning data. You can customize how the bin edges are set and how values are apportioned into the bins.

References:

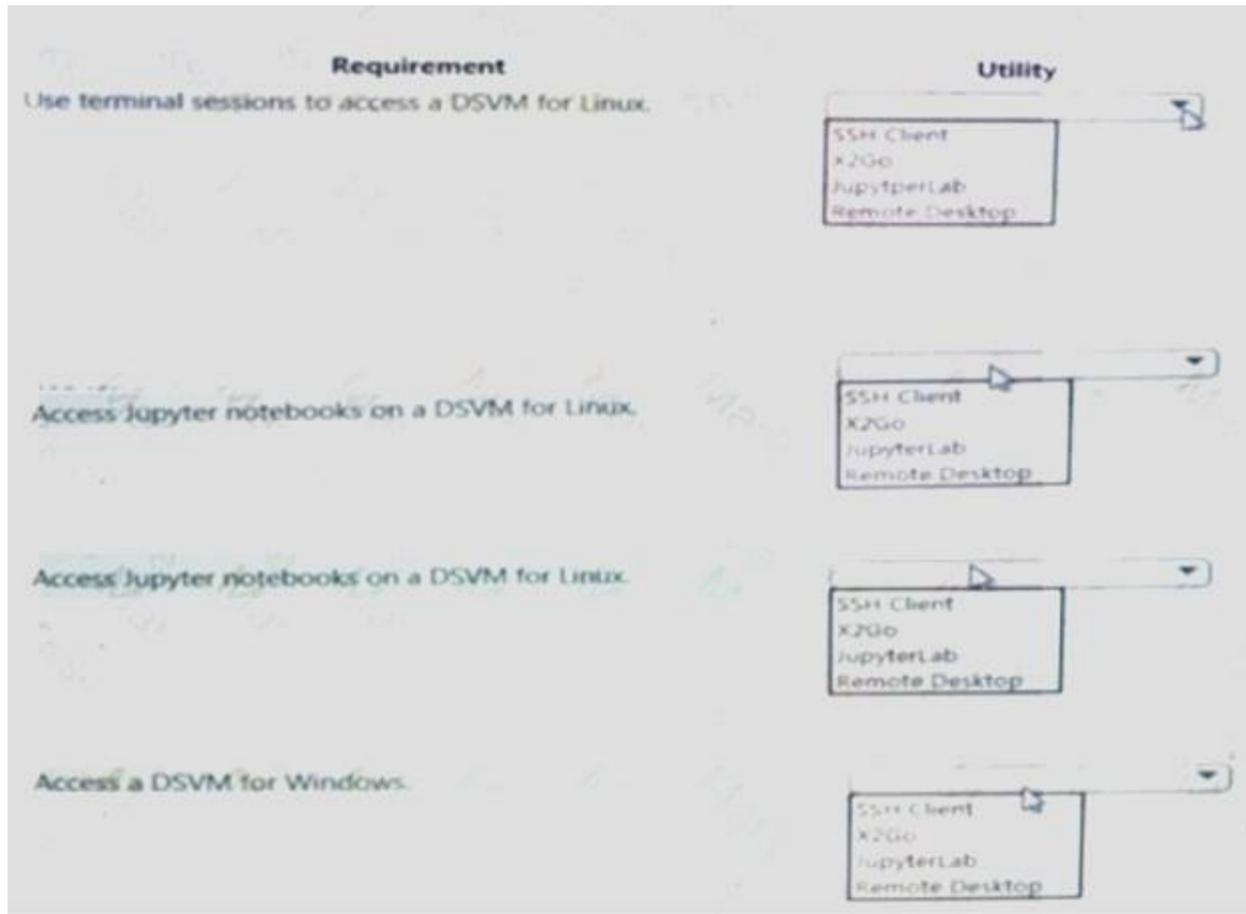
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins>

**NEW QUESTION 215**

- (Exam Topic 3)

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

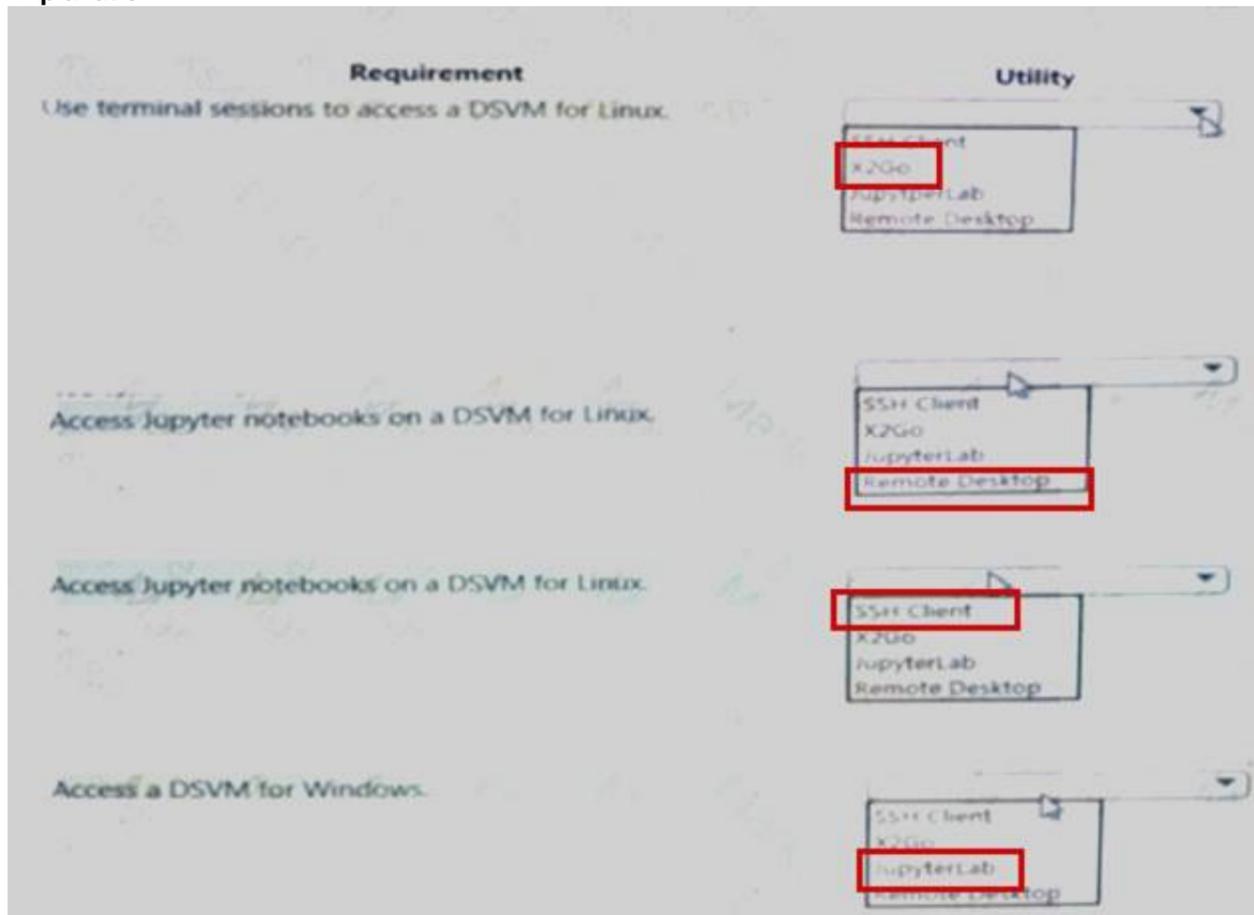
Which utilities should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**



**NEW QUESTION 217**

- (Exam Topic 3)

You are creating a machine learning model. You have a dataset that contains null rows.

You need to use the Clean Missing Data module in Azure Machine Learning Studio to identify and resolve the null and missing data in the dataset.

Which parameter should you use?

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

**Answer: B**

**Explanation:**

Remove entire row: Completely removes any row in the dataset that has one or more missing values. This is useful if the missing value can be considered randomly missing.

References:

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

**NEW QUESTION 220**

- (Exam Topic 3)

You create a binary classification model to predict whether a person has a disease. You need to detect possible classification errors. Which error type should you choose for each description? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

| Description                                                                           | Error type                                                                                                                                                                                                                                     |
|---------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A person has a disease. The model classifies the case as having a disease.            | <div style="border: 1px solid black; padding: 5px;"> <div style="text-align: right;">▼</div> <ul style="list-style-type: none"> <li>True Positives</li> <li>True Negatives</li> <li>False Positives</li> <li>False Negatives</li> </ul> </div> |
| A person does not have a disease. The model classifies the case as having no disease. | <div style="border: 1px solid black; padding: 5px;"> <div style="text-align: right;">▼</div> <ul style="list-style-type: none"> <li>True Positives</li> <li>True Negatives</li> <li>False Positives</li> <li>False Negatives</li> </ul> </div> |
| A person does not have a disease. The model classifies the case as having a disease.  | <div style="border: 1px solid black; padding: 5px;"> <div style="text-align: right;">▼</div> <ul style="list-style-type: none"> <li>True Positives</li> <li>True Negatives</li> <li>False Positives</li> <li>False Negatives</li> </ul> </div> |
| A person has a disease. The model classifies the case as having no disease.           | <div style="border: 1px solid black; padding: 5px;"> <div style="text-align: right;">▼</div> <ul style="list-style-type: none"> <li>True Positives</li> <li>True Negatives</li> <li>False Positives</li> <li>False Negatives</li> </ul> </div> |

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: True Positive

A true positive is an outcome where the model correctly predicts the positive class Box 2: True Negative

A true negative is an outcome where the model correctly predicts the negative class. Box 3: False Positive

A false positive is an outcome where the model incorrectly predicts the positive class. Box 4: False Negative

A false negative is an outcome where the model incorrectly predicts the negative class. Note: Let's make the following definitions:

"Wolf" is a positive class. "No wolf" is a negative class.

We can summarize our "wolf-prediction" model using a 2x2 confusion matrix that depicts all four possible outcomes:

Reference:

<https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative>

**NEW QUESTION 221**

- (Exam Topic 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?

Each 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.

**Answer: CE**

**Explanation:**

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>

**NEW QUESTION 222**

- (Exam Topic 3)

You plan to run a script as an experiment using a Script Run Configuration. The script uses modules from the scipy library as well as several Python packages that are not typically installed in a default conda environment

You plan to run the experiment on your local workstation for small datasets and scale out the experiment by running it on more powerful remote compute clusters for larger datasets.

You need to ensure that the experiment runs successfully on local and remote compute with the least administrative effort.

What should you do?

- A. Create and register an Environment that includes the required package
- B. Use this Environment for all experiment runs.
- C. Always run the experiment with an Estimator by using the default packages.
- D. Do not specify an environment in the run configuration for the experimen
- E. Run the experiment by using the default environment.
- F. Create a config.yaml file defining the conda packages that are required and save the file in the experiment folder.
- G. Create a virtual machine (VM) with the required Python configuration and attach the VM as a compute targe
- H. Use this compute target for all experiment runs.

**Answer:** A

**Explanation:**

If you have an existing Conda environment on your local computer, then you can use the service to create an environment object. By using this strategy, you can reuse your local interactive environment on remote runs.

Reference:

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

**NEW QUESTION 225**

- (Exam Topic 3)

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

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

You are analyzing a numerical dataset which contain 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: Use the last Observation Carried Forward (IOCF) method to impute the missing data points. Does the solution meet the goal?

- A. Yes
- B. No

**Answer:** B

**Explanation:**

Instead use the Multiple Imputation by Chained Equations (MICE) method.

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: Last observation carried forward (LOCF) is a method of imputing missing data in longitudinal studies. If a person drops out of a study before it ends, then his or her last observed score on the dependent variable is used for all subsequent (i.e., missing) observation points. LOCF is used to maintain the sample size and to reduce the bias caused by the attrition of participants in a study.

References:

<https://methods.sagepub.com/reference/encyc-of-research-design/n211.xml> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/>

**NEW QUESTION 229**

- (Exam Topic 3)

You are a data scientist creating a linear regression model.

You need to determine how closely the data fits the regression line. Which metric should you review?

- A. Coefficient of determination
- B. Recall
- C. Precision
- D. Mean absolute error
- E. Root Mean Square Error

**Answer:** A

**Explanation:**

Coefficient of determination, often referred to as R<sup>2</sup>, 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 R<sup>2</sup> values, as low values can be entirely normal and high values can be suspect.

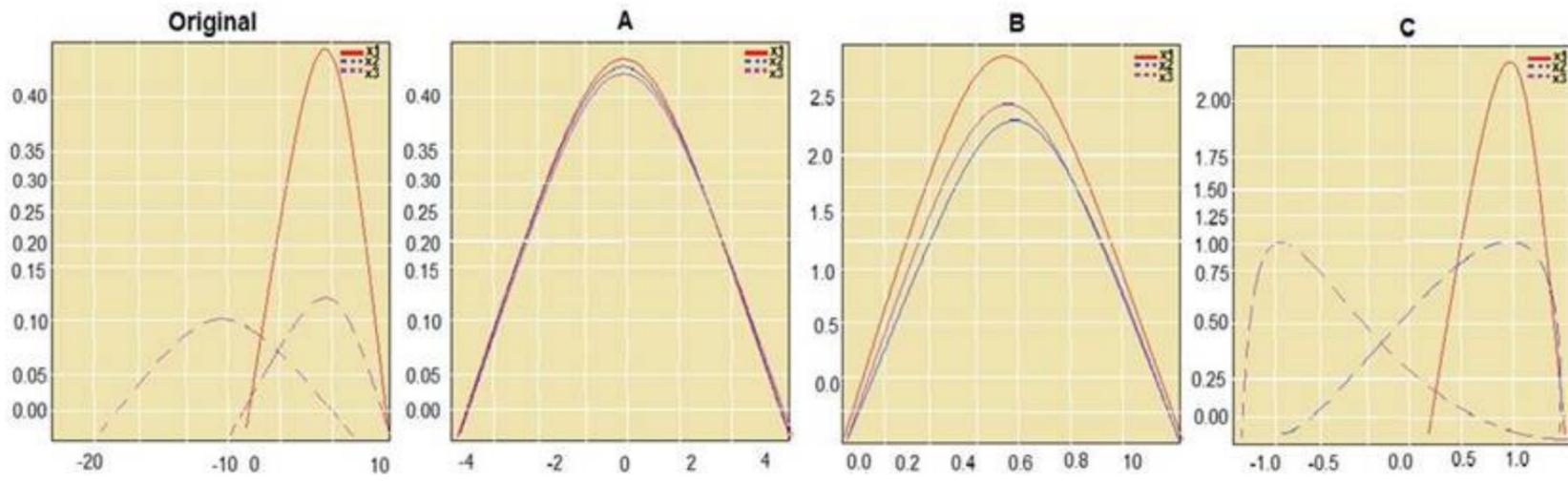
References:

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

**NEW QUESTION 231**

- (Exam Topic 3)

You are performing feature scaling by using the scikit-learn Python library for x<sub>1</sub>, x<sub>2</sub>, and x<sub>3</sub> features. Original and scaled data is shown in the following image.



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

**Question**

**Answer choice**

Which scaler is used in graph A?

▼  
 Standard Scaler  
 Min Max Scale  
 Normalizer

Which scaler is used in graph B?

▼  
 Standard Scaler  
 Min Max Scale  
 Normalizer

Which scaler is used in graph C?

▼  
 Standard Scaler  
 Min Max Scale  
 Normalizer

- A. Mastered
- B. Not Mastered

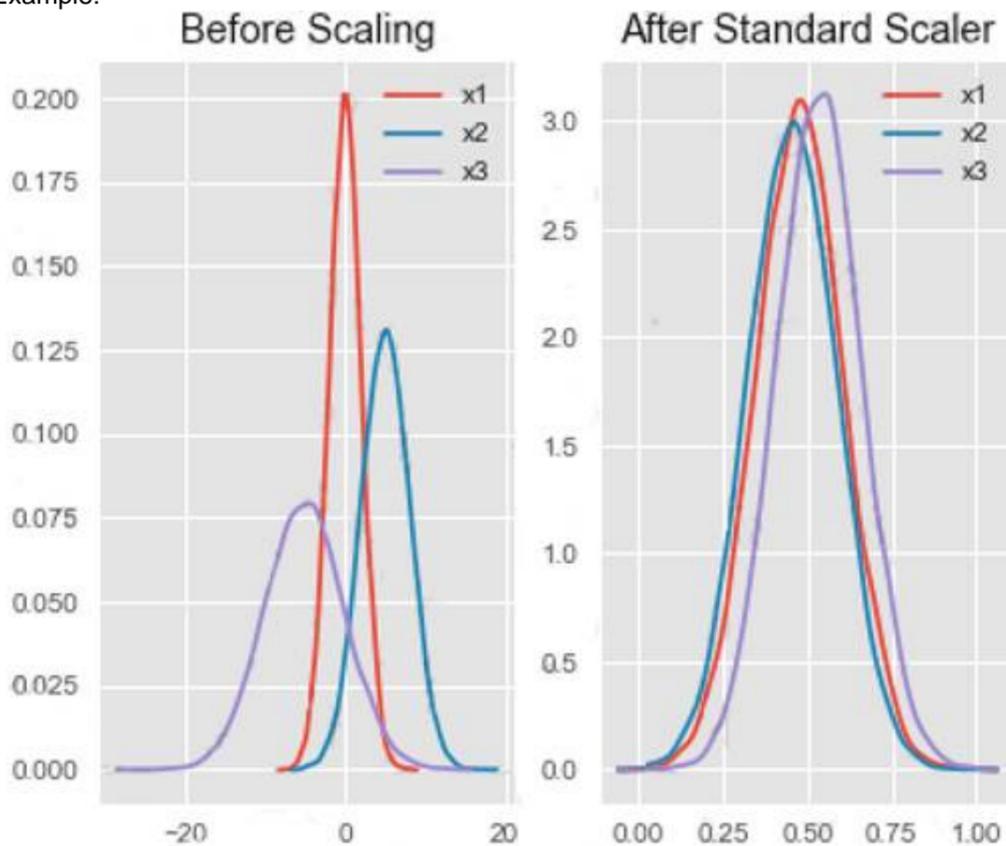
**Answer:** A

**Explanation:**

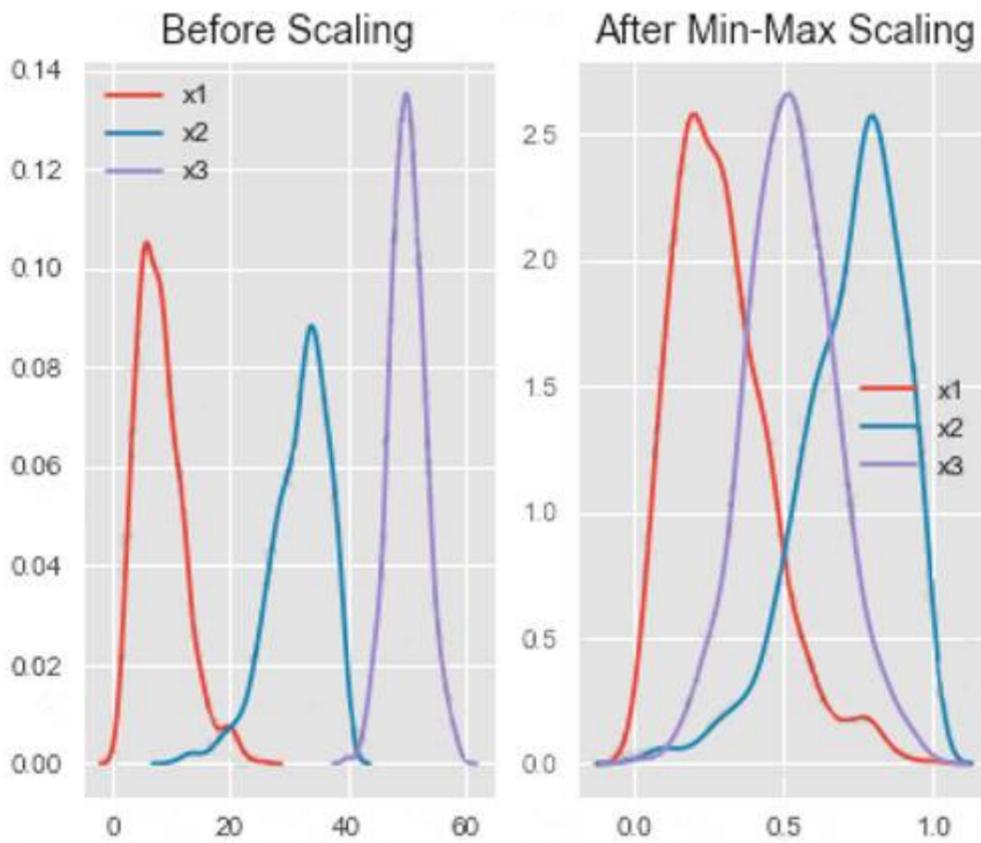
Box 1: StandardScaler

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

Example:



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



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

Box 3: Normalizer

References:

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

**NEW QUESTION 234**

- (Exam Topic 3)

You create a multi-class image classification deep learning model. You train the model by using PyTorch version 1.2. You need to ensure that the correct version of PyTorch can be identified for the inferencing environment when the model is deployed. What should you do?

- A. Save the model locally as a .pt file, and deploy the model as a local web service.
- B. Deploy the model on computer that is configured to use the default Azure Machine Learning conda environment.
- C. Register the model with a .pt file extension and the default version property.
- D. Register the model, specifying the model\_framework and model\_framework\_version properties.

**Answer:** D

**Explanation:**

framework\_version: The PyTorch version to be used for executing training code. Reference: <https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.dnn.pytorch?view=azure-ml-py>

**NEW QUESTION 236**

- (Exam Topic 3)

You run a script as an experiment in Azure Machine Learning. You have a Run object named run that references the experiment run. You must review the log files that were generated during the experiment run. You need to download the log files to a local folder for review. Which two code segments can you run to achieve this goal? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. run.get\_details()
- B. run.get\_file\_names()
- C. run.get\_metrics()
- D. run.download\_files(output\_directory='./runfiles')
- E. run.get\_all\_logs(destination='./runlogs')

**Answer:** AE

**Explanation:**

The run Class get\_all\_logs method downloads all logs for the run to a directory. The run Class get\_details gets the definition, status information, current log files, and other details of the run. Reference: [https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run\(class\)](https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run(class))

**NEW QUESTION 238**

- (Exam Topic 3)

An organization creates and deploys a multi-class image classification deep learning model that uses a set of labeled photographs. The software engineering team reports there is a heavy inferencing load for the prediction web services during the summer. The production web service for the model fails to meet demand despite having a fully-utilized compute cluster where the web service is deployed. You need to improve performance of the image classification web service with minimal downtime and minimal administrative effort. What should you advise the IT Operations team to do?

- A. Increase the minimum node count of the compute cluster where the web service is deployed.
- B. Create a new compute cluster by using larger VM sizes for the nodes, redeploy the web service to that cluster, and update the DNS registration for the service endpoint to point to the new cluster.

- C. Increase the VM size of nodes in the compute cluster where the web service is deployed.
- D. Increase the node count of the compute cluster where the web service is deployed.

**Answer:** D

**Explanation:**

The Azure Machine Learning SDK does not provide support scaling an AKS cluster. To scale the nodes in the cluster, use the UI for your AKS cluster in the Azure Machine Learning studio. You can only change the node count, not the VM size of the cluster.

Reference:

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

**NEW QUESTION 240**

- (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 ParallelRunStep 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 243**

- (Exam Topic 2)

You need to configure the Edit Metadata module so that the structure of the datasets match.

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

NOTE: Each correct selection is worth one point.

Properties      Project

▲ Edit Metadata

Column

Selected columns:

Column names: MedianValue

Launch column selector

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Floating point

Need floating point for Median values.

Scenario: An initial investigation shows that the datasets are identical in structure apart from the MedianValue column. The smaller Paris dataset contains the MedianValue in text format, whereas the larger London dataset contains the MedianValue in numerical format.

Box 2: Unchanged

Note: Select the Categorical option to specify that the values in the selected columns should be treated as categories.

For example, you might have a column that contains the numbers 0,1 and 2, but know that the numbers actually mean "Smoker", "Non smoker" and "Unknown". In that case, by flagging the column as categorical you can ensure that the values are not used in numeric calculations, only to group data.

**NEW QUESTION 247**

- (Exam Topic 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

Stratification key column

Selected columns:  
 Column names: NextToRiver

Launch column selector

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Scenario: Testing

You must produce multiple partitions of a dataset based on sampling using the Partition and Sample module in Azure Machine Learning Studio.

Box 1: Assign to folds

Use Assign to folds 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.

Not Head: Use Head mode to get only the first n rows. This option is useful if you want to test a pipeline on a small number of rows, and don't need the data to be balanced or sampled in any way.

Not Sampling: The Sampling option supports simple random sampling or stratified random sampling. This is useful if you want to create a smaller representative sample dataset for testing.

Box 2: Partition evenly

Specify the partitioner method: Indicate how you want data to be apportioned to each partition, using these options:

> Partition evenly: Use this option to place an equal number of rows in each partition. To specify the number of output partitions, type a whole number in the Specify number of folds to split evenly into text box.

Reference:

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

**NEW QUESTION 252**

- (Exam Topic 1)

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

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

Answer: D

**NEW QUESTION 255**

- (Exam Topic 1)

You need to implement a model development strategy to determine a user's tendency to respond to an ad. Which technique should you use?

- A. Use a Relative Expression Split module to partition the data based on centroid distance.
- B. Use a Relative Expression Split module to partition the data based on distance travelled to the event.
- C. Use a Split Rows module to partition the data based on distance travelled to the event.
- D. Use a Split Rows module to partition the data based on centroid distance.

Answer: A

**Explanation:**

Split Data partitions the rows of a dataset into two distinct sets.

The Relative Expression Split option in the Split Data module of Azure Machine Learning Studio is helpful when you need to divide a dataset into training and testing datasets using a numerical expression.

Relative Expression Split: Use this option whenever you want to apply a condition to a number column. The number could be a date/time field, a column containing age or dollar amounts, or even a percentage. For example, you might want to divide your data set depending on the cost of the items, group people by age ranges, or separate data by a calendar date.

Scenario:

Local market segmentation models will be applied before determining a user's propensity to respond to an advertisement.

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

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

**NEW QUESTION 257**

- (Exam Topic 1)

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

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

| Actions                                                                            |   | Answer Area |
|------------------------------------------------------------------------------------|---|-------------|
| Add new features for retraining supervised models.                                 |   |             |
| Filter labeled cases for retraining using the shortest distance from centroids.    |   |             |
| Evaluate the changes in correlation between model error rate and centroid distance | ⬅ | ⬆           |
| Impute unavailable features with centroid aligned models                           | ➡ | ⬇           |
| Filter labeled cases for retraining using the longest distance from centroids.     |   |             |
| Remove features before retraining supervised models.                               |   |             |

- A. Mastered
- B. Not Mastered

Answer: A

**Explanation:**

Scenario:

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.

Note: Evaluate the changed in correlation between model error rate and centroid distance

In machine learning, a nearest centroid classifier or nearest prototype classifier is a classification model that assigns to observations the label of the class of training samples whose mean (centroid) is closest to the observation.

References: [https://en.wikipedia.org/wiki/Nearest\\_centroid\\_classifier](https://en.wikipedia.org/wiki/Nearest_centroid_classifier)

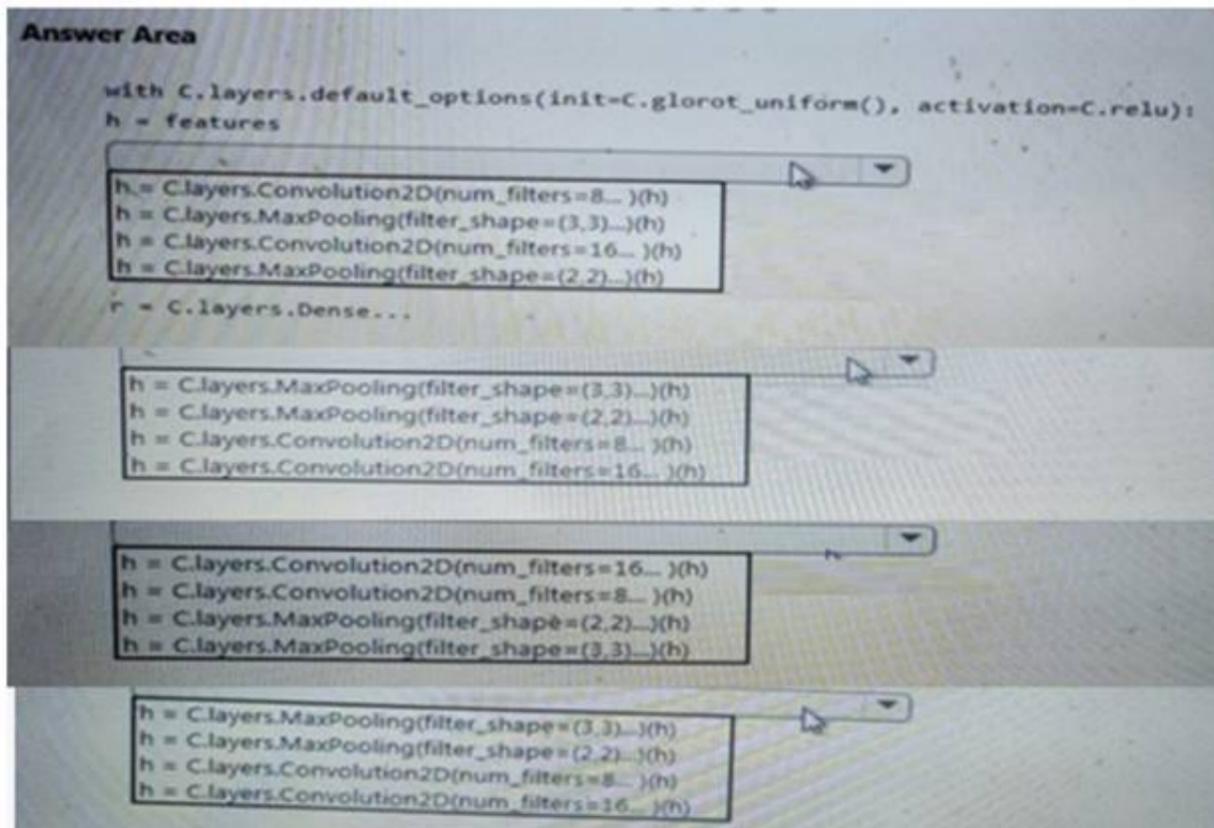
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/sweep-clustering>

**NEW QUESTION 258**

- (Exam Topic 1)

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

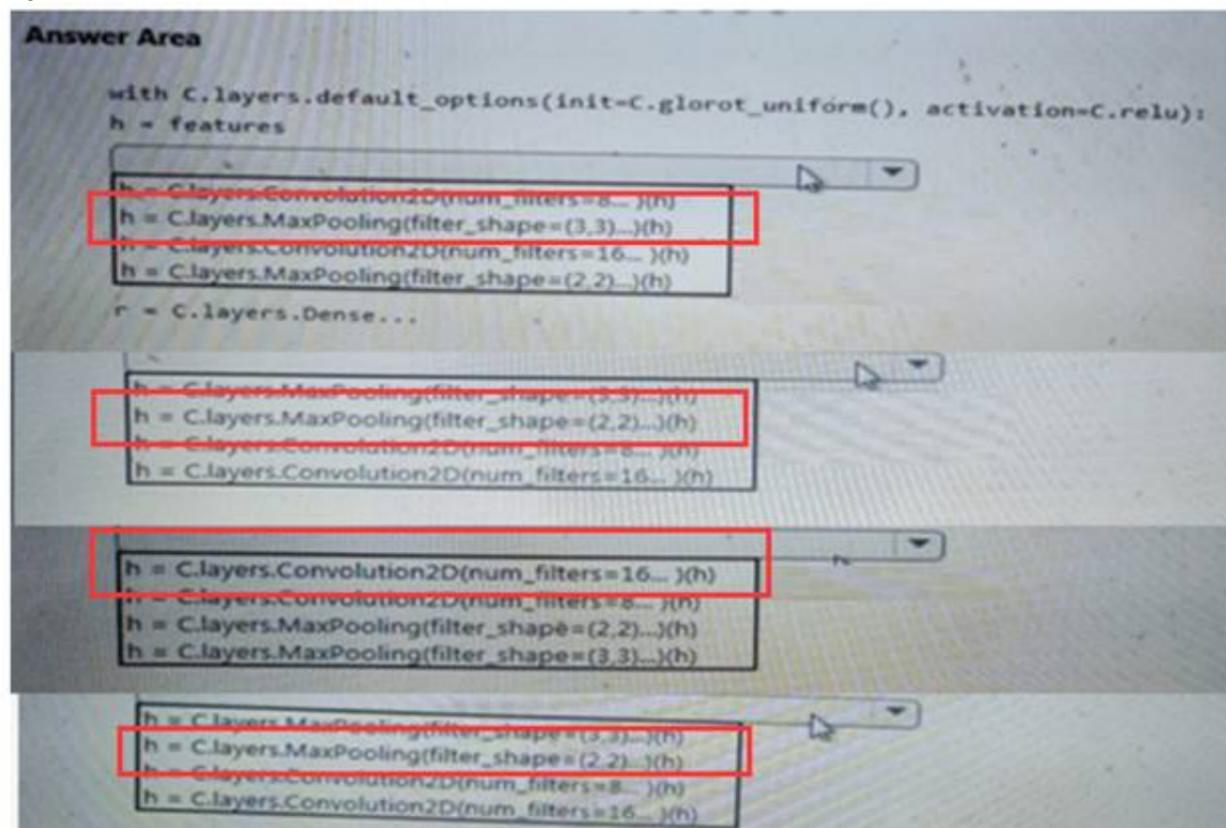
How should you complete the code segment? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



**NEW QUESTION 259**

- (Exam Topic 1)

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

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

Answer: C

Explanation:

Post batch normalization statistics (PBN) is the Microsoft Cognitive Toolkit (CNTK) version of how to evaluate the population mean and variance of Batch Normalization which could be used in inference Original Paper.

In CNTK, custom networks are defined using the BrainScriptNetworkBuilder and described in the CNTK network description language "BrainScript."

Scenario:

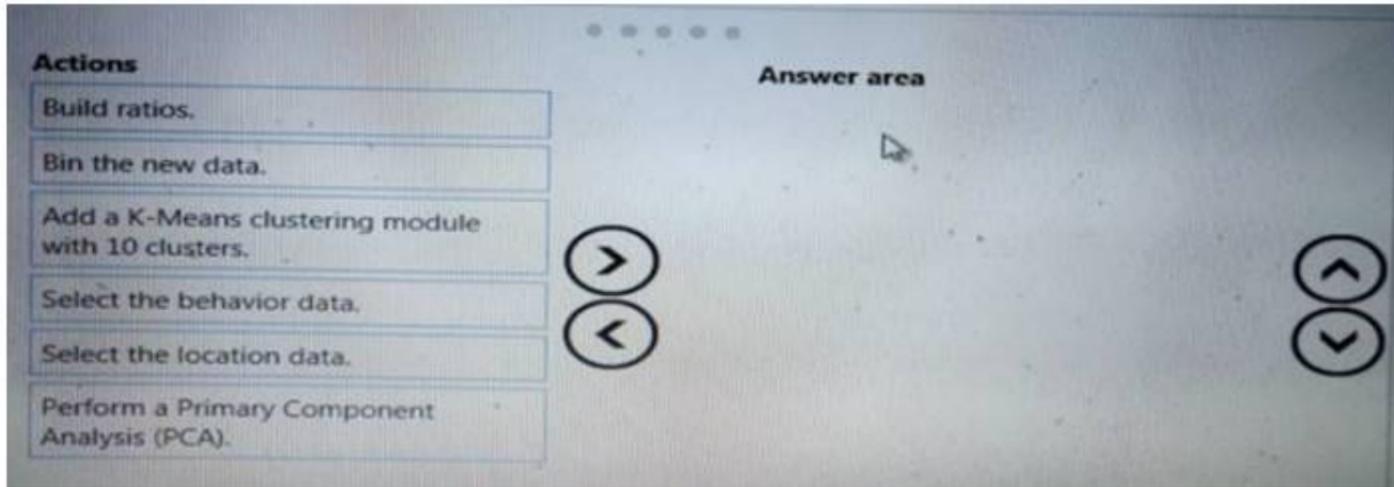
Local penalty detection models must be written by using BrainScript. References:

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

**NEW QUESTION 264**

- (Exam Topic 1)

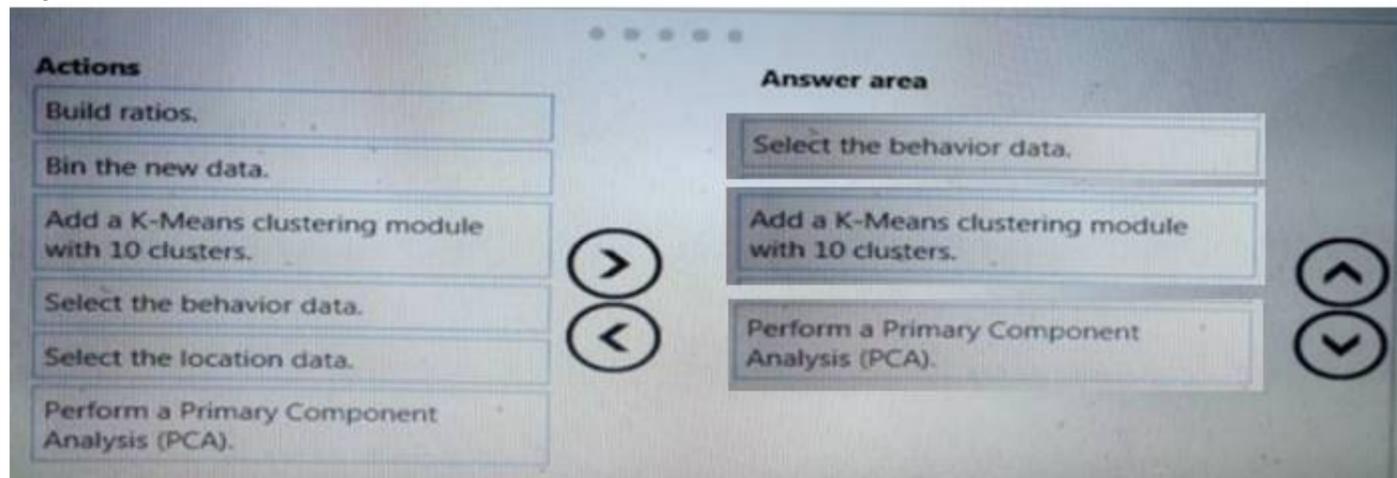
You need to modify the inputs for the global penalty event model to address the bias and variance issue. Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**



**NEW QUESTION 267**

- (Exam Topic 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).

**Answer:** A

**NEW QUESTION 269**

- (Exam Topic 1)

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

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

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

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

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

```
model = deeplearninglib.parallel.Distributed(DataParallel(model))
model = deeplearninglib.nn.parallel.DistributedDataParallelCPU(model)
model = deeplearninglib.keras.Model([
model = deeplearninglib.keras.Sequential([
...
train_sampler.set_epoch(epoch)
for data, target in train_loader:
 data, target = data.to(device), target.to(device)
```

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

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

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

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

Box 3: optimizer = deeplearninglib.train.GradientDescentOptimizer(learning\_rate=0.10)

**NEW QUESTION 274**

.....

## **Thank You for Trying Our Product**

### **We offer two products:**

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

2nd - Questions and Answers in PDF Format

### **DP-100 Practice Exam Features:**

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

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