QUESTION 141
You deploy a model as an Azure Machine Learning real-time web service using the following code.

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

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

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

Answer: A
Explanation:
You can print out detailed Docker engine log messages from the service object. You can view the log for ACI, AKS, and Local deployments. The following example demonstrates how to print the logs.
# if you already have the service object handy
print(service.get_logs())
# if you only know the name of the service (note there might be multiple services with the same name but different version number)
print(ws.webservices['mysvc'].get_logs())
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment

QUESTION 142
You create a multi-class image classification deep learning model.
You train the model by using PyTorch version 1.2.
You need to ensure that the correct version of PyTorch can be identified for the inferencing environment when the model is deployed.
What should you do?

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

Answer: D
Explanation:
framework_version: The PyTorch version to be used for executing training code.
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.dnn.pytorch?view=azure-ml-py

QUESTION 143
You are a lead data scientist for a project that tracks the health and migration of birds. You create a multi- class image classification deep learning model that uses a set of labeled bird photographs collected by experts.
You have 100,000 photographs of birds. All photographs use the JPG format and are stored in an Azure blob container in an Azure subscription.
You need to access the bird photograph files in the Azure blob container from the Azure Machine Learning service workspace that will be used for deep learning model training. You must minimize data movement.
What should you do?

A. Create an Azure Data Lake store and move the bird photographs to the store.
B. Create an Azure Cosmos DB database and attach the Azure Blob containing bird photographs storage to the database.
C. Create and register a dataset by using TabularDataset class that references the Azure blob storage containing bird photographs.
D. Register the Azure blob storage containing the bird photographs as a datastore in Azure Machine Learning service.
E. Copy the bird photographs to the blob datastore that was created with your Azure Machine Learning service workspace.

Answer: D
Explanation:
We recommend creating a datastore for an Azure Blob container. When you create a workspace, an Azure blob container and an Azure file share are automatically registered to the workspace.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-access-data

QUESTION 144
You use the Azure Machine Learning service to create a tabular dataset named **training_data**. You plan to use this dataset in a training script.
You create a variable that references the dataset using the following code:
`training_ds = workspace.datasets.get("training_data")`
You define an estimator to run the script.
You need to set the correct property of the estimator to ensure that your script can access the training_data dataset.
Which property should you set?

A. environment_definition = {"training_data":training_ds}
B. inputs = [training_ds.as_named_input('training_ds')]
C. script_params = {"--training_ds":training_ds}
D. source_directory = training_ds

Answer: B
Explanation:
Example:
# Get the training dataset
diabetes_ds = ws.datasets.get("Diabetes Dataset")
# Create an estimator that uses the remote compute
hyper_estimator = SKLearn(source_directory=experiment_folder, inputs=[diabetes_ds.as_named_input('diabetes')], # Pass the dataset as an input compute_target = cpu_cluster,
conda_packages=['pandas','ipykernel','matplotlib'],
pip_packages=['azureml-sdk','argparse','pyarrow'],
entry_script='diabetes_training.py')
Reference:

https://notebooks.azure.com/GraemeMalcolm/projects/azureml-primers/html/04%20-%20Optimizing%20Model%20Training.ipynb

QUESTION 145
You register a file dataset named csv_folder that references a folder. The folder includes multiple comma-separated values (CSV) files in an Azure storage blob container.
You plan to use the following code to run a script that loads data from the file dataset. You create and instantiate the following variables:

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

You have the following code:

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

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

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

A. inputs=[file_dataset.as_named_input('training_files')],
B. inputs=[file_dataset.as_named_input('training_files').as_mount()],
C. inputs=[file_dataset.as_named_input('training_files').to_pandas_dataframe ()],
D. script_params={'--training_files': file_dataset},

Answer: B
Explanation:
Example:
from azureml.train.estimator import Estimator
script_params = {
# to mount files referenced by mnist dataset
'--data-folder': mnist_file_dataset.as_named_input('mnist_opendataset').as_mount(), '--regularization': 0.5
}
est = Estimator(source_directory=script_folder,
script_params=script_params,
compute_target=compute_target,
environment_definition=env,
entry_script='train.py')
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-train-models-with-aml

QUESTION 146
You are creating a new Azure Machine Learning pipeline using the designer.
The pipeline must train a model using data in a comma-separated values (CSV) file that is published on a website. You have not created a dataset for this file.
You need to ingest the data from the CSV file into the designer pipeline using the minimal administrative effort.
Which module should you add to the pipeline in Designer?

A. Convert to CSV
B. Enter Data Manually

C. Import Data
D. Dataset

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

QUESTION 147
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'))
```

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

QUESTION 148
You run an automated machine learning experiment in an Azure Machine Learning workspace. Information about the run is listed in the table below:

Experiment	Run ID	Status	Created on	Duration
auto_ml_clasification	AutoML_1234567890-123	Completed	11/11/2019 11:00:00 AM	00:27:11

You need to write a script that uses the Azure Machine Learning SDK to retrieve the best iteration of the experiment run.
Which Python code segment should you use?

A.
```
from azureml.core import Workspace
from azureml.train.automl.run import AutoMLRun
ws = Workspace.from_config()
automl_ex = ws.experiments.get('auto_ml_classification')
best_iter = automl_ex.archived_time.find('11/11/2019 11:00:00 AM')
```

B.
```
from azureml.core import Workspace
from azureml.train.automl.run import AutoMLRun
automl_ex = ws.experiments.get('auto_ml_classification')
automl_run = AutoMLRun(automl_ex, 'AutoML_1234567890-123')
best_iter = automl_run.current_run
```

C.
```
from azureml.core import Workspace
from azureml.train.automl.run import AutoMLRun
ws = Workspace.from_config()
automl_ex = ws.experiments.get('auto_ml_classification')
best_iter = list(automl_ex.get_runs())[0]
```

D.
```
from azureml.core import Workspace
from azureml.train.automl.run import AutoMLRun
ws = Workspace.from_config()
automl_ex = ws.experiments.get('auto_ml_classification')
automl_run = AutoMLRun(automl_ex, 'AutoML_1234567890-123')
best_iter = automl_run.get_output()[0]
```

E.
```
from azureml.core import Workspace
from azureml.train.automl.run import AutoMLRun
ws = Workspace.from_config()
automl_ex = ws.experiments.get('auto_ml_classification')
best_iter = automl_ex.get_runs('AutoML_1234567890-123')
```

Answer: D
Explanation:
The get_output method on automl_classifier returns the best run and the fitted model for the last invocation. Overloads on get_output allow you to retrieve the best run and fitted model for any logged metric or for a particular iteration.
In [ ]:
best_run, fitted_model = local_run.get_output()
Reference:
https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/automated-machine-learning/classification-with-deployment/auto-ml-classification-with-deployment.ipynb

QUESTION 149
You have a comma-separated values (CSV) file containing data from which you want to train a classification model.
You are using the Automated Machine Learning interface in Azure Machine Learning studio to train the classification model. You set the task type to Classification.
You need to ensure that the Automated Machine Learning process evaluates only linear models.
What should you do?

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

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

QUESTION 150
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:
When you compute model explanations and visualize them, you're not limited to an existing model explanation for an automated ML model. You can also get an explanation for your model with different test data. The steps in this section show you how to compute and visualize engineered feature importance based on your test data.
Incorrect Answers:
A: In the context of machine learning, data drift is the change in model input data that leads to model performance degradation. It is one of the top reasons where model accuracy degrades over time, thus monitoring data drift helps detect model performance issues.
B: A confusion matrix is used to describe the performance of a classification model. Each row displays the instances of the true, or actual class in your dataset, and each column represents the instances of the class that was predicted by the model.

C: Hyperparameters are adjustable parameters you choose for model training that guide the training process. The HyperDrive package helps you automate choosing these parameters.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability-automl

QUESTION 151
You create a multi-class image classification deep learning model that uses the PyTorch deep learning framework. You must configure Azure Machine Learning Hyperdrive to optimize the hyperparameters for the classification model. You need to define a primary metric to determine the hyperparameter values that result in the model with the best accuracy score.
Which three actions must you perform? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

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

Answer: ADF
Explanation:
AD:
primary_metric_name="accuracy",
primary_metric_goal=PrimaryMetricGoal.MAXIMIZE
Optimize the runs to maximize "accuracy". Make sure to log this value in your training script.
Note:
primary_metric_name: The name of the primary metric to optimize. The name of the primary metric needs to exactly match the name of the metric logged by the training script.
primary_metric_goal: It can be either PrimaryMetricGoal.MAXIMIZE or PrimaryMetricGoal.MINIMIZE and determines whether the primary metric will be maximized or minimized when evaluating the runs.
F: The training script calculates the val_accuracy and logs it as "accuracy", which is used as the primary metric.