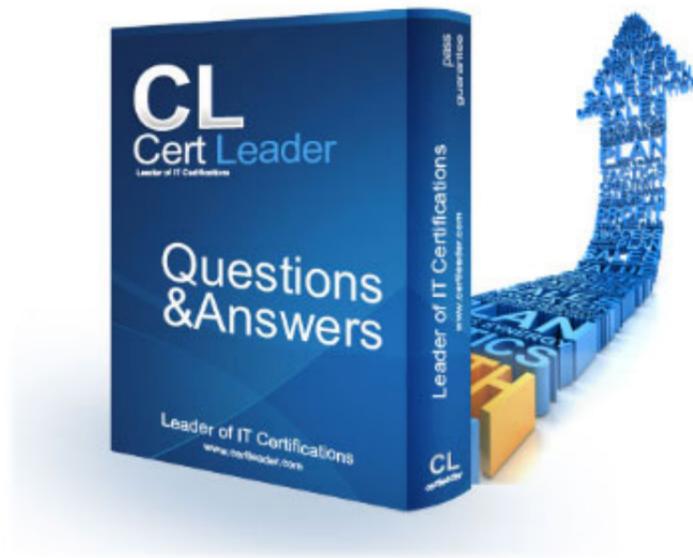


## DP-100 Dumps

# Designing and Implementing a Data Science Solution on Azure

<https://www.certleader.com/DP-100-dumps.html>



**NEW QUESTION 1**

- (Exam Topic 3)

You are using C-Support Vector classification to do a multi-class classification with an unbalanced training dataset. The C-Support Vector classification using Python code shown below:

```
from sklearn.svm import svc
import numpy as np
svc = SVC(kernel= 'linear', class_weight= 'balanced', C=1.0, random_state=0)
modell = svc.fit(X_train, y)
```

You need to evaluate the C-Support Vector classification code.

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

Code Segment	Evaluation Statement
class_weight=balanced	<ul style="list-style-type: none"> <li>Automatically select the performance metrics for the classification.</li> <li>Automatically adjust weights directly proportional to class frequencies in the input data.</li> <li>Automatically adjust weights inversely proportional to class frequencies in the input data.</li> </ul>
C parameter	<ul style="list-style-type: none"> <li>Penalty parameter</li> <li>Degree of polynomial kernel function</li> <li>Size of the kernel cache</li> </ul>

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: Automatically adjust weights inversely proportional to class frequencies in the input data

The "balanced" mode uses the values of y to automatically adjust weights inversely proportional to class frequencies in the input data as  $n\_samples / (n\_classes * np.bincount(y))$ .

Box 2: Penalty parameter

Parameter: C : float, optional (default=1.0) Penalty parameter C of the error term. References:

<https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>

**NEW QUESTION 2**

- (Exam Topic 3)

You have a dataset that includes confidential data. You use the dataset to train a model.

You must use a differential privacy parameter to keep the data of individuals safe and private. You need to reduce the effect of user data on aggregated results.

What should you do?

- A. Decrease the value of the epsilon parameter to reduce the amount of noise added to the data
- B. Increase the value of the epsilon parameter to decrease privacy and increase accuracy
- C. Decrease the value of the epsilon parameter to increase privacy and reduce accuracy
- D. Set the value of the epsilon parameter to 1 to ensure maximum privacy

**Answer: C**

**Explanation:**

Differential privacy tries to protect against the possibility that a user can produce an indefinite number of reports to eventually reveal sensitive data. A value known as epsilon measures how noisy, or private, a report is. Epsilon has an inverse relationship to noise or privacy. The lower the epsilon, the more noisy (and private) the data is.

Reference:

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

**NEW QUESTION 3**

- (Exam Topic 3)

You create a Python script named train.py and save it in a folder named scripts. The script uses the scikit-learn framework to train a machine learning model.

You must run the script as an Azure Machine Learning experiment on your local workstation. You need to write Python code to initiate an experiment that runs the train.py script.

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

NOTE: Each correct selection is worth one point.

### Answer Area

```

from azureml.core import Experiment, ScriptRunConfig, Environment
from azureml.core.conda_dependencies import CondaDependencies
from azureml.core import Workspace

ws = Workspace.from_config()
py_sk = Environment('sklearn-training')
pkgs = CondaDependencies.create(pip_packages=['scikit-learn', 'azureml-defaults'])
py_sk.python.conda_dependencies = pkgs
script_config = ScriptRunConfig (
    source_directory = 'scripts',
    script = 'train.py',
    environment = py_sk)

experiment = Experiment(workspace=ws, name='training-experiment')
run = experiment.submit(config=script_config)

```

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

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

Box 1: source\_directory

source\_directory: A local directory containing code files needed for a run. Box 2: script

Script: The file path relative to the source\_directory of the script to be run. Box 3: environment

Reference:

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

**NEW QUESTION 4**

- (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 5**

- (Exam Topic 3)

You plan to use automated machine learning to train a regression model. You have data that has features which have missing values, and categorical features with few distinct values.

You need to configure automated machine learning to automatically impute missing values and encode categorical features as part of the training task.

Which parameter and value pair should you use in the AutoMLConfig class?

- A. featurization = 'auto'
- B. enable\_voting\_ensemble = True
- C. task = 'classification'
- D. exclude\_nan\_labels = True
- E. enable\_tf = True

**Answer:** A

**Explanation:**

Featurization str or FeaturizationConfig Values: 'auto' / 'off' / FeaturizationConfig

Indicator for whether featurization step should be done automatically or not, or whether customized featurization should be used.

Column type is automatically detected. Based on the detected column type preprocessing/featurization is done as follows:

Categorical: Target encoding, one hot encoding, drop high cardinality categories, impute missing values. Numeric: Impute missing values, cluster distance, weight of evidence.

DateTime: Several features such as day, seconds, minutes, hours etc. Text: Bag of words, pre-trained Word embedding, text target encoding. Reference:

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

**NEW QUESTION 6**

- (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 7**

- (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 8**

- (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 9

- (Exam Topic 3)

You use the Two-Class Neural Network module in Azure Machine Learning Studio to build a binary classification model. You use the Tune Model Hyperparameters module to tune accuracy for the model.

You need to select the hyperparameters that should be tuned using the Tune Model Hyperparameters module. Which two hyperparameters should you use? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Number of hidden nodes
- B. Learning Rate
- C. The type of the normalizer
- D. Number of learning iterations
- E. Hidden layer specification

**Answer:** DE

#### Explanation:

D: For Number of learning iterations, specify the maximum number of times the algorithm should process the training cases.

E: For Hidden layer specification, select the type of network architecture to create.

Between the input and output layers you can insert multiple hidden layers. Most predictive tasks can be accomplished easily with only one or a few hidden layers.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-neural-network>

### NEW QUESTION 10

- (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 10

- (Exam Topic 3)

You create a new Azure subscription. No resources are provisioned in the subscription. You need to create an Azure Machine Learning workspace.

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. Run Python code that uses the Azure ML SDK library and calls the Workspace.create method with name, subscription\_id, resource\_group, and location parameters.
- B. Use an Azure Resource Management template that includes a Microsoft.MachineLearningServices/workspaces resource and its dependencies.
- C. Use the Azure Command Line Interface (CLI) with the Azure Machine Learning extension to call the az group create function with --name and --location parameters, and then the az ml workspace create function, specifying -w and -g parameters for the workspace name and resource group.
- D. Navigate to Azure Machine Learning studio and create a workspace.
- E. Run Python code that uses the Azure ML SDK library and calls the Workspace.get method with name, subscription\_id, and resource\_group parameters.

**Answer:** BCD

#### Explanation:

B: You can use an Azure Resource Manager template to create a workspace for Azure Machine Learning. Example:

```
{"type": "Microsoft.MachineLearningServices/workspaces",
```

...

C: You can create a workspace for Azure Machine Learning with Azure CLI Install the machine learning extension.

Create a resource group: `az group create --name <resource-group-name> --location <location>`

To create a new workspace where the services are automatically created, use the following command: `az ml workspace create -w <workspace-name> -g <resource-group-name>`

D: You can create and manage Azure Machine Learning workspaces in the Azure portal.

- > Sign in to the Azure portal by using the credentials for your Azure subscription.
- > In the upper-left corner of Azure portal, select + Create a resource.
- > Use the search bar to find Machine Learning.
- > Select Machine Learning.
- > In the Machine Learning pane, select Create to begin.

[Home](#) > [New](#) > [Machine Learning](#) >

## Machine Learning

Create a machine learning workspace

[Basics](#) [Networking](#) [Advanced](#) [Tags](#) [Review + create](#)

### Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription \*   

Resource group \*     
[Create new](#)

### Workspace details

Specify the name, region, and edition for the workspace.

Workspace name \*   

Region \*   

Workspace edition \*     
 Basic  
 Basic  
 Enterprise

 For your convenience, these resources are available in your workspace: [Application Insights](#), [Azure Key Vault](#)

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-workspace-template> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace-cli> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace>

### NEW QUESTION 14

- (Exam Topic 3)

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

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

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

**Answer: B**

### Explanation:

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

References:

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

### NEW QUESTION 17

- (Exam Topic 3)

You are moving a large dataset from Azure Machine Learning Studio to a Weka environment. You need to format the data for the Weka environment.

Which module should you use?

- A. Convert to CSV
- B. Convert to Dataset
- C. Convert to ARFF
- D. Convert to SVMLight

**Answer:** C

**Explanation:**

Use the Convert to ARFF module in Azure Machine Learning Studio, to convert datasets and results in Azure Machine Learning to the attribute-relation file format used by the Weka toolset. This format is known as ARFF.

The ARFF data specification for Weka supports multiple machine learning tasks, including data preprocessing, classification, and feature selection. In this format, data is organized by entities and their attributes, and is contained in a single text file.

References:

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

**NEW QUESTION 20**

- (Exam Topic 3)

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

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

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

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

**Answer:** AB

**Explanation:**

The box-plot algorithm can be used to display outliers.

One other way to quickly identify Outliers visually is to create scatter plots. References:

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

**NEW QUESTION 25**

- (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)
```

You plan to use this configuration to run a script that trains a random forest model and then tests it with validation data. The label values for the validation data are stored in a 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:

```
import numpy as np  
from sklearn.metrics import roc_auc_score  
# code to train model omitted  
auc = roc_auc_score(y_test, y_predicted)  
print(np.float(auc))
```

Does the solution meet the goal?

- A. Yes
- B. No

**Answer:** B

**Explanation:**

Use a solution with `logging.info(message)` instead. Note: 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 29**

- (Exam Topic 3)

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

Which visualization should you use?

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

**Answer:** A

**Explanation:**

References:

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

**NEW QUESTION 30**

- (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 35**

- (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.

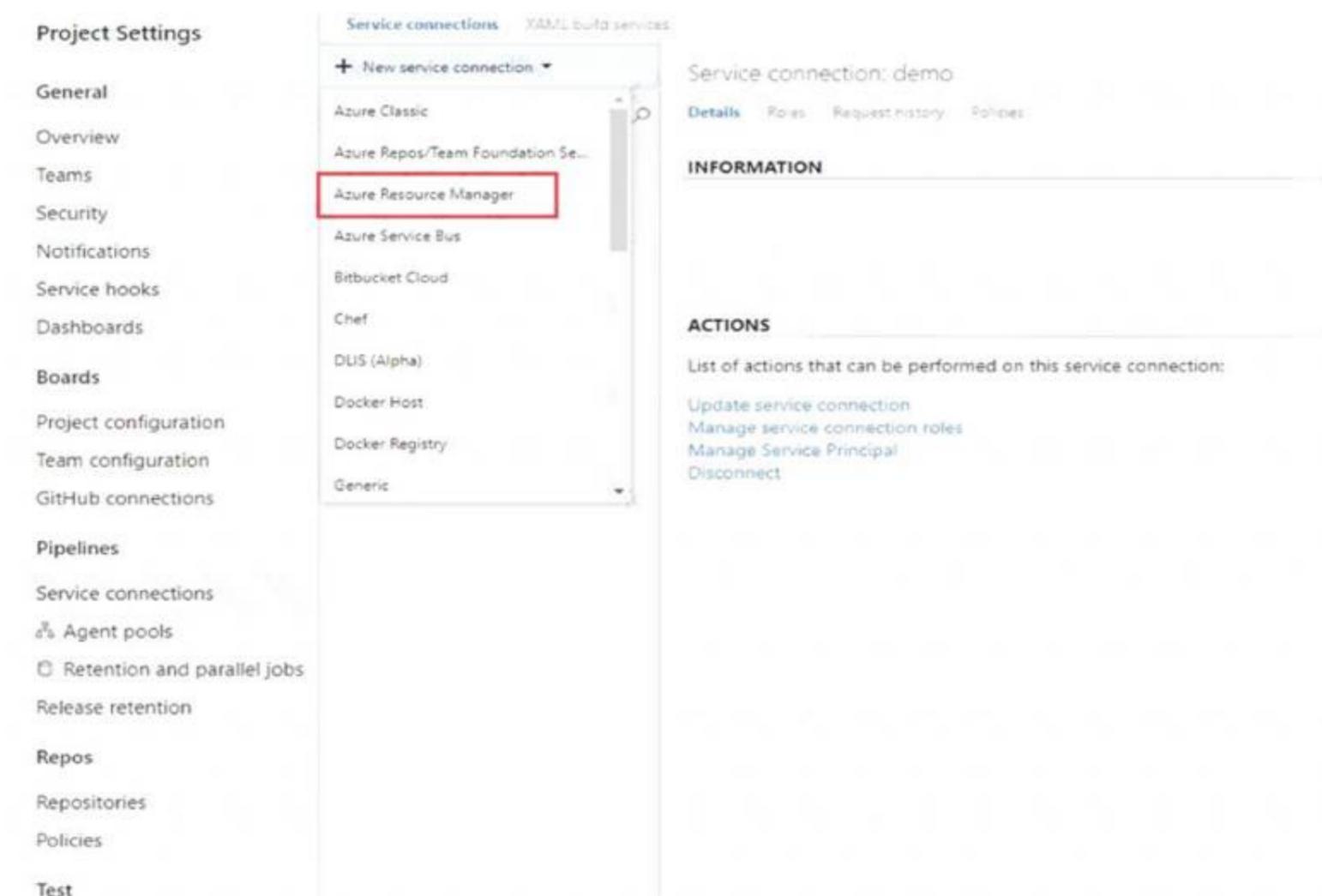
Step 3: Install the Machine Learning extension for Azure Pipelines

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



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 37

- (Exam Topic 3)

You create a datastore named `training_data` that references a blob container in an Azure Storage account. The blob container contains a folder named `csv_files` in which multiple comma-separated values (CSV) files are stored.

You have a script named `train.py` in a local folder named `./script` that you plan to run as an experiment using an estimator. The script includes the following code to read data from the `csv_files` folder:

```
import os
import argparse
import pandas as pd

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from azureml.core import Run

run = Run.get_context()
parser = argparse.ArgumentParser()
parser.add_argument('--data-folder', type=str, dest='data_folder', help='data reference')
args = parser.parse_args()

data_folder = args.data_folder
csv_files = os.listdir(data_folder)
training_data = pd.concat((pd.read_csv(os.path.join(data_folder, csv_file)) for csv_file in csv_files))

# Code goes on to split the training data and train a logistic regression model
```

You have the following script.

```
from azureml.core import Workspace, Datastore, Experiment
from azureml.train.sklearn import SKLearn
```

```
ws = Workspace.from_config()
exp = Experiment(workspace=ws, name='csv_training')
ds = Datastore.get(ws, datastore_name='training_data')
data_ref = ds.path('csv_files')
```

# Code to define estimator goes here

```
run = exp.submit(config=estimator)
run.wait_for_completion(show_output=True)
```

You need to configure the estimator for the experiment so that the script can read the data from a data reference named data\_ref that references the csv\_files folder in the training\_data datastore.

Which code should you use to configure the estimator?

- A. estimator = SKLearn(source\_directory='./script',  
inputs=[data\_ref.as\_named\_input('data-folder').to\_pandas\_dataframe()],  
compute\_target='local',  
entry\_script='train.py')
- B. script\_params = {  
    '--data-folder': data\_ref.as\_mount()  
}  
estimator = SKLearn(source\_directory='./script',  
script\_params=script\_params,  
compute\_target='local',  
entry\_script='train.py')
- C. estimator = SKLearn(source\_directory='./script',  
inputs=[data\_ref.as\_named\_input('data-folder').as\_mount()],  
compute\_target='local',  
entry\_script='train.py')
- D. script\_params = {  
    '--data-folder': data\_ref.as\_download(path\_on\_compute='csv\_files')  
}  
estimator = SKLearn(source\_directory='./script',  
script\_params=script\_params,  
compute\_target='local',  
entry\_script='train.py')
- E. estimator = SKLearn(source\_directory='./script',  
inputs=[data\_ref.as\_named\_input('data-folder').as\_download(path\_on\_compute='csv\_files')],  
compute\_target='local',  
entry\_script='train.py')

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

**Answer:** B

**Explanation:**

Besides passing the dataset through the inputs parameter in the estimator, you can also pass the dataset through script\_params and get the data path (mounting point) in your training script via arguments. This way, you can keep your training script independent of azureml-sdk. In other words, you will be able use the same training script for local debugging and remote training on any cloud platform.

Example:

```
from azureml.train.sklearn import SKLearn
script_params = {
# mount the dataset on the remote compute and pass the mounted path as an argument to the training script '--data-folder':
mnist_ds.as_named_input('mnist').as_mount(),
'--regularization': 0.5
}
est = SKLearn(source_directory=script_folder, script_params=script_params, compute_target=compute_target, environment_definition=env,
entry_script='train_mnist.py')
# Run the experiment
run = experiment.submit(est)
run.wait_for_completion(show_output=True)
Reference:
https://docs.microsoft.com/es-es/azure/machine-learning/how-to-train-with-datasets
```

**NEW QUESTION 41**

- (Exam Topic 3)

You use Azure Machine Learning to train and register a model.

You must deploy the model into production as a real-time web service to an inference cluster named service-compute that the IT department has created in the Azure Machine Learning workspace.

Client applications consuming the deployed web service must be authenticated based on their Azure Active Directory service principal.

You need to write a script that uses the Azure Machine Learning SDK to deploy the model. The necessary modules have been imported.

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

```
# Assume the necessary modules have been imported
deploy_target = (ws, "service-compute")
deployment_config = .deploy_configuration(cpu_cores=1, memory_gb=1,
service = Model.deploy(ws, "ml-service",
    [model], inference_config, deployment_config, deploy_target)
service.wait_for_deployment(show_output = True)
```

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: AksCompute Example:

```
aks_target = AksCompute(ws,"myaks")
```

# 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)
```

Box 2: AksWebservice

Box 3: token\_auth\_enabled=Yes

Whether or not token auth is enabled for the Webservice.

Note: A Service principal defined in Azure Active Directory (Azure AD) can act as a principal on which authentication and authorization policies can be enforced in Azure Databricks.

The Azure Active Directory Authentication Library (ADAL) can be used to programmatically get an Azure AD access token for a user.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-azure-kubernetes-service> <https://docs.microsoft.com/en-us/azure/databricks/dev-tools/api/latest/aad/service-prin-aad-token>

**NEW QUESTION 44**

- (Exam Topic 3)

A coworker registers a datastore in a Machine Learning services workspace by using the following code:

```
Datastore.register_azure_blob_container(workspace=ws,
    datastore_name='demo_datastore',
    container_name='demo_datacontainer',
    account_name='demo_account',
    account_key='0A0A0A-0A0A00A-0A00A0A0A0A0A',
    create_if_not_exists=True)
```

You need to write code to access the datastore from a notebook.

**Answer Area**

```
import azureml.core
from azureml.core import Workspace, Datastore
ws = Workspace.from_config()
datastore = Workspace.get('ws', 'demo_datastore', create_if_not_exists=True)
```

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: DataStore

To get a specific datastore registered in the current workspace, use the get() static method on the Datastore class:

# Get a named datastore from the current workspace

datastore = Datastore.get(ws, datastore\_name='your datastore name') Box 2: ws

Box 3: demo\_datastore Reference:

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

**NEW QUESTION 47**

- (Exam Topic 3)

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

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

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

**Answer: C**

**Explanation:**

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

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

References:

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

**NEW QUESTION 50**

- (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
  ],
  "NotActions": [
    Microsoft.MachineLearningServices/workspaces/*/write
    Microsoft.MachineLearningServices/workspaces/computes/*/write
    Microsoft.MachineLearningServices/workspaces/delete
  ],
  "AssignableScopes": [
    "/subscriptions/<subscription_id>"
  ]
}
```

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

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

- 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 54

- (Exam Topic 3)

You create and register a model in an Azure Machine Learning workspace.

You must use the Azure Machine Learning SDK to implement a batch inference pipeline that uses a ParallelRunStep to score input data using the model. You must specify a value for the ParallelRunConfig compute\_target setting of the pipeline step.

You need to create the compute target. Which class should you use?

- A. BatchCompute
- B. AdlaCompute
- C. AmlCompute
- D. Aks Compute

**Answer: C**

#### Explanation:

Compute target to use for ParallelRunStep. This parameter may be specified as a compute target object or the string name of a compute target in the workspace. The compute\_target target is of AmlCompute or string.

Note: An Azure Machine Learning Compute (AmlCompute) is a managed-compute infrastructure that allows you to easily create a single or multi-node compute. The compute is created within your workspace region as a resource that can be shared with other users

Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-contrib-pipeline-steps/azureml.contrib.pipeline.steps.parall> [https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute\(class\)](https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute(class))

#### NEW QUESTION 57

- (Exam Topic 3)

You plan to run a Python script as an Azure Machine Learning experiment.

The script must read files from a hierarchy of folders. The files will be passed to the script as a dataset argument.

You must specify an appropriate mode for the dataset argument.

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

NOTE: Each correct selection is worth one point.

- A. to\_pandas\_dataframe ()
- B. as\_download()
- C. as\_upload()
- D. as mount ()

**Answer: B**

#### Explanation:

Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.data.filedataset?view=azure-ml-py>

#### NEW QUESTION 58

- (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 59**

- (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 63**

- (Exam Topic 3)

You plan to use Hyperdrive to optimize the hyperparameters selected when training a model. You create the following code to define options for the hyperparameter experiment

```
import azureml.train.hyperdrive.parameter_expressions as pe
from azureml.train.hyperdrive import GridParameterSampling, HyperDriveConfig

param_sampling = GridParameterSampling({
    "max_depth" : pe.choice(6, 7, 8, 9),
    "learning_rate" : pe.choice(0.05, 0.1, 0.15)
})

hyperdrive_run_config = HyperDriveConfig(
    estimator = estimator,
    hyperparameter_sampling = param_sampling,
    policy = None,
    primary_metric_name = "auc",
    primary_metric_goal = PrimaryMetricGoal.MAXIMIZE,

    estimator = estimator,
    hyperparameter_sampling = param_sampling,
    policy = None,
    primary_metric_name = "auc",
    primary_metric_goal = PrimaryMetricGoal.MAXIMIZE,
    max_total_runs = 50,
    max_concurrent_runs = 4)
```

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

**Answer Area**

	Yes	No
There will be 50 runs for this hyperparameter tuning experiment.	<input type="radio"/>	<input type="radio"/>
You can use the policy parameter in the HyperDriveConfig class to specify a security policy.	<input type="radio"/>	<input type="radio"/>
The experiment will create a run for every possible value for the learning rate parameter between 0.05 and 0.15.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: No

max\_total\_runs (50 here)

The maximum total number of runs to create. This is the upper bound; there may be fewer runs when the sample space is smaller than this value.

Box 2: Yes

Policy EarlyTerminationPolicy

The early termination policy to use. If None - the default, no early termination policy will be used. Box 3: No

Discrete hyperparameters are specified as a choice among discrete values. choice can be: one or more comma-separated values

> a range object

> any arbitrary list object Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.hyperdriveconfig> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters>

**NEW QUESTION 67**

- (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 71**

- (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 73**

- (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. 

|                               |
|-------------------------------|
| register_azure_blob_container |
| register_azure_file_share     |
| register_azure_data_lake      |
| register_azure_sql_database   |

 (workspace = ws,

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 77**

- (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 78**

- (Exam Topic 3)

You are building a recurrent neural network to perform a binary classification. You review the training loss, validation loss, training accuracy, and validation accuracy for each training epoch.

You need to analyze model performance.

Which observation indicates that the classification model is over fitted?

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

**Answer:** B

**NEW QUESTION 82**

- (Exam Topic 3)

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

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

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

NOTE: Each correct selection is worth one point.

```
# Assume required modules are imported
```

```
ws = Workspace.from_config()
```

```
feature_importances = explanation.
```

▼
( workspace = ws,

- from\_run
- list\_model\_explanations
- from\_run\_id
- download\_model\_explanation

```
experiment_name='train_and_explain',  
run_id='train_and_explain_12345')
```

```
explanation = client.
```

▼
()

- upload\_model\_explanation
- list\_model\_explanations
- run
- download\_model\_explanation

```
feature_importances = explanation.
```

▼
()

- explanation
- explanation\_client
- get\_feature\_important\_dict
- download\_model\_explanation

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

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: from\_run\_id

from\_run\_id(workspace, experiment\_name, run\_id) Create the client with factory method given a run ID. Returns an instance of the explanations Client.

Parameters

- > Workspace Workspace An object that represents a workspace.
- > experiment\_name str The name of an experiment.
- > run\_id str A GUID that represents a run.

Box 2: list\_model\_explanations

list\_model\_explanations returns a dictionary of metadata for all model explanations available.

Returns

A dictionary of explanation metadata such as id, data type, explanation: method, model type, and upload time, sorted by upload time

Box 3: explanation:

Reference:

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

**NEW QUESTION 85**

- (Exam Topic 3)

You plan to provision an Azure Machine Learning Basic edition workspace for a data science project. You need to identify the tasks you will be able to perform in the workspace.

Which three tasks will you be able to perform? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point. D

- A. Create a Compute Instance and use it to run code in Jupyter notebooks.
- B. Create an Azure Kubernetes Service (AKS) inference cluster.
- C. Use the designer to train a model by dragging and dropping pre-defined modules.
- D. Create a tabular dataset that supports versioning.
- E. Use the Automated Machine Learning user interface to train a model.

**Answer: ABD**

**Explanation:**

Reference:

<https://azure.microsoft.com/en-us/pricing/details/machine-learning/>

**NEW QUESTION 89**

- (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 94**

- (Exam Topic 3)

You define a datastore named ml-data for an Azure Storage blob container. In the container, you have a folder named train that contains a file named data.csv.

You plan to use the file to train a model by using the Azure Machine Learning SDK.

You plan to train the model by using the Azure Machine Learning SDK to run an experiment on local compute.

You define a DataReference object by running the following code:

```
from azureml.core import Workspace, Datastore, Environment
from azureml.train.estimator import Estimator
ws = Workspace.from_config()
ml_data = Datastore.get(ws, datastore_name='ml-data')
data_ref = ml_data.path('train').as_download(path_on_compute='train_data')
estimator = Estimator(source_directory='experiment_folder',
    script_params={'--data-folder': data_ref},
    compute_target = 'local',
    entry_script='training.py')
run = experiment.submit(config=estimator)
run.wait_for_completion(show_output=True)
```

You need to load the training data. Which code segment should you use?

- A. 

```
import os
import argparse
import pandas as pd

parser = argparse.ArgumentParser()
parser.add_argument('--data-folder', type=str, dest='data_folder')
data_folder = args.data_folder
data = pd.read_csv(os.path.join(data_folder, 'ml-data', 'train_data', 'data.csv'))
```
- B. 

```
import os
import argparse
import pandas as pd

parser = argparse.ArgumentParser()
parser.add_argument('--data-folder', type=str, dest='data_folder')
data_folder = args.data_folder
data = pd.read_csv(os.path.join(data_folder, 'train', 'data.csv'))
```
- C. 

```
import pandas as pd

data = pd.read_csv('./data.csv')
```
- D. 

```
import os
import argparse
import pandas as pd

parser = argparse.ArgumentParser()
parser.add_argument('--data-folder', type=str, dest='data_folder')
data_folder = args.data_folder
data = pd.read_csv(os.path.join('ml_data', data_folder, 'data.csv'))
```
- E. 

```
import os
import argparse
import pandas as pd

parser = argparse.ArgumentParser()
parser.add_argument('--data-folder', type=str, dest='data_folder')
data_folder = args.data_folder
data = pd.read_csv(os.path.join(data_folder, 'data.csv'))
```

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

**Answer:** E

**Explanation:**

Example:  
`data_folder = args.data_folder`  
`# Load Train and Test data`  
`train_data = pd.read_csv(os.path.join(data_folder, 'data.csv'))` Reference:  
<https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai>

**NEW QUESTION 95**

- (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 100**

- (Exam Topic 3)

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

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

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

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

The IT department creates an Azure Kubernetes Service (AKS)-based inference compute target named aks-cluster in the Azure Machine Learning workspace.

You have a Microsoft Surface Book computer with a GPU. Python 3.6 and Visual Studio Code are installed. You need to run a script that trains a deep neural network (DNN) model and logs the loss and accuracy metrics.

Solution: Attach the mlvm virtual machine as a compute target in the Azure Machine Learning workspace. Install the Azure ML SDK on the Surface Book and run Python code to connect to the workspace. Run the training script as an experiment on the mlvm remote compute resource.

- A. Yes
- B. No

**Answer:** A

**Explanation:**

Use the VM as a compute target.

Note: A compute target is a designated compute resource/environment where you run your training script or host your service deployment. This location may be your local machine or a cloud-based compute resource.

Reference:

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

**NEW QUESTION 105**

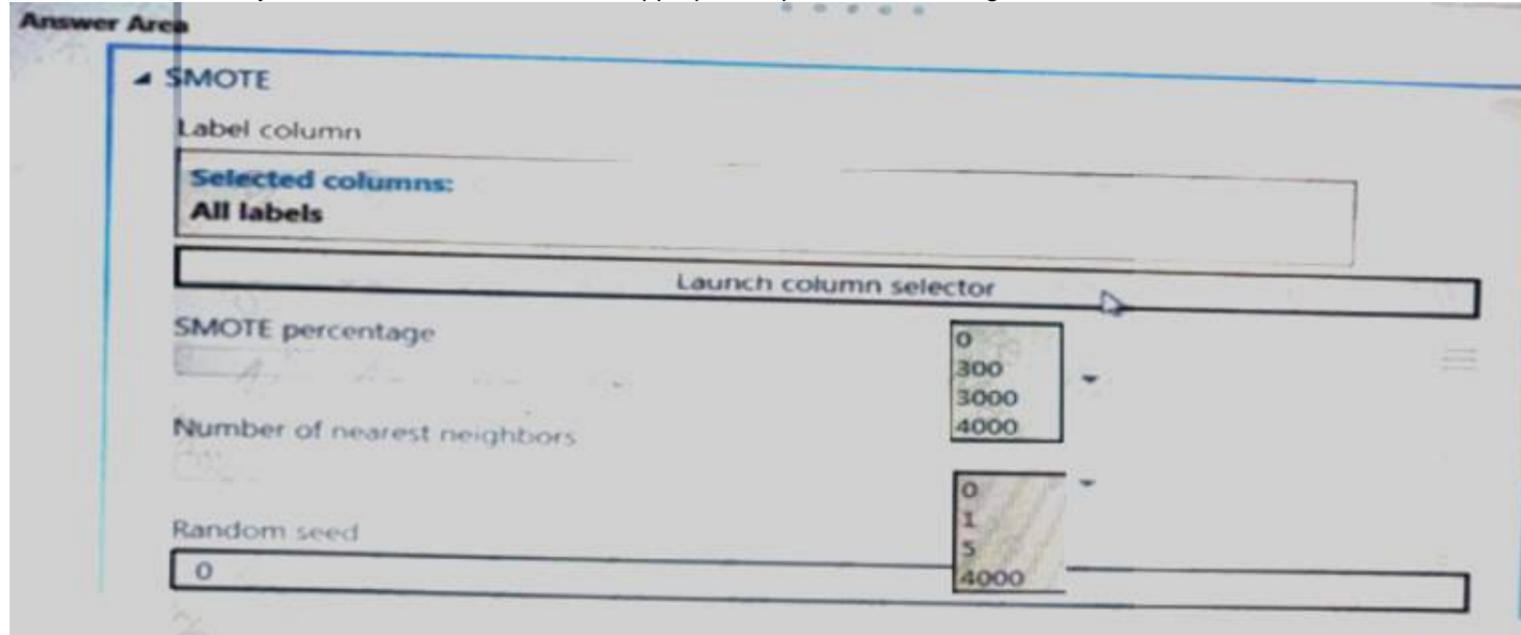
- (Exam Topic 3)

You create an experiment in Azure Machine Learning Studio- You add a training dataset that contains 10,000 rows. The first 9,000 rows represent class 0 (90 percent). The first 1,000 rows represent class 1 (10 percent).

The training set is unbalanced between two Classes. You must increase the number of training examples for class 1 to 4,000 by using data rows. You add the Synthetic Minority Oversampling Technique (SMOTE) module to the experiment.

You need to configure the module.

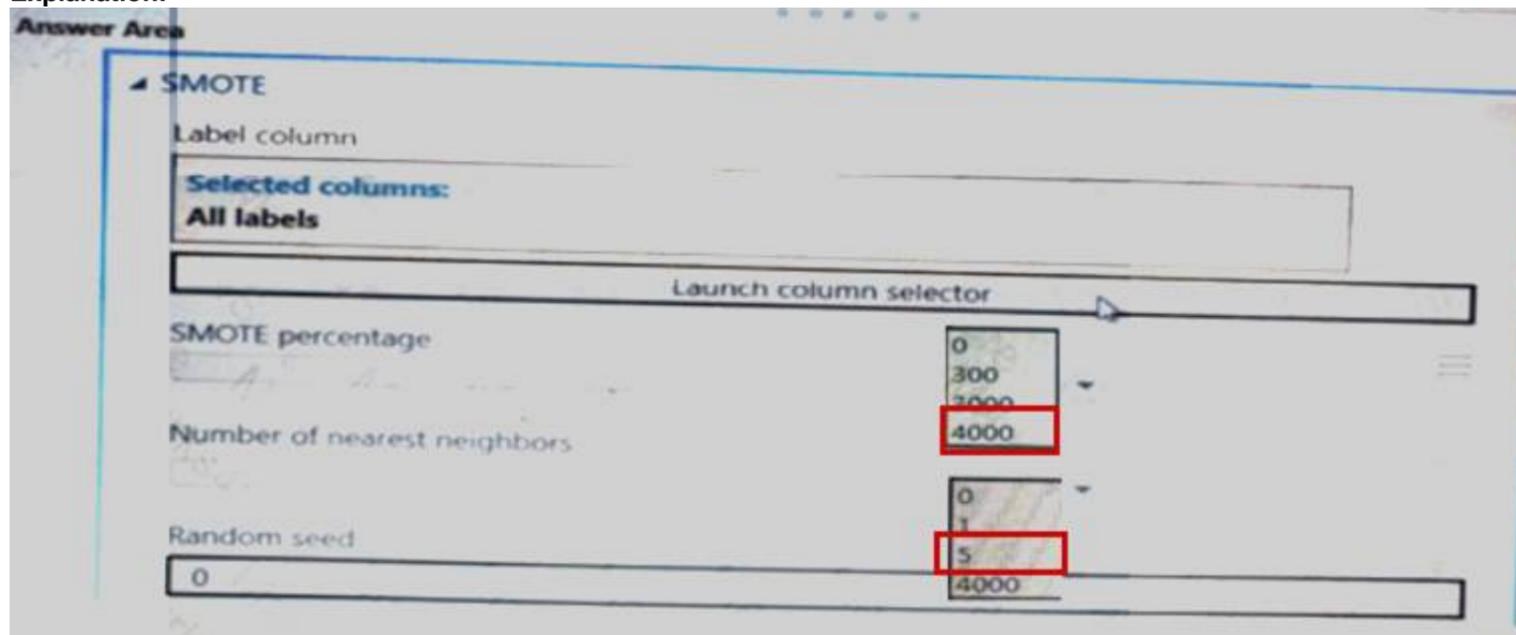
Which values should you use? To answer, select the appropriate options in the dialog box in the answer area. NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

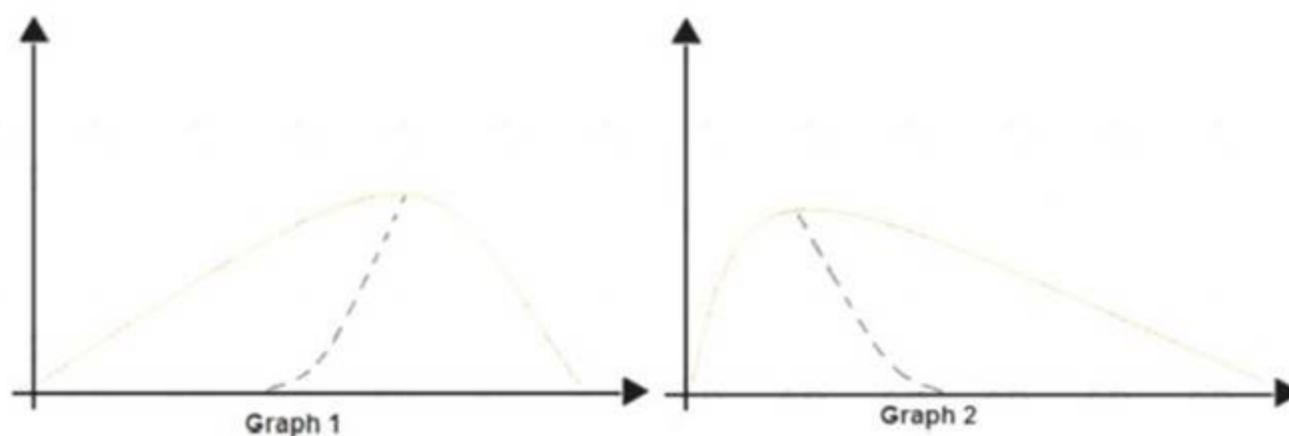


**NEW QUESTION 110**

- (Exam Topic 3)

You are analyzing the asymmetry in a statistical distribution.

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



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

NOTE: Each correct selection is worth one point.

**Question**

**Answer choice**

Which type of distribution is shown for the dataset density curve of Graph 1?

▼

Negative skew

Positive skew

Normal distribution

Bimodal distribution

Which type of distribution is shown for the dataset density curve of Graph 2?

▼

Negative skew

Positive skew

Normal distribution

Bimodal distribution

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Positive skew

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

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

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

**NEW QUESTION 114**

- (Exam Topic 3)

You are analyzing a raw dataset that requires cleaning.

You must perform transformations and manipulations by using Azure Machine Learning Studio. You need to identify the correct modules to perform the transformations.

Which modules should you choose? To answer, drag the appropriate modules to the correct scenarios. Each module 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.

**Answer Area**

Methods	Scenario	Module
Clean Missing Data	Replace missing values by removing rows and columns.	
SMOTE	Increase the number of low-incidence examples in the dataset.	
Convert to Indicator Values	Convert a categorical feature into a binary indicator.	
Remove Duplicate Rows	Remove potential duplicates from a dataset.	
Threshold Filter		

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Clean Missing Data Box 2: SMOTE

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

Box 3: Convert to Indicator Values

Use the Convert to Indicator Values module in Azure Machine Learning Studio. The purpose of this module is to convert columns that contain categorical values into a series of binary indicator columns that can more easily be used as features in a machine learning model.

Box 4: Remove Duplicate Rows References:

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

**NEW QUESTION 119**

- (Exam Topic 3)

You are using Azure Machine Learning to train machine learning models. You need a compute target on which to remotely run the training script. You run the following Python code:

```
from azureml.core.compute import ComputeTarget, AmlCompute
from azureml.core.compute_target import ComputeTargetException
the_cluster_name = "NewCompute"
config = AmlCompute.provisioning_configuration(vm_size='STANDARD_D2', max_nodes=3)
the_cluster = ComputeTarget.create(ws, the_cluster_name, config)
```

**Answer Area**

	Yes	No
The compute is created in the same region as the Machine Learning service workspace.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
The compute resource created by the code is displayed as a compute cluster in Azure Machine Learning studio.	<input type="checkbox"/>	<input type="checkbox"/>
The minimum number of nodes will be zero.	<input type="checkbox"/>	<input type="checkbox"/>

- A. Mastered
- B. Not Mastered

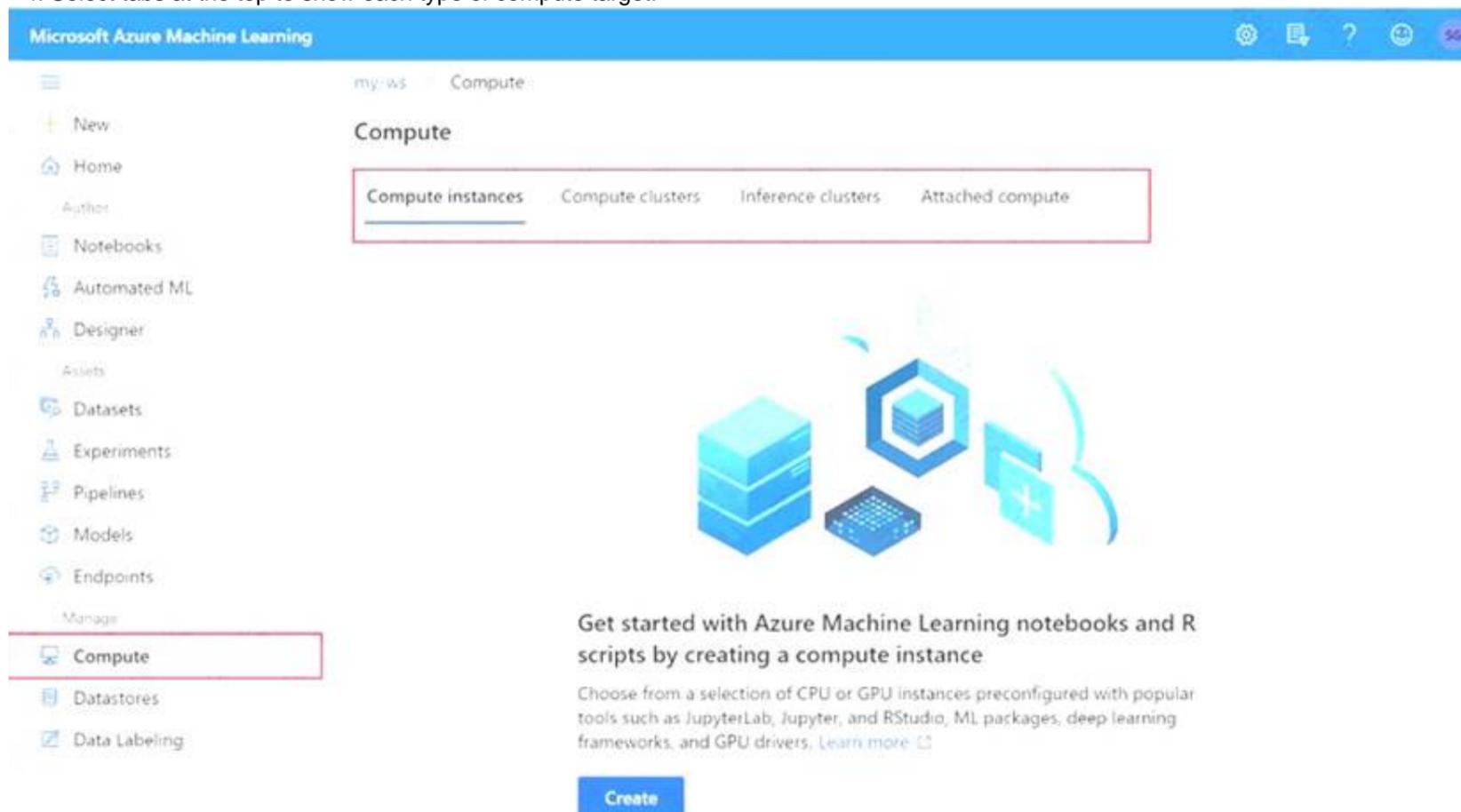
**Answer: A**

**Explanation:**

Box 1: Yes

The compute is created within your workspace region as a resource that can be shared with other users. Box 2: Yes  
It is displayed as a compute cluster. View compute targets

- \* 1. To see all compute targets for your workspace, use the following steps:
- \* 2. Navigate to Azure Machine Learning studio.
- \* 3. Under Manage, select Compute.
- \* 4. Select tabs at the top to show each type of compute target.



Box 3: Yes

min\_nodes is not specified, so it defaults to 0. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute.amlcomputeprovi> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-studio>

**NEW QUESTION 123**

- (Exam Topic 3)

You are creating a new experiment in Azure Machine Learning Studio. You have a small dataset that has missing values in many columns. The data does not require the application of predictors for each column. You plan to use the Clean Missing Data module to handle the missing data. You need to select a data cleaning method. Which method should you use?

- A. Synthetic Minority Oversampling Technique (SMOTE)
- B. Replace using MICE
- C. Replace using; Probabilistic PCA
- D. Normalization

**Answer: C**

**Explanation:**

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

References:

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

**NEW QUESTION 126**

- (Exam Topic 3)

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

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

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

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

**Modules**

- Export Data
- Tune Model Hyperparameters
- Cross Validate Model
- Evaluate Model
- Score Model
- Train Model

**Answer Area**

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Step 1: Train Model

Two-Class Boosted Decision Tree

First, set up the boosted decision tree model.

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

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

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

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

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



Step 2: Score Model

Score and evaluate the models

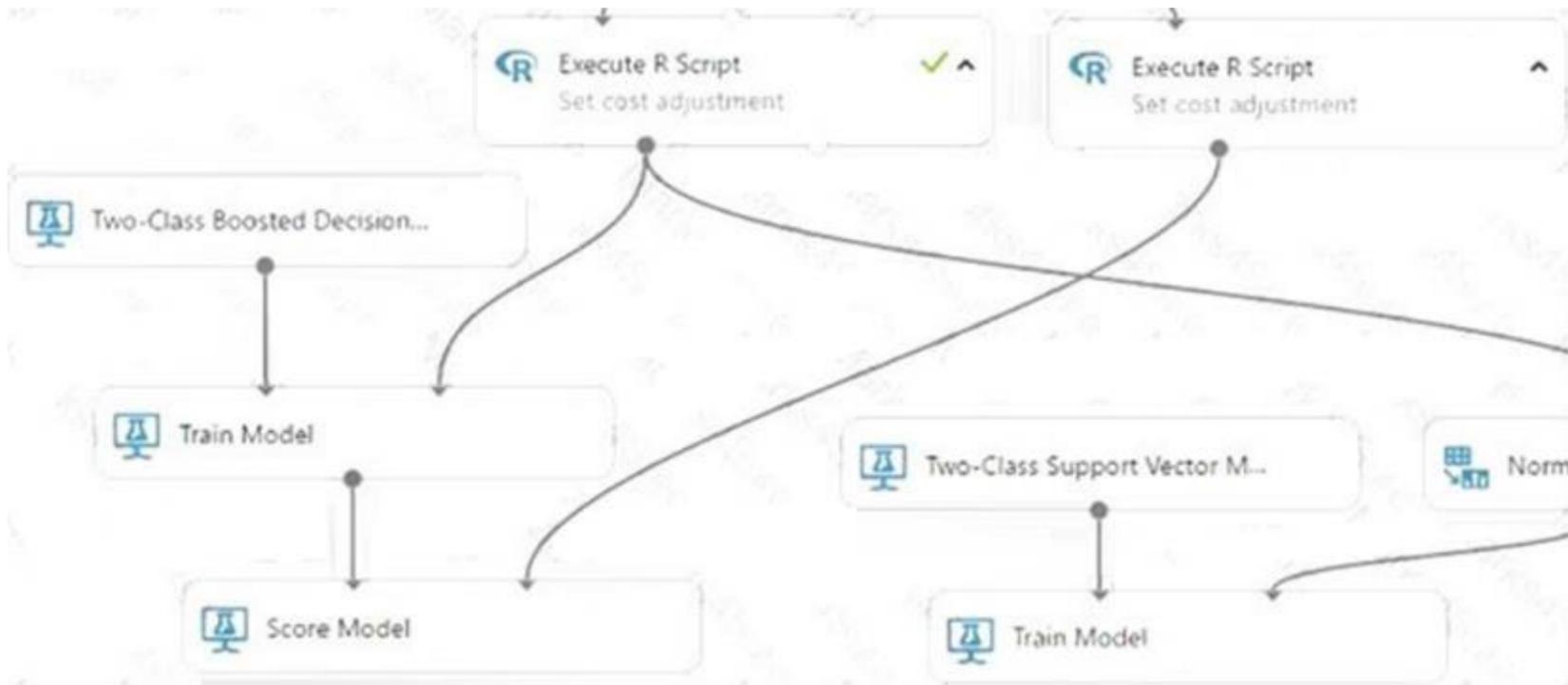
You use the testing data that was separated out by the Split Data module to score our trained models. You can then compare the results of the two models to see which generated better results.

Add the Score Model modules

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

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

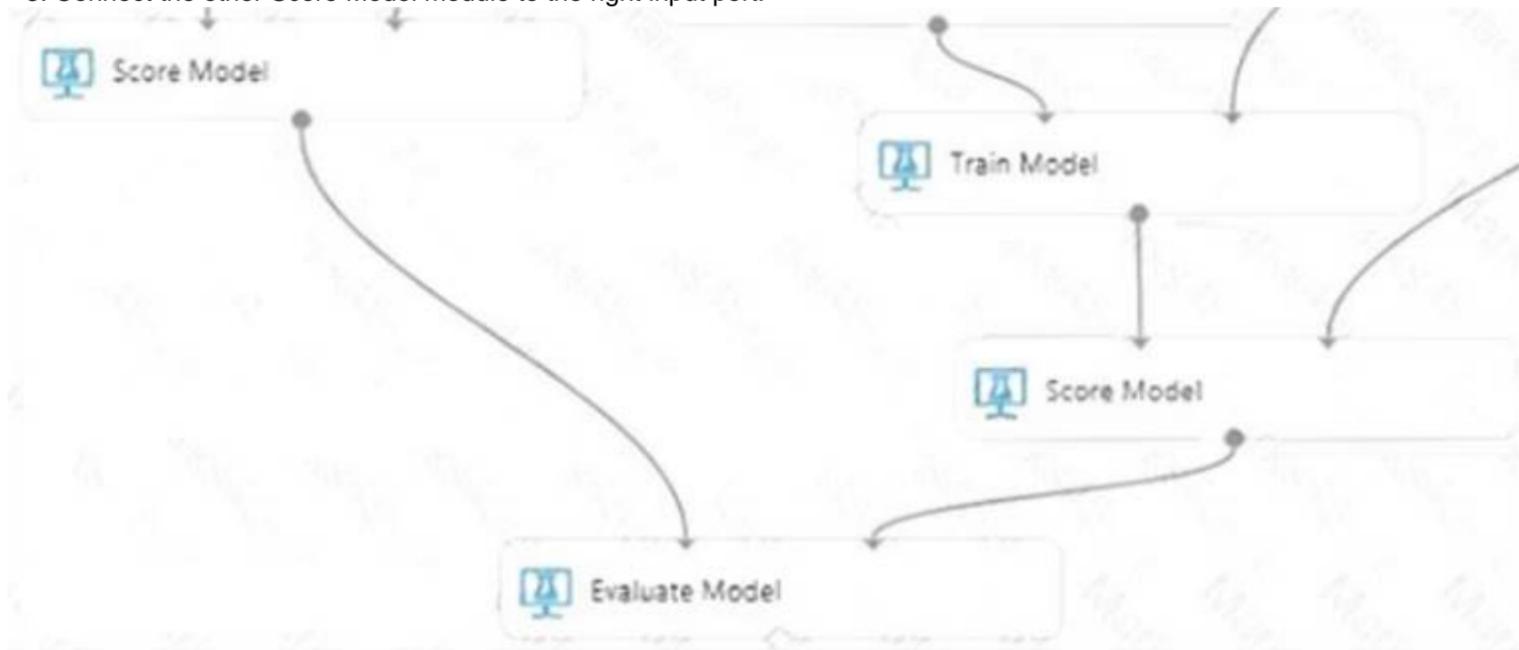
\* 3. Connect the right Execute R Script module (our testing data) to the right input port of the Score Model module.



**Step 3: Evaluate Model**

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

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



**NEW QUESTION 128**

- (Exam Topic 3)

You publish a batch inferencing pipeline that will be used by a business application.

The application developers need to know which information should be submitted to and returned by the REST interface for the published pipeline.

You need to identify the information required in the REST request and returned as a response from the published pipeline.

Which values should you use in the REST request and to expect in the response? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

REST Request	Value
Request Header	JSON containing the run ID JSON containing the pipeline ID JSON containing the experiment name JSON containing an OAuth bearer token
Response	JSON containing the run ID JSON containing the pipeline ID JSON containing the experiment name JSON containing an OAuth bearer token
Response	JSON containing the run ID JSON containing a list of predictions JSON containing the experiment name JSON containing a path to the parallel_run_step.txt output file

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: JSON containing an OAuth bearer token Specify your authentication header in the request.  
To run the pipeline from the REST endpoint, you need an OAuth2 Bearer-type authentication header. Box 2: JSON containing the experiment name  
Add a JSON payload object that has the experiment name. Example:  
rest\_endpoint = published\_pipeline.endpoint response = requests.post(rest\_endpoint, headers=auth\_header, json={"ExperimentName": "batch\_scoring", "ParameterAssignments": {"process\_count\_per\_node": 6}}) run\_id = response.json()["Id"]  
Box 3: JSON containing the run ID  
Make the request to trigger the run. Include code to access the Id key from the response dictionary to get the value of the run ID.  
Reference:  
<https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-pipeline-batch-scoring-classification>

**NEW QUESTION 130**

- (Exam Topic 3)

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

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

You train and register a machine learning model.

You plan to deploy the model as a real-time web service. Applications must use key-based authentication to use the model.

You need to deploy the web service.

Solution:

Create an AksWebservice instance.

Set the value of the auth\_enabled property to False.

Set the value of the token\_auth\_enabled property to True.

Deploy the model to the service. Does the solution meet the goal?

- A. Yes
- B. No

**Answer: B**

**Explanation:**

Instead use only auth\_enabled = TRUE Note: Key-based authentication.

Web services deployed on AKS have key-based auth enabled by default. ACI-deployed services have key-based auth disabled by default, but you can enable it by setting auth\_enabled = TRUE when creating the ACI web service. The following is an example of creating an ACI deployment configuration with key-based auth enabled.

deployment\_config <- aci\_websevice\_deployment\_config(cpu\_cores = 1, memory\_gb = 1,

auth\_enabled = TRUE) Reference:

<https://azure.github.io/azureml-sdk-for-r/articles/deploying-models.html>

**NEW QUESTION 135**

- (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	Answer Area						
Vowpal Wabbit	<table border="1"> <thead> <tr> <th>Task</th> <th>Tool</th> </tr> </thead> <tbody> <tr> <td>Build DNN models</td> <td>Tool</td> </tr> <tr> <td>Enable interactive data exploration and visualization</td> <td>Tool</td> </tr> </tbody> </table>	Task	Tool	Build DNN models	Tool	Enable interactive data exploration and visualization	Tool
Task	Tool						
Build DNN models	Tool						
Enable interactive data exploration and visualization	Tool						
PowerBI Desktop							
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 139**

- (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.

**Answer Area**

	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="checkbox"/>	<input type="checkbox"/>
The estimator will mount the local data-folder folder and make it available to the script through a parameter.	<input type="checkbox"/>	<input type="checkbox"/>
The train.py script file will be created if it does not exist.	<input type="checkbox"/>	<input type="checkbox"/>

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

**Answer Area**

	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="checkbox"/>	<input type="checkbox"/>
The estimator will mount the local data-folder folder and make it available to the script through a parameter.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
The train.py script file will be created if it does not exist.	<input type="checkbox"/>	<input checked="" type="checkbox"/>

**NEW QUESTION 142**

- (Exam Topic 3)

You create a classification model with a dataset that contains 100 samples with Class A and 10,000 samples with Class B. The variation of Class B is very high. You need to resolve imbalances. Which method should you use?

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

**Answer: D**

**NEW QUESTION 145**

- (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:**

s 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 150**

- (Exam Topic 3)

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

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

You train and register a machine learning model.

You plan to deploy the model as a real-time web service. Applications must use key-based authentication to use the model.

You need to deploy the web service. Solution:

Create an AciWebservice instance.

Set the value of the ssl\_enabled property to True.

Deploy the model to the service. Does the solution meet the goal?

- A. Yes
- B. No

**Answer:** B

**Explanation:**

Instead use only auth\_enabled = TRUE Note: Key-based authentication.

Web services deployed on AKS have key-based auth enabled by default. ACI-deployed services have

key-based auth disabled by default, but you can enable it by setting auth\_enabled = TRUE when creating the ACI web service. The following is an example of creating an ACI deployment configuration with key-based auth enabled.

deployment\_config <- aci\_webservice\_deployment\_config(cpu\_cores = 1,

memory\_gb = 1, auth\_enabled = TRUE) Reference:

<https://azure.github.io/azureml-sdk-for-r/articles/deploying-models.html>

**NEW QUESTION 151**

- (Exam Topic 3)

You have a model with a large difference between the training and validation error values. You must create a new model and perform cross-validation.

You need to identify a parameter set for the new model using Azure Machine Learning Studio.

Which module you should use for each step? To answer, drag the appropriate modules to the correct steps. Each module may be used once or 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	Step	Module
Two-Class Boosted Decision Tree	Define the parameter scope	
Partition and Sample	Define the cross-validation settings	
Tune Model Hyperparameters	Define the metric	
Split Data	Train, evaluate, and compare	

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Split data

Box 2: Partition and Sample

Box 3: Two-Class Boosted Decision Tree Box 4: Tune Model Hyperparameters

Integrated train and tune: You configure a set of parameters to use, and then let the module iterate over multiple combinations, measuring accuracy until it finds a "best" model. With most learner modules, you can choose which parameters should be changed during the training process, and which should remain fixed.

We recommend that you use Cross-Validate Model to establish the goodness of the model given the specified

parameters. Use Tune Model Hyperparameters to identify the optimal parameters. References:

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

**NEW QUESTION 154**

- (Exam Topic 3)

You have an Azure Machine Learning workspace that contains a CPU-based compute cluster and an Azure Kubernetes Services (AKS) inference cluster. You create a tabular dataset containing data that you plan to use to create a classification model.

You need to use the Azure Machine Learning designer to create a web service through which client applications can consume the classification model by submitting new data and getting an immediate prediction as a response.

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
Create and run a batch inference pipeline on the compute cluster.	
Deploy a real-time endpoint on the inference cluster.	
Create and run a real-time inference pipeline on the compute cluster.	⬅️ ⬆️
Create and run a training pipeline that prepares the data and trains a classification model on the compute cluster.	➡️ ⬇️
Use the automated ML user interface to train a classification model on the compute cluster.	
Create and start a Compute Instance.	

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Step 1: Create and start a Compute Instance

To train and deploy models using Azure Machine Learning designer, you need compute on which to run the training process, test the model, and host the model in a deployed service.

There are four kinds of compute resource you can create:

Compute Instances: Development workstations that data scientists can use to work with data and models. Compute Clusters: Scalable clusters of virtual machines for on-demand processing of experiment code. Inference Clusters: Deployment targets for predictive services that use your trained models.

Attached Compute: Links to existing Azure compute resources, such as Virtual Machines or Azure Databricks clusters.

Step 2: Create and run a training pipeline..

After you've used data transformations to prepare the data, you can use it to train a machine learning model. Create and run a training pipeline

Step 3: Create and run a real-time inference pipeline

After creating and running a pipeline to train the model, you need a second pipeline that performs the same data transformations for new data, and then uses the trained model to inference (in other words, predict) label values based on its features. This pipeline will form the basis for a predictive service that you can publish for applications to use.

Reference:

<https://docs.microsoft.com/en-us/learn/modules/create-classification-model-azure-machine-learning-designer/>

**NEW QUESTION 156**

- (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 158**

- (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 160**

- (Exam Topic 3)

You create an Azure Machine Learning workspace named workspace1. You assign a custom role to a user of workspace1.

The custom role has the following JSON definition:

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

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

NOTE: Each correct selection is worth one point.

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

A.

**Answer:**

**Explanation:**

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

Box 1: No

The actions listed in NotActions are prohibited.

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

Box 2: No

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

Writing metrics is not listed in NotActions. Reference:

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

**NEW QUESTION 163**

- (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 165**

- (Exam Topic 3)

You are developing a linear regression model in Azure Machine Learning Studio. You run an experiment to compare different algorithms.

The following image displays the results dataset output:

Algorithm	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error
Bayesian Liner	3.276025	4.655442	0.511436	0.282138
Neural Network	2.676538	3.621476	0.417847	0.17073
Boosted Decision Tree	2.168847	2.878077	0.338589	0.107831
Linear	6.350005	8.720718	0.99133	0.99002
Decision Forest	2.390206	3.315 164	0.373146	0.14307

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

NOTE: Each correct selection is worth one point.

**Question**

**Answer choice**

Which algorithm minimizes differences between actual and predicted values?

▼

- Bayesian Linear Regression
- Neural Network Regression
- Boosted Decision Tree Regression
- Linear Regression
- Decision Forest Regression

Which approach should you use to find the best parameters for a Linear Regression model for the Online Gradient Descent method?

▼

- Set the Decrease learning rate option to True.
- Set the Decrease learning rate option to True.
- Set the Create trainer mode option to Parameter Range.
- Increase the number of epochs.
- Decrease the number of epochs.

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Boosted Decision Tree Regression

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

Box 2:

Online Gradient Descent: If you want the algorithm to find the best parameters for you, set Create trainer mode option to Parameter Range. You can then specify multiple values for the algorithm to try.

References:

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

### NEW QUESTION 167

- (Exam Topic 3)

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

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

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

```
from azureml.core import Run
import pandas as pd

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

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

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

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

Does the solution meet the goal?

- A. Yes
- B. No

**Answer: B**

#### Explanation:

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

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

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

### NEW QUESTION 168

- (Exam Topic 3)

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

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

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

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

All files store data in the following format:

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

You run the following code:

```
data_store = Datastore.register_azure_blob_container(workspace=ws,
    datastore_name='data_store',
    container_name='quarterly_data',
    account_name='companydata',
    account_key='NRPxk8duxBM3...'
    create_if_not_exists=False)
```

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

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

Solution: Run the following code:

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

Does the solution meet the goal?

- A. Yes
- B. No

**Answer: A**

#### Explanation:

Use two file paths.

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

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

Reference:

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

### NEW QUESTION 173

- (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 177

- (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 182

- (Exam Topic 3)

You are creating a new Azure Machine Learning pipeline using the designer.

The pipeline must train a model using data in a comma-separated values (CSV) file that is published on a website. You have not created a dataset for this file.

You need to ingest the data from the CSV file into the designer pipeline using the minimal administrative effort.

Which module should you add to the pipeline in Designer?

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

**Answer:** D

#### Explanation:

The preferred way to provide data to a pipeline is a Dataset object. The Dataset object points to data that lives in or is accessible from a datastore or at a Web URL. The Dataset class is abstract, so you will create an instance of either a FileDataset (referring to one or more files) or a TabularDataset that's created by from one or more files with delimited columns of data.

Example:

```
from azureml.core import Dataset
```

```
iris_tabular_dataset = Dataset.Tabular.from_delimited_files([(def_blob_store, 'train-dataset/iris.csv')])
```

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-your-first-pipeline>

### NEW QUESTION 183

- (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 binning mode with a PQuantile 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 184**

- (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 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.

References:

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

**NEW QUESTION 187**

- (Exam Topic 3)

You create a binary classification model. The model is registered in an Azure Machine Learning workspace. You use the Azure Machine Learning Fairness SDK to assess the model fairness.

You develop a training script for the model on a local machine.

You need to load the model fairness metrics into Azure Machine Learning studio. What should you do?

- A. Implement the `download_dashboard_by_upload_id` function
- B. Implement the `create_group_metric_sec` function
- C. Implement the `upload_dashboard_dictionary` function
- D. Upload the training script

**Answer: C**

**Explanation:**

import `azureml.contrib.fairness` package to perform the upload:

from `azureml.contrib.fairness` import `upload_dashboard_dictionary`, `download_dashboard_by_upload_id` Reference:

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

**NEW QUESTION 188**

- (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 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 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)
```

You plan to use this configuration to run a script that trains a random forest model and then tests it with validation data. The label values for the validation data are stored in a variable named y\_test variable, and the predicted probabilities from the model are stored in a variable named y\_predicted.

Solution: Run the following code:

```
import numpy as np
from sklearn.metrics import roc_auc_score
from azureml.core.run import Run
run = Run.get_context()
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
run.log("AUC", np.float(auc))
```

Does the solution meet the goal?

- A. Yes
- B. No

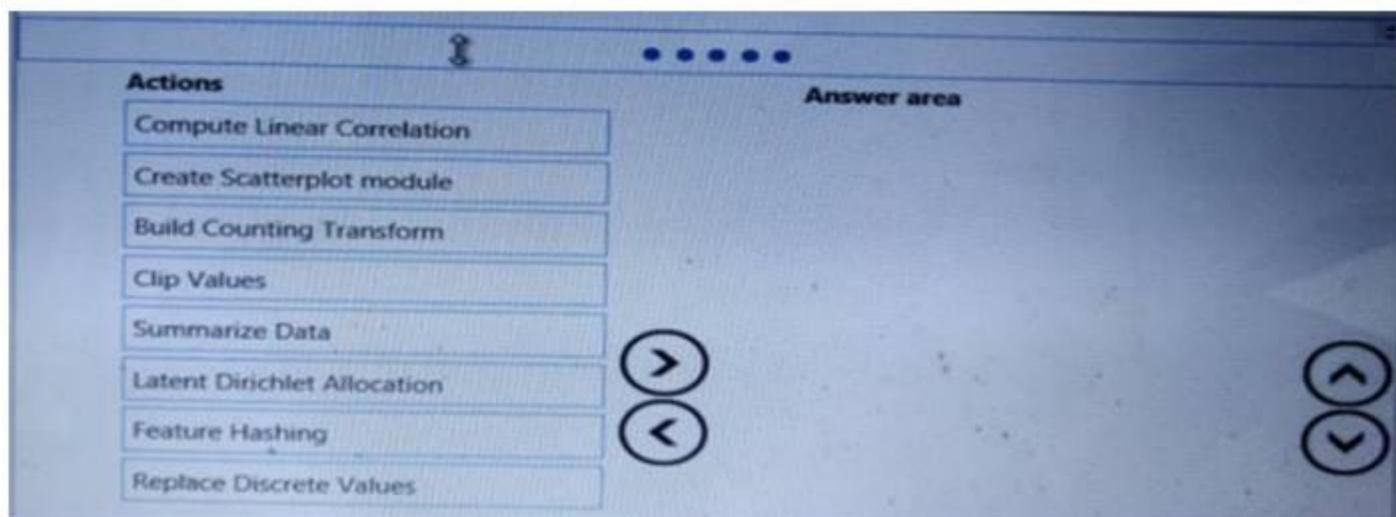
**Answer:** A

**NEW QUESTION 191**

- (Exam Topic 2)

You need to visually identify whether outliers exist in the Age column and quantify the outliers before the outliers are removed.

Which three Azure Machine Learning Studio modules should you use in sequence? To answer, move the appropriate modules from the list of modules to the answer area and arrange them in the correct order.



- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Create Scatterplot Summarize Data Clip Values

You can use the Clip Values module in Azure Machine Learning Studio, to identify and optionally replace data values that are above or below a specified threshold. This is useful when you want to remove outliers or replace them with a mean, a constant, or other substitute value.

References:

<https://blogs.msdn.microsoft.com/azuredev/2017/05/27/data-cleansing-tools-in-azure-machine-learning/> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clip-values>

**NEW QUESTION 196**

- (Exam Topic 2)

You need to correct the model fit issue.

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

**Actions**

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

**Answer Area**

⏪ ⏩

⏴ ⏵

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Step 1: Augment the data

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

Step 2: Add the Bayesian Linear Regression module.

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

Step 3: Configure the regularization weight.

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

Scenario:

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

**NEW QUESTION 201**

- (Exam Topic 2)

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

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

**Answer:** A

**Explanation:**

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

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

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

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

**NEW QUESTION 202**

- (Exam Topic 2)

You need to configure the Feature Based Feature Selection module based on the experiment requirements and datasets.

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

NOTE: Each correct selection is worth one point.

Filter Based Feature Selection

Feature scoring method

Fisher Score

Chi-squared

Mutual information

Counts

Operate on feature columns only

Target column

MedianValue

AvgRooms/nHouse

Launch column selector

Number of desired features

1

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Mutual Information.

The mutual information score is particularly useful in feature selection because it maximizes the mutual information between the joint distribution and target variables in datasets with many dimensions.

Box 2: MedianValue

MedianValue is the feature column, , it is the predictor of the dataset.

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

References:

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

**NEW QUESTION 203**

- (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 205**

- (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:

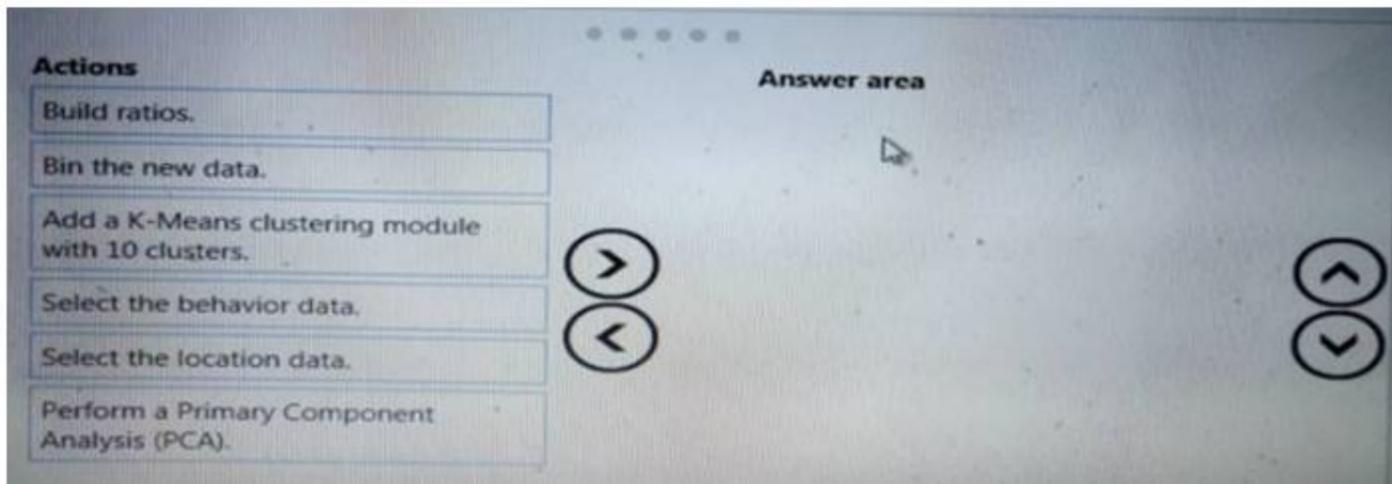
Local penalty detection models must be written by using BrainScript. References:

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

**NEW QUESTION 206**

- (Exam Topic 1)

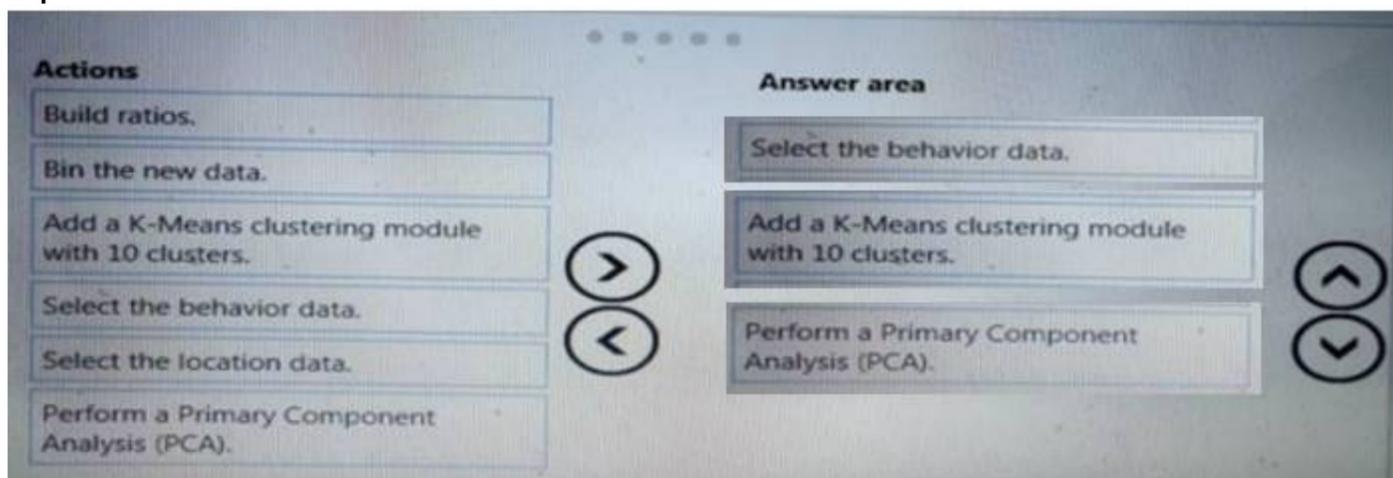
You need to modify the inputs for the global penalty event model to address the bias and variance issue. Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**



**NEW QUESTION 210**

- (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).

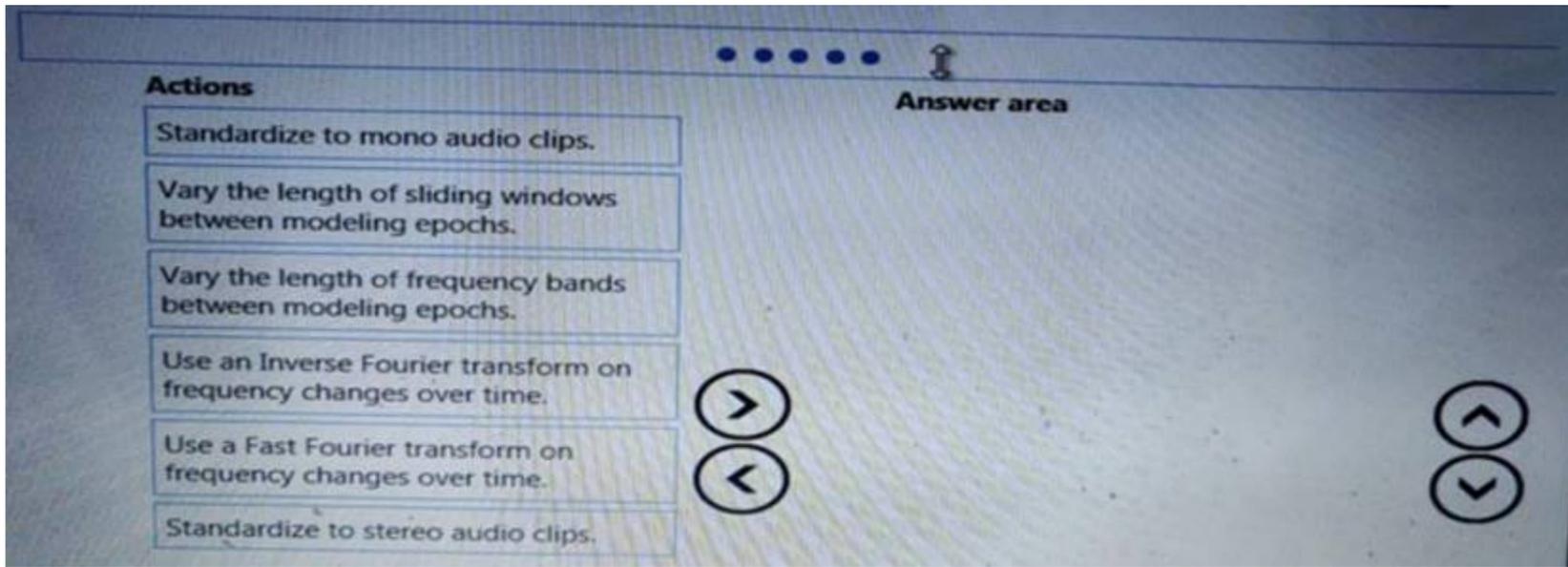
**Answer: A**

**NEW QUESTION 213**

- (Exam Topic 1)

You need to define a process for penalty event detection.

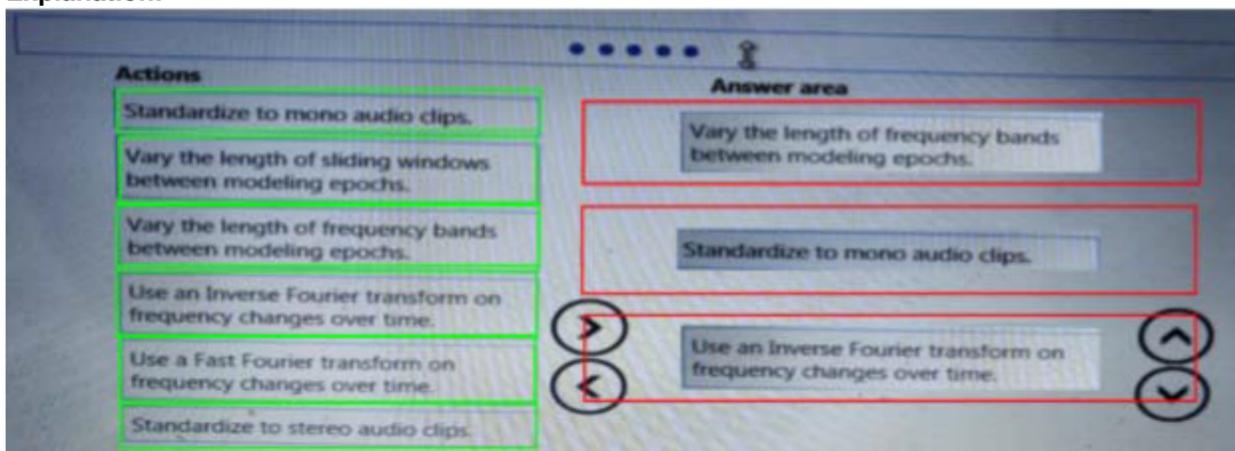
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**



**NEW QUESTION 215**

- (Exam Topic 1)

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

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

**Actions**

**Answer Area**

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



- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Step 1: Define a cross-entropy function activation

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

Step 2: Add cost functions for each target state. Step 3: Evaluated the distance error metric. References:

<https://www.analyticsvidhya.com/blog/2018/04/fundamentals-deep-learning-regularization-techniques/>

**NEW QUESTION 220**

- (Exam Topic 1)

You need to define a modeling strategy for ad response.

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.

Action	Answer area
Implement a K-Means Clustering model.	
Use the raw score as a feature in a Score Matchbox Recommender model.	
Use the cluster as a feature in a Decision Jungle model.	⬅️ ⬆️
Use the raw score as a feature in a Logistic Regression model.	⬆️
Implement a Sweep Clustering model.	

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Step 1: Implement a K-Means Clustering model

Step 2: Use the cluster as a feature in a Decision jungle model.

Decision jungles are non-parametric models, which can represent non-linear decision boundaries. Step 3: Use the raw score as a feature in a Score Matchbox Recommender model

The goal of creating a recommendation system is to recommend one or more "items" to "users" of the system. Examples of an item could be a movie, restaurant, book, or song. A user could be a person, group of persons, or other entity with item preferences.

Scenario:

Ad response rated declined.

Ad response models must be trained at the beginning of each event and applied during the sporting event. Market segmentation models must optimize for similar ad response history.

Ad response models must support non-linear boundaries of features. References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/multiclass-decision-jungle> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/score-matchbox-recommende>

**NEW QUESTION 225**

- (Exam Topic 1)

You need to implement a new cost factor scenario for the ad response models as illustrated in the performance curve exhibit.

Which technique should you use?

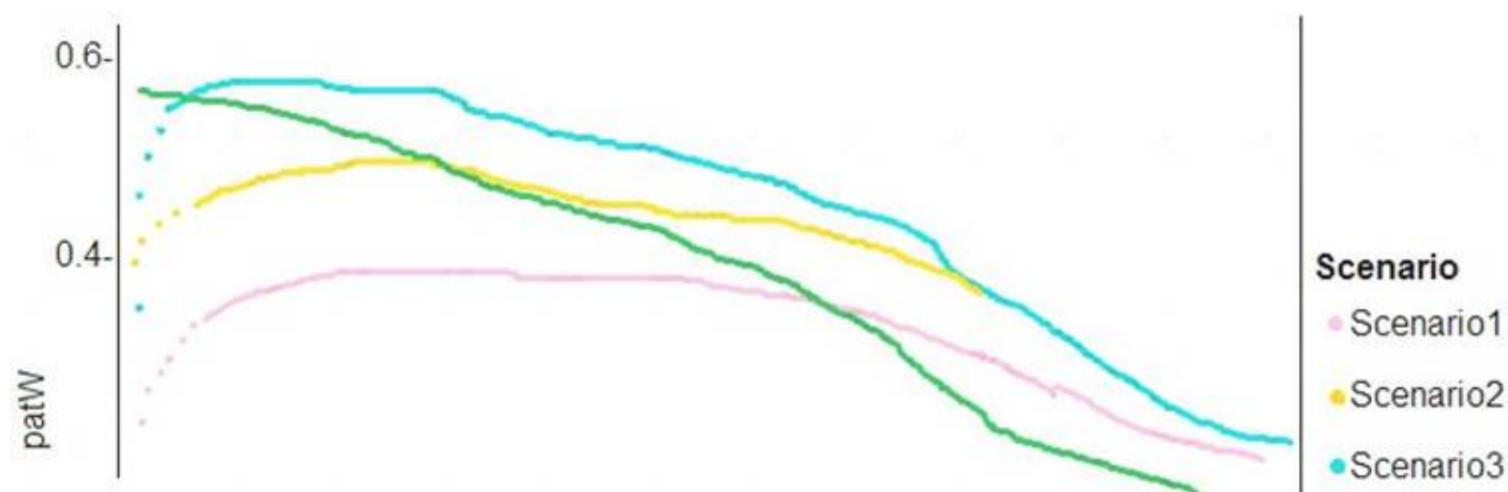
- A. Set the threshold to 0.5 and retrain if weighted Kappa deviates +/- 5% from 0.45.
- B. Set the threshold to 0.05 and retrain if weighted Kappa deviates +/- 5% from 0.5.
- C. Set the threshold to 0.2 and retrain if weighted Kappa deviates +/- 5% from 0.6.
- D. Set the threshold to 0.75 and retrain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

**Answer:** A

**Explanation:**

Scenario:

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



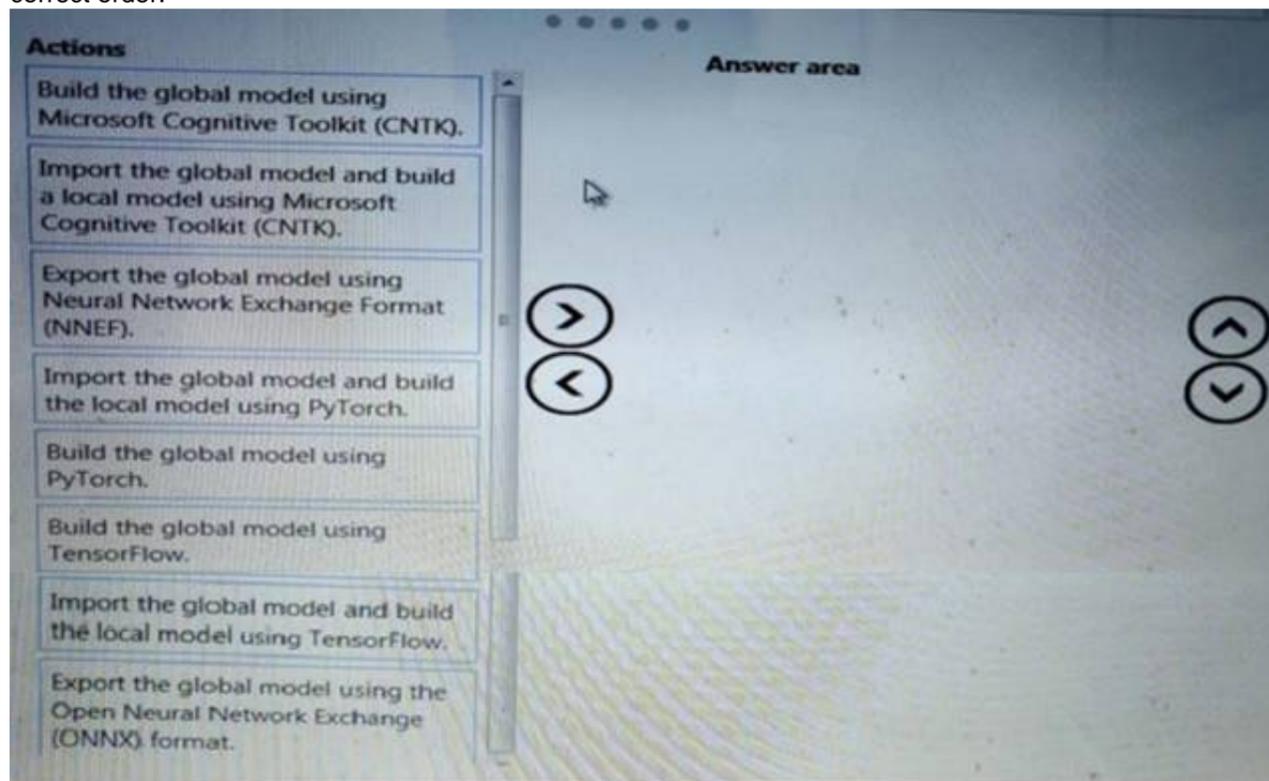
The ad propensity model uses a cut threshold is 0.45 and retrains occur if weighted Kappa deviated from 0.1 +/- 5%.

**NEW QUESTION 227**

- (Exam Topic 1)

You need to define a process for penalty event detection.

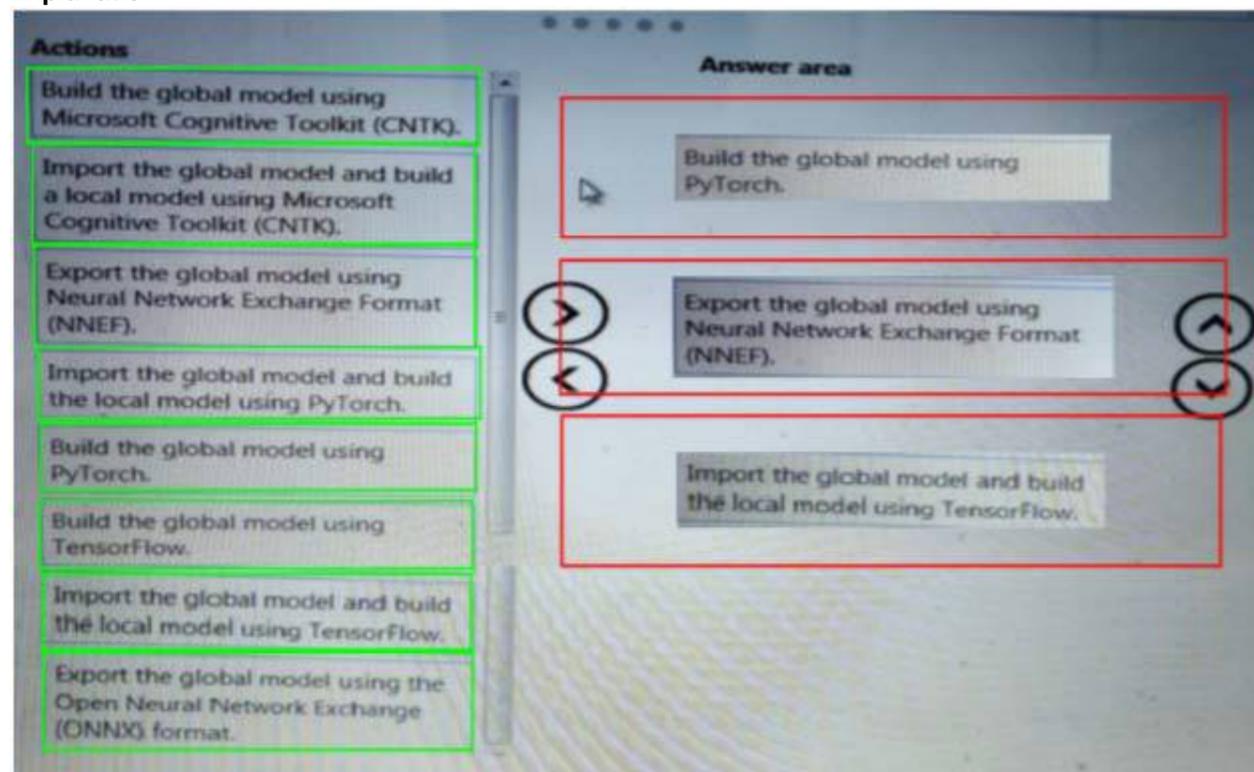
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**



**NEW QUESTION 228**

- (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 231**

.....

## Thank You for Trying Our Product

\* 100% Pass or Money Back

All our products come with a 90-day Money Back Guarantee.

\* One year free update

You can enjoy free update one year. 24x7 online support.

\* Trusted by Millions

We currently serve more than 30,000,000 customers.

\* Shop Securely

All transactions are protected by VeriSign!

**100% Pass Your DP-100 Exam with Our Prep Materials Via below:**

<https://www.certleader.com/DP-100-dumps.html>