Question 1

You configure an Azure Machine Learning (ML) workspace for your organization. Your organization wants to further leverage the platform for additional ML projects.

You need to create a compute instance that would support running ML pipeline training using the Azure Machine Learning designer.

Solution: You create a Machine Learning compute cluster.

Does the solution meet the goal?

- No
- Yes

**Explanation**

This solution meets the goal. With a Machine Learning compute cluster, you can easily create a single or multi-node compute instance. A Machine Learning compute cluster is the only type supported by Azure Machine Learning studio to associate a Machine Learning designer training pipeline.

**References**

Set up and use compute targets for model training

Question 2

You configure an Azure Machine Learning (ML) workspace for your organization. Your organization wants to further leverage the platform for additional ML projects.

You need to create a compute instance that would support running ML pipeline training using the Azure Machine Learning designer.

Solution: You create a Azure Databricks environment.

Does this solution meet the goal?

- [x] No
- [ ] Yes

**Explanation**

This solution does not meet the goal. Azure Databricks is an Apache Spark-based environment in the Azure cloud. It can be used as a target with an Azure Machine Learning pipeline, but you cannot set it as a target in Azure Machine Learning designer.

**References**

[Create and run machine learning pipelines with Azure Machine Learning SDK]

[Set up and use compute targets for model training]

Question 3

You configure an Azure Machine Learning (ML) workspace for your organization. Your organization wants to further leverage the platform for additional ML projects.

You need to create a compute instance that would support running ML pipeline training using the Azure Machine Learning designer.

Solution: You create an Azure HDInsight environment.

Does the solution meet the goal?

- Yes
- No

**Explanation**

This solution does not meet the goal. Azure HDInsight is a platform for big-data analytics. It provides Apache Spark, which can be used to train your model, but you cannot set it as a target in Azure Machine Learning designer.

**References**

Set up and use compute targets for model training

Question 4

You have a Python script named sample.py in your local folder. The script trains a model using the SKLearn library, and the script and training data are located in the same folder. You create a compute cluster named train-compute. You plan to run the script on your local machine targeting the compute cluster you created. In your script, you define a local variable named compute_cluster, which is set to reference the train-compute instance in Azure. You have a parameter named local_folder that points to the folder where the scripts are stored.

You need to run the code in sample.py.

Solution: Execute the following code to run the scikit-learn estimator:

```
from azureml.train.sklearn import SKLearn
script_params = {
'--kernel': 'linear'
'--penalty': 1.0
}
sklearn01 = SKLearn(source_directory=local_folder,
script_params=script_params
compute_target=compute_cluster,
entry_script='sample.py')
```

Does the solution meet the goal?

- No
- Yes

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

**References**

SKLearn class
Build scikit-learn models at scale with Azure Machine Learning
Train and hyperparameter tune on Iris Dataset with Scikit-learn

Question 5

You have a Python script named sample.py in your local folder. The script trains a model using the SKLearn library, and the script and training data are located in the same folder.

You create a compute cluster named train-compute. You plan to run the script on your local machine targeting the compute cluster you created. In your script, you define a local variable named compute_cluster, which is set to reference the train-compute instance in Azure. You have a parameter named local_folder that points to the folder where the scripts are stored.

You need to run the code in sample.py.

Solution: Execute the following code to run the TensorFlow estimator.

```
from azureml.train.dnn import TensorFlow
script_params = {
 '--data-folder':
dataset.as_named_input('mnist').as_mount(),
 '--batch-size': 50,
 '--first-layer-neurons': 300,
 '--second-layer-neurons': 100,
 '--learning-rate': 0.01
}
sklearn01 = TensorFlow(source_directory=local_folder,
script_params=script_params
compute_target=compute_cluster,
entry_script='sample.py')
```

Does the solution meet the goal?

- No
- Yes

This solution does not meet the goal. The TensorFlow estimator will not leverage the scikit-learn estimator. The code will invoke the TensorFlow estimator to execute the Python training model.

**References**


Build a TensorFlow deep learning model at scale with Azure Machine Learning

SKLearn class

Build scikit-learn models at scale with Azure Machine Learning

Train and hyperparameter tune on Iris Dataset with Scikit-learn

Question 6

You have a Python script named sample.py in your local folder. The script trains a model using the SKLearn library, and the script and training data are located in the same folder.

You create a compute cluster named train-compute. You plan to run the script on your local machine targeting the compute cluster you created. In your script, you define a local variable named compute_cluster, which is set to reference the train-compute instance in Azure. You have a parameter named local_folder that points to the folder where the scripts are stored.

You need to run the code in sample.py.

Solution: Execute the following code to run the PyTorch estimator.

```
from azureml.train.dnn import PyTorch
script_params = {
     '--num_epochs': 30,
     '--output_dir': './outputs'
}
estimator = PyTorch(source_directory=local_folder,
                    script_params=script_params,
                    compute_target=compute_cluster,
                    entry_script='sample.py',
                    use_gpu=True,
                    pip_packages=['pillow==5.4.1'])
```

Does the solution meet the goal?

- No
- Yes

This solution does not meet the goal. The PyTorch estimator will not leverage the scikit-learn estimator. The code will invoke the PyTorch estimator to execute the Python training model.

**References**

Train models with Azure Machine Learning using estimator

SKLearn class

Build scikit-learn models at scale with Azure Machine Learning

Train and hyperparameter tune on Iris Dataset with Scikit-learn

Question 7

You need to submit a pipeline run that has the following requirements:

- Continue execution, even if a step in the pipeline fails.
- Create the outputs of all steps again for every run, even if outputs are not stale.
- Pass a parameter named param1 with a value of value1 to the pipeline.
- Not refer to any other run id.

Your experiment already has an additional pipeline run, with run ids 1, 2, and 3, respectively. You want to add this run to the same experiment.
Solution: Run the following script:

```
from azureml.pipeline.core import Pipeline

pipeline = Pipeline(workspace=ws, steps=steps)
pipeline_run = experiment.submit(pipeline,
                                    continue_on_step_fail
ure=True,
                                    regenerate_outputs=Tr
ue,
                                    pipeline_parameters={
"param1": "value1"},
                                    parent_run_id="3")
```

Does this solution meet the goal?

No

Yes

## Explanation

This solution does not meet the goal. The script passes the parent_run_id. You need to run a script that does not refer to any other run id.

## References

[Pipeline class]

Question 8

You need to submit a pipeline run that has the following requirements:

- Continue execution, even if a step in the pipeline fails.
- Create the outputs of all steps again for every run, even if outputs are not stale.
- Pass a parameter named param1 with a value of value1 to the pipeline.
- Not refer to any other run id.

Your experiment already has an additional pipeline run, with run ids 1, 2, and 3, respectively. You want to add this run to the same experiment.

Solution: Run the following script:

```
from azureml.pipeline.core import Pipeline

pipeline = Pipeline(workspace=ws, steps=steps)
pipeline_run = experiment.submit(pipeline,
                                 continue_on_step_fail
ure=True,
                                 regenerate_outputs=Tr
ue,
                                 pipeline_parameters={
"param1": "value1"})
```

Does this solution meet the goal?

- [x] Yes
- [ ] No

This solution meets the goal. In the experiment's submit method, you set continue_on_step_failure to True to ensure that pipeline execution continues, even if a step fails. You set regenerate_outputs to True, because you need to create outputs of the steps every time the pipeline is executed. You also pass the parameter and its value as a dictionary object to the pipeline_parameters parameter.

**References**

[Pipeline class]

Question 9

You need to submit a pipeline run that has the following requirements:

- Continue execution, even if a step in the pipeline fails.
- Create the outputs of all steps again for every run, even if outputs are not stale.
- Pass a parameter named param1 with a value of value1 to the pipeline.
- Not refer to any other run id.

Your experiment already has an additional pipeline run, with run ids 1, 2, and 3, respectively. You want to add this run to the same experiment.

Solution: Run the following script:

```
from azureml.pipeline.core import Pipeline

pipeline = Pipeline(workspace=ws, steps=steps)
pipeline_run = experiment.submit(pipeline,
                                    continue_on_step_fail
ure=True,
                                    pipeline_parameters={
"param1": "value1"})
```

Does this solution meet the goal?

- [ ] Yes
- [x] No

**Explanation**

This solution does not meet the goal. You should not submit the experiment without setting the regenerate_outputs parameter. The default value for this parameter is False. The requirements state that outputs be regenerated on every run of the pipeline.

## References

[Pipeline class]

## Question 10

You want to use the interpretability package to explain Machine Learning (ML) models and predictions in Python. You run the following code to install the interpretability package on your personal machine:

```
pip install azureml-interpret
pip install azureml-contrib-interpret
```

Next, you train a sample model in a local Jupyter notebook. You want to call the explainer locally.

You need to write your code to call SHAP explainers and leverage the most appropriate one for the dataset you are running ML on.

Solution: You run the following code to initialize an explainer object and to pass your model and some training data:

```
from interpret.ext.blackbox import TabularExplainer
explainer = TabularExplainer(model, x_train,
                             features=breast_cancer_data.
feature_names,
                             classes=classes)
```

Does the solution meet the goal?

- Yes
- No

**Explanation**

This solution meets the goal. TabularExplainer calls one of the three SHAP explainers (TreeExplainer, DeepExplainer, or KernelExplainer). TabularExplainer automatically selects the most appropriate one for your use case.

**Reference :** Use the interpretability package to explain ML models & predictions in Python (preview) | SHAP

Question 11

You want to use the interpretability package to explain Machine Learning (ML) models and predictions in Python.
You run the following code to install the interpretability package on your person machine:

```
pip install azureml-interpret
pip install azureml-contrib-interpret
```

Next, you train a sample model in a local Jupyter notebook. You want to call the explainer locally.

You need to write your code to call SHAP explainers and leverage the most appropriate one for the dataset you are running ML on.

Solution: You run the following code to initialize an explainer object and to pass your model and some training data:

```
from interpret.ext.blackbox import MimicExplainer
explainer = MimicExplainer(model, x_train,
                           LGBMExplainableModel,
                           augment_data=True,
                           max_num_of_augmentations=10,
                           features=breast_cancer_data.fe
ature_names,
                           classes=classes)
```

Does the solution meet the goal?

- Yes
- No

**Explanation**

This solution does not meet the goal. MimicExplainer does not select one of the SHAP explainers. You have to explicitly pass the model in the initialization process of the explainer.

**References**

[Use the interpretability package to explain ML models & predictions in Python (preview)]

Question 12

You want to use interpretability package to explain Machine Learning (ML) models and predictions in Python.

You run the following code to install the interpretability package on your person machine:

```
pip install azureml-interpret
pip install azureml-contrib-interpret
```

Next, you train a sample model in a local Jupyter notebook. You want to call the explainer locally.

You need to write your code to call SHAP explainers and leverage the most appropriate one for the dataset you are running ML on.

Solution: You run the following code to initialize an explainer object and to pass your model and some training data:

```
from interpret.ext.blackbox import PFIExplainer

explainer = PFIExplainer(model,
                         features=breast_cancer_data.feat
ure_names,
                         classes=classes)
```

Does the solution meet the goal?

- [x] No
- [ ] Yes

**Explanation**

This solution does not meet the goal. MimicExplainer does not select one of the SHAP explainers. You have to explicitly pass the model in the initialization process of the explainer.

**References**

[Use the interpretability package to explain ML models & predictions in Python (preview)]

Question 13

You want to deploy your trained model to a production environment. You want to ensure that once the web service endpoints are deployed, only authorized users are able to reach the service using the token auth method. You decide to use Azure Kubernetes Service (AKS) to deploy your model.

You need to write the script so that when it is executed, the token auth method is enabled.

Solution: Execute AksWebservice.deploy_configuration(token_auth_enabled=True, auth_enabled=False)

Does this solution meet the goal?

- No
- Yes

**Explanation**

This solution meets the goal. You should call the AksWebservice.deploy_configuration method with the token_auth_enabled parameter set to True. You then need to set the auth_enabled parameter to False, because both token and key-based authentication cannot be enabled at the same time. By default, the auth_enabled parameter is set to True to provide key authentication.

**References**

Deploy a model to an Azure Kubernetes Service cluster

AksWebservice class

Question 14

You want to deploy your trained model to a production environment. You want to ensure that once the web service endpoints are deployed, only authorized users are able to reach the service using the token auth method. You decide to use Azure Kubernetes Service (AKS) to deploy your model.

You need to write the script so that when it is executed, the token auth method is enabled.

Solution: Execute AksWebservice.deploy_configuration(auth_enabled=True)

Does this solution meet the goal?

- Yes
- No

**Explanation**

This solution does not meet the goal. Calling AksWebservice.deploy_configuration with the auth_enabled parameter set to True will enable key-based authentication. It will not enable token based authentication.

**References**

Deploy a model to an Azure Kubernetes Service cluster

AksWebservice class Definition

Question 15

You want to deploy your trained model to a production environment. You want to ensure that once the web service endpoints are deployed, only authorized users are able to reach the service using the token auth method. You decide to use Azure Kubernetes Service (AKS) to deploy your model.

You need to write the script so that when it is executed, the token auth method is enabled.

Solution: Execute AksWebservice.deploy_configuration()

Does this solution meet the goal?

- Yes
- No

**Explanation**

This solution does not meet the goal. Calling AksWebservice.deploy_configuration without any parameters set will enable key-based authentication by default. It will not enable token-based authentication.

**References**

[Deploy a model to an Azure Kubernetes Service cluster]()

[AksWebservice class]()

Question 16

```
from azureml.core import Experiment
from azureml.core import Model
import pandas as pd
import numpy as np
import joblib
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import roc_auc_score
from sklearn.metrics import roc_curve

experiment = Experiment(workspace = ws, name = "diabetes-training")
diabetes = pd.read_csv('data/diabetes.csv')

X, y = diabetes[['Pregnancies','PlasmaGlucose','DiastolicBloodPressure',
                 'TricepsThickness','SerumInsulin','BMI',
                 'DiabetesPedigree','Age']].values,
                 diabetes['Diabetic'].values

X_train, X_test, y_train, y_test =
            train_test_split(X, y, test_size=0.30, random_state=0)

model = DecisionTreeClassifier().fit(X_train, y_train)
model_file = 'diabetes_model.pkl'
joblib.dump(value=model, filename=model_file)
run.upload_file(name = 'outputs/' + model_file, path_or_stream = './' +
model_file)

run.complete()

run.register_model(model_path='outputs/diabetes_model.pkl',
model_name='diabetes_model',
                    tags={'Training context':'Inline Training'},
                    properties={'AUC': run.get_metrics()['AUC'],
                             'Accuracy': run.get_metrics()['Accuracy']})
```

You design and train a classification model to categorize individuals that are at a high risk of developing diabetes based on various features of the individual in question.

The code you use to train and register the model is shown in Exhibit 1.

You want to deploy the model as a web service using Azure Machine Learning SDK. You create a scoring script called score_diabetes.py and write the code below to deploy the model:

```
from azureml.core.webservice import AciWebservice
from azureml.core.model import InferenceConfig

inference_config = InferenceConfig(runtime= "python",
                                   source_directory =
folder_name,
                                   entry_script="score_di
abetes.py")

deployment_config =
AciWebservice.deploy_configuration(cpu_cores = 1,
memory_gb = 1)
service_name = "diabetes-service"
service = Model.deploy(ws, service_name, [model],
inference_config, deployment_config)
```

The web service fails to deploy.

You need to fix the problem and redeploy.

Solution: You run the following script:

```
from azureml.core.webservice import AKSWebservice
from azureml.core.model import InferenceConfig

inference_config = InferenceConfig(runtime= "python",
                                   source_directory =
folder_name,
                                   entry_script="score_di
abetes.py")

deployment_config =
AKSWebservice.deploy_configuration(cpu_cores = 1,
memory_gb = 1)
service_name = "diabetes-service"
service = Model.deploy(ws, service_name, [model],
inference_config, deployment_config)
```

Does the solution meet the goal?

- Yes
- No

**Explanation**

This solution does not meet the goal. The script will not deploy the service because it is not deployed in a Azure Container Instance. You need to deploy a web service in an Azure Container Instance or an Azure Kubernetes Service.

## References

[Deploy a model to Azure Machine Learning compute instances]

[Creating a Real-Time Inferencing Service]

Question 17

You design and train a classification model to categorize individuals that are at a high risk of developing diabetes based on various features of the individual in question.

The code you use to train and register the model is shown in Exhibit 1.

You want to now deploy the model as a web service using Azure Machine Learning SDK. You create a scoring script called score_diabetes.py and write the code below to deploy the model:

```
from azureml.core.webservice import AciWebservice
from azureml.core.model import InferenceConfig

inference_config = InferenceConfig(runtime= "python",
                                   source_directory =
folder_name,
                                   entry_script="score_di
abetes.py")

deployment_config =
AciWebservice.deploy_configuration(cpu_cores = 1,
memory_gb = 1)
service_name = "diabetes-service"
service = Model.deploy(ws, service_name, [model],
inference_config, deployment_config)
```

The web service fails to deploy.

You need to fix the problem and redeploy.

Solution: You run the following script:

```
from azureml.core.webservice import AciWebservice
from azureml.core.model import InferenceConfig

inference_config = InferenceConfig(runtime= "python",
                                   source_directory =
folder_name,
```

```
                                                              entry_script="score_di
abetes.py")

deployment_config =
AciWebservice.deploy_configuration(cpu_cores = 4,
memory_gb = 16)
service_name = "diabetes-service"
service = Model.deploy(ws, service_name, [model],
inference_config, deployment_config)
```

Does the solution meet the goal?

- [x] No
- [ ] Yes

**Explanation**

This solution does not meet the goal. Increasing the compute specifications is not required. This might improve the run times, but it will not fix the error.

**References**

[Deploy a model to Azure Machine Learning compute instances]

[Creating a Real-Time Inferencing Service]

Question 18

You design and train a classification model to categorize individuals that are at a high risk of developing diabetes based on various features of the individual in question.

The code you use to train and register the model is shown Exhibit 1.

You want to now deploy the model as a web service using Azure Machine Learning SDK. You create a scoring script called score_diabetes.py and write the code below to deploy the model:

```
from azureml.core.webservice import AciWebservice
from azureml.core.model import InferenceConfig

inference_config = InferenceConfig(runtime= "python",
                                   source_directory =
folder_name,
                                   entry_script="score_di
abetes.py")

deployment_config =
AciWebservice.deploy_configuration(cpu_cores = 1,
memory_gb = 1)
service_name = "diabetes-service"
service = Model.deploy(ws, service_name, [model],
inference_config, deployment_config)
```

The web service fails to deploy.

You need to fix the problem and redeploy.

Solution: You run the following script:

```
from azureml.core.conda_dependencies import
CondaDependencies
from azureml.core.webservice import AciWebservice
from azureml.core.model import InferenceConfig
```

```
myenv = CondaDependencies()
myenv.add_conda_package("scikit-learn")

env_file = folder_name + "/diabetes_env.yml"
with open(env_file,"w") as f:
        f.write(myenv.serialize_to_string())

inference_config = InferenceConfig(runtime= "python",
                                   source_directory =
folder_name,
                                   entry_script="score_di
abetes.py")

deployment_config =
AciWebservice.deploy_configuration(cpu_cores = 1,
memory_gb = 1)
service_name = "diabetes-service"
service = Model.deploy(ws, service_name, [model],
inference_config, deployment_config)
```

Does the solution meet the goal?

- No
- Yes

**Explanation**

This solution meets the goal. The model leverages the SKLearn library in the training process. The web service is hosted in a container, and the container will need to install scikit-learn as a Python dependency. You have to create the necessary YML file and deploy the dependent package.

**References**

[Deploy a model to Azure Machine Learning compute instances]

[Creating a Real-Time Inferencing Service]

Question 19

You publish an Azure Machine Learning (ML) designer service with a REST endpoint using default options. The service provides access to a batch inferencing model. The endpoint for the service is stored in a variable rest_endpoint.

You need to consume the service in one of your experiments.
Solution: You write the following script:

```
from azureml.core.authentication import
InteractiveLoginAuthentication
import requests

interactive_auth = InteractiveLoginAuthentication()
auth_header =
interactive_auth.get_authentication_header()

response = requests.post(rest_endpoint,
                         headers=auth_header,
                         json={"ExperimentName":
"Batch_Pipeline_via_REST"})
run_id = response.json()["Id"]
```

Does the solution meet the goal?

- No
- Yes

**Explanation**

This solution meets the goal. In order to access an ML designer service using the rest endpoint, you would first get an authentication token to the service. Once you are authenticated, you can pass the auth_header generated as a result of the interactive login process. The request.post method can be used to make REST calls to the published service.

**References**

[Creating a Batch Inferencing Service](#)
[InteractiveLoginAuthentication class](#)

Question 20

You publish an Azure Machine Learning (ML) designer service with a REST endpoint using default options. The service provides access to a batch inferencing model. The endpoint for the service is stored in a variable rest_endpoint.

You need to consume the service in one of your experiments.

Solution: You write the following script:

```
import requests

response = requests.post(rest_endpoint,
                         headers={'content_type':'JSON'},
                         json={"ExperimentName":
"Batch_Pipeline_via_REST"})
run_id = response.json()["Id"]
```

Does the solution meet the goal?

- Yes
- No

## Explanation

This solution does not meet the goal. The call only works if there is no authentication required by the service. Since the service was deployed using Azure ML designer, the default configuration requires an authentication header to be passed as the headers parameter for the requests.post method.

## References

[Creating a Batch Inferencing Service]

[InteractiveLoginAuthentication class Definition]

Question 21

You publish an Azure Machine Learning (ML) designer service with a REST endpoint using default options. The service provides access to a batch inferencing model. The endpoint for the service is stored in a variable rest_endpoint.

You need to consume the service in one of your experiments.

Solution: You write the following script:

```
from azureml.core.authentication import
InteractiveLoginAuthentication
import requests

interactive_auth = InteractiveLoginAuthentication()
response = requests.post(rest_endpoint,
                         json={"ExperimentName":
"Batch_Pipeline_via_REST"})
run_id = response.json()["Id"]
```

Does the solution meet the goal?

- Yes
- No

**Explanation**

This solution does not meet the goal. The call only works if there is no authentication required by the service. Even though a call to interactively login is made using the InteractiveLoginAuthentication method, the auth token generated is not passed as part of the call to the rest endpoint: the requests.post method call. Since the service was deployed using Azure ML designer, the default configuration requires an authentication header to be passed as the headers parameter for the requests.post method.

**References**

[Creating a Batch Inferencing Service]

[InteractiveLoginAuthentication class]

Question 22

You are creating an automated machine learning experiment that generates models that are used to identify faces in images. You create the AutoMLConfig object listed below.

```
automl_experiment = AutoMLConfig(
task='classification',
primary_metric: 'spearman_correlation',
debug_log='experiment.log',
training_data=imgs_faces,
label_column_name="identity",
)
```

You need to ensure that your experiment functions properly.

What should you do?

Choose the correct answer

- Replace the task='classification' line with task='regression'.
- Add the experiment_timeout_hours=24 line.
- Add the n_cross_validations=3 line.
- Remove the primary_metric: 'spearman_correlation' line.

**Explanation**

You should remove the primary_metric: 'spearman_correlation' line. Automated machine learning in Azure Machine Learning uses the primary metric to optimize model training. The metrics you can configure are dependent on the machine learning task type, such as regression or classification. In this scenario, you will use image classification to identify faces in images. However, the specified primary metric, spearman_correlation, is used for regression tasks and works on numeric and logical data only. If the primary_metric line is removed, image classification experiments will use the accuracy metric by default. The accuracy metric calculates the proportion of instances that have been correctly classified.

You should not add the experiment_timeout_hours=24 line. This parameter specifies the maximum time an experiment will be allowed to run. The default timeout is six days.

You should not replace the task='classification' line with task='regression'. This experiment will perform image classification. Regression is used in machine learning to identify relationships between data.

You should not add the n_cross_validations=3 line. Cross-validation is a machine learning technique designed to verify dataset variability as well as the reliability of models trained using that dataset. In practice, cross-validation divides a dataset into folds and builds a model for each fold. Accuracy statistics for each fold are then compared.

**References**

Train models with automated machine learning in the cloud

AutoMLConfig class

Configure automated ML experiments in Python

Cross Validate Model

Question 23

You create Azure Machine Learning workspaces for different projects.

You need to create each workspace using the minimum edition required for each project.

Based on each stated requirement, which Azure Machine Learning edition should you provision?

Choose the correct options

| Requirement | Edition |
|---|---|
| Use notebooks to create and run experiments. | |
| Create pipelines using Azure Machine Learning Designer. | |
| Share compute instances. | |

Basic or Enterprise

Select your answ er

Select your answ er

## Explanation

Azure Machine Learning workspaces can be provisioned as one of two editions, Basic or Enterprise, and the edition determines which features will be available for a workspace. You define the machine learning workspace edition when the workspace is created. You can also upgrade any existing Basic edition to Enterprise edition at any time.

Basic or Enterprise edition can use notebooks to create and run experiments. A notebook is essentially a development environment that has been specialized for data science tasks. Unlike a traditional object-oriented coding environment, notebooks support features such as data visualization and equation writing. The most popular notebook application is Jupyter Notebook, which supports popular data science programming languages such as Python and R.

Enterprise edition is required to create pipelines using Azure Machine Learning Designer. Azure Machine Learning pipelines are workflows that represent a series of machine learning tasks. Pipeline tasks can be executed independently from the underlying data, and you can register new datasets for each pipeline run, if necessary. While Enterprise edition is required for this task, you can still create and manage pipelines using Azure Machine Learning SDK with the Basic edition.

Basic or Enterprise edition can be used to share compute instances. A compute instance is a single Azure-homed virtual machine (VM). Azure Machine Learning compute instances are highly scalable cloud compute resources, which support multiple CPUs and large amounts of RAM based on the VM size you select at deployment.

**References**

Azure Machine Learning pricing

Enterprise and Basic Editions of Azure Machine Learning

Jupyter Notebook: An Introduction

What is Azure Machine Learning designer (preview)?

What are compute targets - Azure Machine Learning?

Question 24

You use Azure Machine Learning to create a machine learning pipeline. Your dataset includes sparse string and numeric data. While working with pipeline modules, you receive an error indicating that a value is required.

You need to resolve this issue.

What should you do?

Choose the correct answer

- Use the Select Columns in Dataset module to choose a column.
- Specify the columns to be cleaned in the Clean Missing Data module.
- Configure a custom substitution value in the Clean Missing Data module.
- Configure the Select Columns in Dataset module to exclude string data.

**Explanation**

You should choose a column using the Select Columns in Dataset module. Azure Machine Learning pipeline modules can validate, analyze, and transform your data to ensure your models are accurate. As you link pipeline modules, you control what data flows from one module to the next in your workflow. You can use the Select Columns in Dataset module to identify the columns that should be processed and sent to the next pipeline element. By default, no columns are selected, and you will receive an error indicating that a value is required until you select at least one column.

You should not configure the Select Columns in Dataset module to exclude string data. The Select Columns in Dataset module supports several methods for identifying columns. You can select columns by name, type, or column index. As part of this process, you can define conditions that can filter your data based on type. For example, you could create a condition to exclude all string column types.

You should not specify the columns to be cleaned in the Clean Missing Data module. You use the Clean Missing Data module to ensure that your data is as complete as possible prior to machine learning processing. When configuring this module, you need to specify the columns that contain missing values that you need to modify.

You should not configure a custom substitution value in the Clean Missing Data module. A substitution value is an explicit value that is used to replace missing values. For example, you may replace missing values with a zero.

**References**

What is Azure Machine Learning designer (preview)?

Select Columns in Dataset

Clean Missing Data module

Question 25

You complete training a linear regression model using Azure Machine Learning (ML) studio. You want to test the model by having users of your organization call an endpoint. The outcomes of the model should be provided in real time for users. All workload will be CPU based.

You need to determine a deployment target compute resource that can be used for testing and debugging while incurring minimal cost.
What two deployment targets should you select? Each correct answer presents a complete solution.

Choose the correct answers

- [ ] Azure Kubernetes Service (AKS)
- [ ] Local web service
- [ ] Azure ML compute clusters
- [ ] Azure Container Instances (ACI)

**Explanation**

You should select Local web service or Azure Container Instances (ACI). Both deployment targets provide low cost instances that can be used for testing and debugging CPU based workloads.

You should not select AKS. AKS is used for production workloads. AKS provides fast response times and autoscaling of the deployed service, but it is costly compared to ACI and local web service. AKS will not be suitable because the instance is for testing and debugging,

You should not select Azure ML compute clusters. Azure ML compute clusters are suitable for batch inference pipelines. Here the service needs to be deployed as a real-time service.

**References**

[What are compute targets in Azure Machine Learning?]

Question 26

You are preparing to register multiple trained models using Azure Machine Learning.

You need to identify the most appropriate compute targets that will be used for each model.

Which deployment target should you implement based on each model's usage scenario? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

| Model Usage Scenario | Deployment Target |
|---|---|
| Batch inference | |
| Real-time inference | |
| Testing and debugging | |

- Azure Machine Learning compute cluster
- Azure Kubernetes Service (AKS)
- Azure Machine Learning compute instance w eb service

**Explanation**

You should deploy to an Azure Machine Learning compute cluster target for models that perform batch inference. In Azure Machine Learning, inference is also known as model scoring. Azure Machine Learning compute clusters are scalable machine learning platforms consisting of one or more CPU or GPU nodes. Compute clusters can scale from zero to hundreds of nodes, depending on workload. Compute clusters support the use of low-priority virtual machines (VMs), which do not have guaranteed availability. Using low-priority VMs can help reduce machine learning costs.

AKS target could theoretically be used to perform batch inference. However, AKS is designed for compute-intensive operations at scale, while batch inference requires compute resources on an intermittent basis. By contrast, batch inference requires the scalability that an Azure Machine Learning compute instance web service would not provide.

You should deploy to AKS target for models that perform real-time inference. In Azure Machine Learning, inference is also known as model scoring. Such models are trained on a data set and can analyze the data in real-time to provide predictions. For example, you could train a model on stock market data. Once trained, you could use the model to analyze stock prices in real-time and then make predictions on future prices. Inference clusters are built using Azure AKS and are sometimes referred to AKS clusters.

Azure machine learning compute clusters cannot be used for real-time inference because they use low-priority VMs and may not scale to the load required. Additionally, VM availability is not guaranteed, thus the provided service is not real-time. Similarly, Azure Machine Learning compute instance web services cannot be used for real-time inference because they typically lack hardware acceleration capabilities and do not scale to the workloads involved in real-time inference.

You should deploy to an Azure Machine Learning compute instance web service target for models that need to be tested and debugged. Azure Machine Learning compute instances are highly scalable cloud compute resources. Compute instances support AutoML and machine learning pipelines.

**References**

[Deploy models with Azure Machine Learning]

[What are compute targets in Azure Machine Learning?]

Question 27

You use a published Azure Machine Learning pipeline to predict which customers should be targeted for marketing a new service offering.

New customer information has been collected via survey, and you need to tune your model using this data.

What should you do?

Choose the correct answer

- Change the pipeline concurrency parameter to 2.
- Create a new training pipeline. Specify a new dataset.
- Reuse the pipeline. Change the dataset and parameters.
- Convert the training pipeline into a real-time inference pipeline.

**Explanation**

You should reuse the pipeline and change the dataset and parameters. Azure Machine Learning pipelines are workflows that represent a series of machine learning tasks. Pipeline tasks can be executed independently from the underlying data, and you can register new datasets for each pipeline run, if necessary. In this scenario, as you already have a functional prediction model and you only want to tune the model using new data, you should reuse the pipeline.

You should not create a new training pipeline and specify a new dataset. The same pipeline can be used to train multiple models. One of the benefits of defining a machine learning pipeline is its reusability. You would create a new pipeline if the required machine learning workflow had changed, and you did not want to modify an existing workflow.

You should not convert the training pipeline into a real-time inference pipeline. In Azure Machine Learning, inference is also known as model scoring. Such models are trained on a dataset and can then analyze data in real time to provide predictions. For example, you could train a model on stock market data. Once trained, you could use the model to analyze stock prices in real-time and then make predictions on future prices.

You should not change the pipeline concurrency parameter to 2. Pipeline concurrency allows you to specify the maximum number of concurrent pipeline runs. Any runs above the concurrency limit are queued until a run space is freed.

**References**

What is Azure Machine Learning designer (preview)?

Pipelines and activities in Azure Data Factory

Question 28

You configure an Azure Machine Learning experiment using the following code:

```
import azureml.core
from azureml.core import Workspace
from azureml.core import Experiment
ws = Workspace.from_config()
script_params = ["--experiment_ouput", experiment_ouput]
exp = Experiment(ws, experiment_name)
```

You need to ensure output files are uploaded in real time.

Which line of code should you add to your script?

Choose the correct answer

- experiment_ouput = os.path.join(os.curdir, "outputs")
- run.get_file_names()
- run.log(“experiment_output”,0)
- experiment_ouput = os.path.join(os.curdir, "logs")

**Explanation**

You should add the experiment_ouput = os.path.join(os.curdir, "logs") line of code. Azure Machine Learning supports several locations for storing experiment output. Files can be saved to storage on the local compute instance, but these files do not persist across training runs. To store files for later analysis and review, you should use an Azure Machine Learning datastore, or you should write to the outputs or logs folders. Files written to the ./logs folder are uploaded in real time.

You should not use the experiment_ouput = os.path.join(os.curdir, "outputs") line of code. Files written to the outputs folder persist across experiments, but they are not uploaded in real time.

You should not use the run.log(“experiment_output”,0) line of code. This method can be used to log scalar string or numerical values. This method does not upload output files.

You should not use the run.get_file_names() line of code. This method is used to list all files that have been stored by the training run.

**References**

Where to save and write files for Azure Machine Learning experiments

Introduction to datastores

Run class

Question 29

You use Azure Machine Learning to create and publish a batch inference pipeline. To meet regulatory requirements, your pipeline endpoint has been scanned for vulnerabilities.

You want to test some changes, but you need to ensure that the existing pipeline is not modified. You also need to ensure that the pipeline endpoint is not changed.

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

Create a list in the correct order

**Actions in order**

- Create a new pipeline that implements your changes.
- Publish the pipeline to your scanned endpoint.
- Include the test pipeline version in your REST calls.

**Explanation**

You should perform the following actions in order:

1. Create a new pipeline that implements your changes.
2. Publish the pipeline to your scanned endpoint.
3. Include the test pipeline version in your REST calls.

In Azure Machine Learning, inference is also known as model scoring. Such models are trained on a dataset and can analyze data in real-time or on-demand to provide predictions. A batch inference pipeline is used to score datasets on-demand via web service.

This web service is defined when you publish a pipeline, and an HTTP endpoint that external applications and services can consume is created. Azure Machine Learning allows you to publish multiple pipelines under the same endpoint. When this occurs, each published pipeline is assigned to a version number, which can be provided in REST API calls.

In this question, you want to preserve the existing pipeline, so you should create a new pipeline that implements your changes. You can then publish this pipeline under your existing endpoint. Finally, you can test you new pipeline by including its version in your REST calls.

You should not determine the ID of the default published pipeline. Every endpoint has one default pipeline. When you publish a new pipeline under an existing endpoint, you can choose to make it the default pipeline for that endpoint.

**References**

Run batch predictions using Azure Machine Learning designer (preview)

Run batch inference on large amounts of data by using Azure Machine Learning

Question 30

You want to use your local computer as a development environment to work with Azure Machine Learning (ML). You have decided to use Anaconda within your local environment. You want to leverage Automated ML to turn Hyperparameters for your training pipelines.

You have downloaded and installed Anaconda with Python 3.7 version. You are at the Anaconda prompt.

You need to create the environment to work with Azure ML.

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

Create a list in the correct order

**Possible commands**

**Commands in order**

- conda create -n devenv python=3.7.7
- conda activate devenv
- conda install notebook ipykernel
- ipython kernel install --user --name devenv --display-name "Python (devenv)"
- pip install azureml-sdk[notebooks,automl]

**Explanation**

You should execute the commands in the following order:

1. conda create -n devenv python=3.7.7
2. conda activate devenv
3. conda install notebook ipykernel
4. ipython kernel install --user --name devenv --display-name "Python (devenv)
5. pip install azureml-sdk[notebooks,automl]

First you should execute conda create -n devenv python=3.7.7 to create the environment. This code leverages python 3.7.7.

Next you should execute conda activate devenv to activate the environment.

Then, once activated, you should execute conda install notebook ipykernel and ipython kernel install --user --name devenv --display-name "Python (devenv) to enable the ipykernel and create the kernel in your environment.

Finally, you should execute pip install azureml-sdk[notebooks,automl] to install the azureml sdk and automl extras to run automated ML hyperparameter tuning jobs.

You should not execute Jupyter Notebook. This command is used when you want to launch a notebook. It is not required to configure the development environment.

You should not execute conda activate AzureML. You would first have to run pip install because AzureML is not present by default in an Anaconda environment. If you leverage a Data Science Virtual Machine (DSVM), which comes with AzureML installed, you can run this command to activate it.

### References

Configure a development environment for Azure Machine Learning

Question 31

You manage an Azure Machine Learning workspace.

You need to send an HTTP push notification to an external system when a machine learning model is registered or deployed in the workspace.

What should you do?

Choose the correct answer

- Stream Azure Machine Learning metric information to Azure Event Hub.
- Create an event subscription and set the endpoint type to web hook.
- Deploy a real-time endpoint and specify a compute target.
- Create a service principal and grant it access to your workspace.

**Explanation**

You should create an event subscription and set the endpoint type to web hook. An Azure Machine Learning workspace is tightly integrated with Azure Event Grid. While Azure Event Hub is designed for ingesting large amounts of data that can be used to glean business intelligence, Event Grid is focused on discrete events generated by applications and aids in application automation.

Event Grid supports several event endpoint destinations, including web hooks. Webhooks enable applications to receive real-time data from a server or service. When a webhook is triggered, Event Grid sends a notification to a preregistered Uniform Resource Identifier (URI) using HTTP.

You should not deploy a real-time endpoint and specify a compute target. A real-time endpoint is the port-to-service mapping that is created when you deploy a web service. As part of deploying a real-time endpoint, you are required to specify a compute target. Once the real-time endpoint has been deployed, applications and services can consume the endpoint in the same way as they would any other REST API.

You should not create a service principal and grant it access to your workspace. A service principal (SP) is any directory object that can be used for authentication. Once an SP is created, it can be used in Azure Machine Learning to facilitate token-based authentication.

You should not stream Azure Machine Learning metric information to Azure Event Hub. Azure Event Hub is a stand-alone platform that, like Azure Monitor, can ingest logging and other information from a variety of Azure Services. However, Event Hub is focused on data analysis to discover actionable insights, sometimes referred to as business intelligence.

**References**

What is Azure Event Grid?

Tutorial: Deploy a machine learning model with the designer (preview)

Set up authentication for Azure Machine Learning resources and workflows

Event Hubs

Question 32

You are evaluating Azure Machine Learning to run training scripts. To configure and run your experiment, you run the following code:

```
from azureml.core import Workspace, RunConfig,
ScriptRunConfig, Experiment, Run

ws = Workspace.from_config()
src = ScriptRunConfig(source_directory='./',
script='global-warming.py',
run_config=RunConfiguration())
experiment = Experiment(workspace=ws, name='global-
warming-experiment')
run = experiment.submit(config=src)
run.wait_for_completion()
```

You need to add code to show the output files that are generated by the experiment.

Which code line should you add?

Choose the correct answer

- files = run.get_file_names()
- files = run.get_details_with_logs()
- files = run.get_status()
- files = run.get_details()

**Explanation**

You should add a call to the get_file_names() function. This function will list all the files that are associated with the experiment run object.

You should not use files=run.get_status(). The get_status() function is used to fetch the status of the last run.

You should not use files=run.get_details_with_logs(). This function is used to get the status of the last run, along with log file contents.

You should not use files=run.get_details(). This function is used to get the definition, status information, current log files, and other details of the run.

**References**

[RunConfiguration class]

[Enable logging in Azure ML training runs]

[Run class]

Question 33

You train and register an infection detection inference model in Azure Machine Learning (ML) studio. The model is used to predict outcomes on a large volume of data files.

You need to create a batch inferencing pipeline to process these large files that are stored in various storage accounts. You want your pipeline to run the scoring script on multiple nodes within your compute cluster and collate the results.

Which step should you use?

Choose the correct answer

- ParallelRunStep
- ParallelRunConfig
- PythonScriptStep
- AdlaStep

**Explanation**

You should use ParallelRunStep to run the scoring script in your pipeline. ParallelRunStep can be used to process large amounts of data in parallel. ParallelRunStep works by breaking up the data into batches that are processed in parallel.

You should not use PythonScriptStep. PythonScriptStep is a basic, built-in step that is used to run a Python Script on a compute target. It takes a script name and other optional parameters, such as arguments for the script, compute target, inputs, and outputs. PythonScriptStep can be used to execute the scoring script; however, it will not run parallel instances partitioning the data.

You should not use AdlaStep. AdlaStep works only with data stored in the default Data Lake Storage of the the Azure Data Lake Analytics account. In this case, we have multiple storage accounts where you might want to run the pipeline.

You should not use ParallelRunConfig. ParallelRunConfig is the configuration that is provided to the ParallelRunStep. ParallelRunConfig cannot be used as a pipeline run step.

**References**

Run batch inference on large amounts of data by using Azure Machine Learning

AdlaStep class

PythonScriptStep class

ParallelRunStep class

Question 34

Your model training script includes the following code:

```
from azureml.core import Experiment
exp = Experiment(workspace=ws, name='my_experiment')
run = exp.start_logging()
run.log()
```

You need to log a scalar metric for your model on each run. This metric should be stored in the run record for the experiment.

What should you do?

Choose the correct answer

- Tag the run with a string key and value.
- Add code to define a dictionary object and include the log_table method. Specify a metric name.
- Add the desired parameters to the run.log function.
- Replace the run.log() method with run.log_list("my_metric", [0.52, 0.1, 0.4]).

**Explanation**

You should add the desired parameters to the run.log function. Azure Machine Learning allows you to track multiple metrics for your experiments. These metrics are stored in the experiment's run record for later retrieval and analysis, and the same metric can be logged within a run more than once. The run.log method can be used to log string or numerical scalar values and accepts three parameters, the metric name, the value to be logged, and an optional description.

You should not replace the run.log() method with run.log_list("my_metric", [0.52, 0.1, 0.4]). This method allows you to log lists of values in an experiment run. A value list is comparable to a one-dimensional array.

You should not add code to define a dictionary object and include the log_table method and then specify a metric name. You can use the run.log_table method to log a dictionary object to the run. A dictionary is sometimes referred to as an array and allows you to store key values along with associated data.

You should not tag the run with a string key and value. The run.tag method allows you to tag a run with a string key. The value is optional.

**References**

Enable logging in Azure ML training runs

Run class

Question 35

You evaluate Automated Machine Learning (ML) training results based on the metrics generated during the training run.

The classification model you evaluate metrics for is used to predict the possibilities of a person getting a viral infection based on the location, gender, and age of an individual. The dataset available is highly imbalanced.

You need to find the optimal metric to evaluate the efficiency of the model.

Which metric should you use?

Choose the correct answer

- normalized_root_mean_squared_error
- AUC_Weighted
- log_loss
- accuracy

**Explanation**

You should use the AUC_Weighted metric. AUC_Weighted is a metric that can be used for classification models. AUC_Weighted is the arithmetic mean of the score for each class, weighted by the number of true instances in each class. This works well when the datasets are imbalanced.

You should not use the accuracy metric. This metric provides the efficiency of the model against the training data. However, you should not use this metric when the dataset is imbalanced, because it may provide inconsistent weight to certain rows of the dataset.

You should not use the normalized_root_mean_squared_error metric. This metric is well suited for the evaluation of regression algorithms. This metric is not optimal for classification model evaluation.

You should not use the log_loss metric. This metric is used for logistic regressions and similar extensions, such as neural networks. This metric is not optimal for classification model evaluation.

**References**

Understand automated machine learning results

Root-mean-square deviation

Question 36

You use the Apply SQL Transformation module to ingest data in a machine learning pipeline. You process data from two datasets: a list of employees with salary data and a list of departments with department names. Your transformed data should list the department name and the average salary for the department.

You need to generate a SQL query to meet this goal.

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

Create a list in the correct order

**Possible clauses**

**Clauses in order**

- SELECT t2.department, AVG(t1.salary)
- FROM t1
- JOIN t2
- ON t1.departmentID = t2.departmentID
- GROUP BY t1.departmentID

**Explanation**

You should complete the SQL query in the following order:

1. SELECT t2.department, AVG(t1.salary)
2. FROM t1
3. JOIN t2
4. ON t1.departmentID = t2.departmentID
5. GROUP BY t1.departmentID

The basic structure of a SQL query includes SELECT and FROM clauses. The SELECT clause allows you to specify the data you want, and a simple query usually consists of one or more table columns. The FROM clause allows you to specify the database tables that hold the data you need. When using the Apply SQL Transformation module, datasets are represented as t1, t2, or t3.

As you are combining data from two datasets, you include the name of the first dataset in the FROM clause, and then use the JOIN clause to specify the second dataset. The ON clause identifies the column that the two datasets have in common, in this case, the departmentID.

Finally, the GROUP BY clause tells the module how to group data prior to applying the group function, in this case, the average function.

You should not use the WHERE t1.departmentID = t2.departmentID clause. In some SQL implementations, you can join tables in the FROM clause and specify the shared column in the WHERE clause.

**References**

[Apply SQL Transformation]

[SQL Introduction]

Question 37

You use Azure Machine Learning to train models to play video games.

To increase training performance, you need to ensure your training jobs can use multiple compute targets.

What should you do?

Choose the correct answer

- Create a generic Estimator class object. Populate the compute_target parameter.
- Use the SKLearn class to create an Estimator. Use a script_params dictionary to pass parameters to the estimator.
- Specify a virtual network. Ensure head and worker nodes can communicate with each other.
- Create a FileDataset. Download the dataset files to each compute target.

**Explanation**

You should specify a virtual network and ensure head and worker nodes can communicate with each other. Training video game machine learning models uses a process known as reinforcement learning (RL). RL machine learning agents take actions and then observe the results as a way of seeking rewards. Because RL is compute intense, it is typically performed across multiple compute nodes, known as head and worker nodes. In Azure Machine Learning, RL requires that you specify a virtual network that does not block the ports that nodes need to communicate.

You should not create a FileDataset and download the dataset files to each compute target. You create a file dataset to reference the unstructured file or files you want to use in your machine learning experiments. In machine learning experiments, a FileDataset may be downloaded to compute targets or stored elsewhere and mounted to the experiment.

You should not create a generic Estimator class object and populate the compute_target parameter. The Estimator class can be used when a predefined machine learning framework estimator does not already exist in Azure Machine Learning. For reinforcement learning, Azure Machine Learning provides the ReinforcementLearningEstimator class.

You should not use the SKLearn class to create an Estimator and then use a script_params dictionary to pass parameters to the estimator. The SKLearn class is used when running Scikit-learn training scripts. Scikit-learn is a Python-based machine learning library that only supports single-node compute targets.

**References**

[Reinforcement learning (preview) with Azure Machine Learning]()

[Train with datasets in Azure Machine Learning]()

[Train models with Azure Machine Learning using estimator]()

[Estimator class]()

[Build scikit-learn models at scale with Azure Machine Learning]()

Question 38

Your team creates an automobile prediction pipeline as shown in in the exhibit.

You need to submit the run and understand the model.

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

| Statement | Yes | No |
|---|---|---|
| The Split Data module splits the columns in your dataset, based on how you configure the module. | ○ | ◉ |
| The model classifies automobiles into various price categories. | ○ | ◉ |
| The Score Model generates scores for prices predicted versus prices provided for the test data features. | ◉ | ○ |

**Explanation**

The Split Data module does not split the data columns. Instead, it provides a configuration where you can split the rows of data into a training set and a test data set.

The model has a Linear Regression module that runs a regression algorithm. The model does not run a classification algorithm.

The Score Model takes the model as the input and uses the test data from the Split Data module. The test data already has the label values for the features being used to predict the prices. The Score Model runs the model and comes up with a prediction. It then compares the prices predicted with the prices provided from the test data.

**References**

What is Azure Machine Learning designer?

Tutorial: Predict automobile price with the designer

Question 39

Your clients want to be able to make calls to a batch inference pipeline to execute a large volume of data that is periodically uploaded to a storage account. You want to create a pipeline that leverages the ParallelRunStep to run an inferencing script to provide outcomes to your client.

You need to write a batch inference script.

Which two functions should you include in your script? Each correct answer presents part of the solution.

Choose the correct answers

- [ ] load()
- [ ] evaluate(mini_batch)
- [ ] init()
- [ ] execute(mini_batch)
- [ ] run(mini_batch)

**Explanation**

You should include the init() and run(mini_batch) functions. You should use init() for any costly or common preparation for later inference. You should use run(mini_batch) to write code that would evaluate and append the outputs of the evaluation.

You should not include the execute(mini_batch), load(), or evaluate(mini_batch) functions. They are not valid functions for an entry script.

**References**

[Run batch inference on large amounts of data by using Azure Machine Learning]

Question 40

You are configuring Automated Machine Learning (ML) to evaluate a classification experiment based on the following features:

- You want to limit the number of algorithm/parameter combinations to test to 6, because you want to run the experiment in your local environment.
- You have a trained dataset named train_ds and a test dataset named test_ds.
- The features are location, gender, and age.
- The outcome is the infections column in the datasets.
- You want Automated ML to identify features and apply normalization to them.
- The training dataset is highly unbalanced, so you want to ensure to provide the appropriate metric in the Auto ML evaluation.

You need to evaluate the classification experiment.

Which code should you use? To answer, complete the commands by selecting the correct parts from the drop-down menus.

Choose the correct options

```
from azureml.train.automl import AutoMLConfig

automl_config = AutoMLConfig(
                    name='Automated ML Experiment',
                    task='                        ',
                    compute_target='                        ',
                    enable_local_managed=True,
                    training_data = train_ds,
                    validation_data = test_ds,
                    label_column_name='                        ',
                    iterations=6,
                    primary_metric='                        ',
                    max_concurrent_iterations=4,
                    featurization='                        ')
```

classification

local

infections

AUC_w eighted

auto

## Explanation

You should complete the code as follows:

```
from azureml.train.automl import AutoMLConfig

automl_config = AutoMLConfig(name='Automated ML
Experiment',
                             task='classification',
                             compute_target='local',
                             enable_local_managed=True,
```

```
                                 training_data = train_ds,
                                 validation_data = test_ds,
                                 label_column_name='infection
s',
                                 iterations=6,
                                 primary_metric =
'AUC_weighted',
                                 max_concurrent_iterations=4,
                                 featurization='auto')
```

You should select classification as the task, since the Automated ML experiment is evaluating a classification problem.

You should select local as the compute_target. You want to run the experiment in your local environment rather than on a remote compute cluster.

You should select infections as the label_column_name. You want the outcome of the experiment to predict infection possibility. Location, gender, and age are features used to train the model.

You should select AUC_weighted as the primary_metric. AUC_weighted is the correct metric to evaluate unbalanced datasets. It measures the area under the curve with consideration to the mean for each class, weighted by the number of true instance evaluated in each class.

You should set featurization to auto. You want Automated ML to identify the features. Normalization and scaling are feature engineering transformations that the Automated ML experiment will conduct.

**References**


AutoMLConfig class

What is automated machine learning (AutoML)?

## Question 41

You use Azure Machine Learning designer to publish an inference pipeline as a web service. During deployment, you create a service principal (SP) and configure authentication.

You need to use the SP while consuming the endpoint.

What should you do?

Choose the correct answer

- Use the client secret to retrieve an authentication token.
- Use the AciWebservice.deploy_configuration to set auth_enabled to True.
- Use the get_keys method to retrieve authentication keys.
- Use the regen_key method to regenerate the primary key.

**Explanation**

You should use the client secret to retrieve an authentication token. A SP is any directory object that can be used for authentication. Once an SP is created, it can be used in Azure Machine Learning to facilitate token-based authentication. As part of the creation process for an SP, a client secret is generated, which is comparable to a password. You can supply the client secret in an HTTP POST to retrieve an authentication token. This token, also known as a JSON Web Token (JWT), can then be used to authenticate with an Azure Machine Learning web service. By default, a JWT is valid for one hour, and must be refreshed when it expires.

You should not use the regen_key method to regenerate the primary key. In addition to token-based authentication, Azure Machine Learning also supports key-based web-service authentication. Keys, comparable to Application Programming Interface (API) keys, are statically generated. If a key needs to be replaced, you can use the regen_key method to regenerate authentication keys.

You should not use the get_keys method to retrieve authentication keys. This method is used to retrieve the primary and secondary authentication keys. You used keys to authenticate to an Azure Machine Learning web service that has been configured with key-based authentication.

You should not use the AciWebservice.deploy_configuration to set auth_enabled to True. This method is used to enable key-based authentication on an Azure Container Instances (ACI) web service.

**References**

[Consume an Azure Machine Learning model deployed as a web service]

[Create, run, and delete Azure ML resources using REST]

[Set up authentication for Azure Machine Learning resources and workflows]

Question 42

You need to create an Azure Machine Learning workspace using Azure Cloud Shell.

How should you complete the command? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

| | | |
|---|---|---|
| az ml workspace create | AML-Workspace | AML-LearningResources |

-w

-g

**Explanation**

You should add the -w AML-Workspace parameter. This parameter specifies the name of the workspace.

You should add the -g AML-ResourceGroup parameter. This parameter identifies the resource group that will home the Azure Machine Learning workspace. An Azure resource group is a collection of Azure components. A resource group allows you to combine Azure resources based on function, geography, or any other attribute you choose. The resource group must exist before you run the az ml workspace create command.

The following command creates an Azure Machine Learning workspace in the AML-ResourceGroup:

az ml workspace create -w AML-Workspace -g AML-ResourceGroup

You should not add the --sku parameter. This parameter allows you to specify workspace edition: Basic or Enterprise. If omitted, the default option of this parameter is Basic.

You should not add the -l parameter. This parameter allows you to specify the Azure region where the Azure Machine Learning will be created. Each region offers different types of resources used for machine learning; therefore it is highly recommended that you choose a region with the resources you will need for your machine learning operations.

You should not add the --keyvault parameter. This parameter allows you to specify the Key Vault that will be associated with the workspace. Azure Key Vaults are used to manage credentials and authentication keys that will be used by the workspace.

**References**

Create a workspace for Azure Machine Learning with Azure CLI

az ml workspace

Azure Machine Learning workspaces

Question 43

You create Azure Machine Learning workspaces.

You need to perform machine learning assisted image classification for 1,000,000 images. Each image is either of a cat or of a dog.

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

Create a list in the correct order

**Possible actions**

**Actions in order**

- Create an Azure Machine Learning workspace. Set the edition to Enterprise.
- Define two datasets and split the images between each dataset.
- Create a multi-class image classification project.

**Explanation**

You should complete the following actions in order:

1. Create an Azure Machine Learning workspace. Set the edition to Enterprise.
2. Define two datasets and split the images between each dataset.
3. Create a multi-class image classification project.

First, you should create an Azure Machine Learning workspace and set the edition to Enterprise. Azure Machine Learning can be used to create, manage, and monitor image classification projects. This feature is available in the Enterprise edition of Azure Machine Learning. You define the machine learning workspace edition when the workspace is created. You can also upgrade any existing Basic edition to Enterprise edition at any time.

Then, you should define two datasets and split the images between each dataset. Training a machine learning model to correctly classify images requires large datasets of correctly labeled images. Azure Machine Learning supports up to 500,000 images per dataset when performing machine learning assisted image classification. Any images beyond this limit will not be processed.

Finally, you should create a multi-class image classification project. Azure Machine Learning supports several types of image classification projects. Multi-class projects are used when a single class is applied to an image. In this scenario, either the dog or the cat class will be assigned to each image.

You should not use Azure Machine Learning SDK to create an Azure Machine Learning workspace and set the SKU to Basic. The Basic edition of Azure Machine Learning does not support machine learning assisted image classification. However, a Basic edition workspace can be upgraded to the Enterprise edition at any time.

You should not create a multi-label image classification project. Multi-label projects are used when multiple labels might be applied to a single image. For example, if an image included a dog and a cat, it may receive a label for each animal.

### References

Create a data labeling project (preview) and export labels

Enterprise and Basic Editions of Azure Machine Learning

Question 44

You use Azure Machine Learning as part of a project that will evaluate a black box machine learning model.

You need to ensure that the mimic explainer interpretability technique can be used to evaluate the black box model.

What should you do first?

Choose the correct answer

- Create a multi-label image classification project.
- Implement Shapley Additive Explanations (SHAP) linear explainer.
- Train a global surrogate model.
- Configure Permutation Feature Importance Explainer (PFI).

**Explanation**

You should train a global surrogate model. A global surrogate is meant to be an interpretable approximation of a black box model. Black box models are those for which no explanation exists, which means that the public does not know how the model makes its predictions. Once a surrogate model is trained, the mimic explainer interpretability technique can be used to interpret the model.

You should not configure PFI. Features are data fields that are used to train a model. If you need to determine which fields - or features - have the largest impact on a model's predictions, you should use an interpretability technique that calculates and tracks feature importance. Azur Machine Learning supports the PFI for this purpose. PFI randomly shuffles features during model training and then calculates the impact on the model's performance.

You should not implement SHAP linear explainer. SHAP is a model-specific interpretability technique used for linear models. SHAP explainers use calculates based on coalitional game theory.

You should not create a multi-label image classification project. Multi-label projects are used when multiple labels might be applied to a single image. For example, if an image included a dog and a cat, it may receive a label for each, dog and cat.

**References**


Model interpretability in Azure Machine Learning

5.6 Global Surrogate

Create a data labeling project (preview) and export labels

Question 45

You have a sample file with historical data that has columns for features and labels. The data spans multiple years of sales and has over 1 million rows.

To examine the data, you create a training dataset in Azure Machine Learning studio and drop the dataset onto a new designer pipeline window.

In the results visualization for the dataset, you notice that the revenue column has empty values for various rows of the dataset.

You need to eliminate all rows from the dataset where the revenue column does not contain values.

Which Azure Machine Learning module should you use?

Choose the correct answer

- Select Column in Dataset
- Partition and Sample
- Normalize Data
- Clean Missing Data

**Explanation**

You should use the Clean Missing Data module. You should configure the module to remove the entire row when the revenue column is empty. This keeps the rows in the original dataset, but the removed rows are not to be used in the model training.

You should not use the Select Column in Dataset module. Select Column in Dataset is used to specify which columns are included in the next activity of the training pipeline. It will not allow you to remove rows that do not contain values.

You should not use the Normalize Data module. Normalize Data allows you to configure all columns that contain numerical data to have the same scale without losing information. This will allow the training step to eliminate bias that could be associated with higher values because of certain column value units.

You should not use the Partition and Sample module. This module performs sampling on a dataset and analyzes the data without losing meaning. This module does not allow you to specify rows to eliminate from a dataset based on a column value being empty.

**References**

Clean Missing Data module

Select Columns in Dataset module

Normalize Data module

Partition and Sample module

Question 46

An ecommerce company wants you to create a machine learning model that runs a classification algorithm on images they receive from various product vendors. The model will be used to categorize these images into product labels based on what the image shows. You decide to use Azure Machine Learning studio to create a labeling project for an initial set of images provided.
You create an Azure Machine Learning workspace in an Azure subscription. You are now ready to create the labeling project.

You need to meet the following requirements:

- An image can only be given one label during the classification process.
- The labels that are associated with products do not change.
- To reduce errors, you want all labeling to be done manually.
- Additional images will need to be added to the datastore on a weekly basis.

You need to configure your image classification labeling project.

What six actions should to perform in sequence? To answer, move the appropriate actions from the list of possible actions to the answer area and arrange them in the correct order.

Create a list in the correct order

**Possible actions**

**Actions in order**

- Choose Image Classification Multi-class as the labeling task type.
- Select or create a dataset associated with storage where the images reside.
- Ensure that the Incremental refresh option is turned on.
- Create the list of label classes against which you want to categorize your images.
- Provide any instructions for the labelers on the process of classification.
- Ensure that the Enable ML assisted labeling is turned off.

## Explanation

You should configure your labeling project in the following order:

1. Choose Image Classification Multi-class as the labeling task type.
2. Select or create a dataset associated with storage where the images reside.
3. Ensure that Incremental refresh is turned on.
4. Create the list of label classes against which you want to categorize your images.
5. Provide any instructions for the labelers on the process of classification.
6. Ensure that the Enable ML assisted labeling is turned off.

Fisrt, you should select Image Classification Multi-class as the task type. You want all images to have one label and to be classified based on the label class.

Next you need to select or create your dataset of images.

Then, since the images will change in the datastore, you should ensure that incremental refresh is turned on.

Next, you should provide the list of labels against which you want to classify your images.

Finally, you need to provide instructions and ensure that the Enable ML assisted labeling is turned off since the requirement states that all labeling should be done manually.

You should not ensure that the Incremental refresh option is turned off. This is required to be turned on because you want to update the classification dataset as new images are added.

You should not ensure that Enable ML assisted labeling is turned on. You want all classification to be done manually.

You should not choose Image Classification Multi-label as the labeling task type. The requirements state that an image can only have one label class.

You should not choose Object Identification (Bounding Box) as the labeling task type. The requirements are to classify the image. This type of task type allows objects to be classified in an image.

**References**

Create a data labeling project and export labels

Question 47

You work for a healthcare organization and are responsible for configuring datastores for various large data files. You plan to create the datastore using the files stored in Azure Data Lake Gen2 storage.

You need to register the datastore.

Which Python method should you use?

Choose the correct answer

- Datastore.register_azure_data_lake_gen2()
- Datastore.register_azure_blob_container()
- Datastore.register_azure_data_lake()
- Datastore.register_azure_file_share()

**Explanation**

You should use the Datastore.register_azure_data_lake_gen2() method. This method will register a credential datastore connected to Azure DataLake Gen 2 storage with service principal permissions.

You should not use the Datastore.register_azure_blob_container() method. This method is used to register a datastore for blob containers and not for Azure DataLake Gen2 storage.

You should not use the Datastore.register_azure_file_share() method. This method is used to register a datastore for an Azure Storage account file share and not for Azure DataLake Gen2 storage.

You should not use the Datastore.register_azure_data_lake() method. This method is used to register a datastore for Azure DataLake Storage Gen1.

**References**

[Connect to Azure storage services]

Question 48

You plan to configure logging for an experiment that explores data associated with gender distribution within organizations across the world. The experiment must draw a box plot of genders by country.

You need to use the appropriate logging method to render the plot for the experiment run object.

Which run object method should you use?

Choose the correct answer

- log_table()
- log_image()
- log_row()
- log_list()
- log()

**Explanation**

You should use the log_image() method. This method logs a .PNG image file or a matplotlib plot to the run. These images will be visible and comparable in the run record.

You should not use the log() method. This method is used to log a numerical or a string value to the run with a given name. This method will not render a plot.

You should not use the log_list() method. This method is used to log a list of values to the run with a given name. This method will not render a plot.

You should not use the log_row() method. This method creates a metric with multiple columns. Each named parameter generates a column with the value specified. This method will not render a plot.

You should not use the log_table() method. This method can be used to log a dictionary object to the run with the given name. This method will not render a plot.

**References**

[Enable logging in Azure ML training runs]

[Run class]

Question 49

You use Azure Machine Learning to design and train machine learning models. You obtain new data and would like to use the data to retrain an existing machine learning model.

You need to bookmark the state of your data prior to retraining.

What should you do?

Choose the correct answer

- Use the get_by_name method from the Dataset class. Increment the version parameter.
- Create and register a new file dataset. Use the from_files method to specify the bookmarked data.
- Create and register a tabular dataset. Specify the new version name.
- Use the register method from the Dataset class to create a new dataset version.

**Explanation**

You should use the register method from the Dataset class to create a new dataset version. Azure Machine Learning allows you to register a new dataset by using an existing dataset name using versioning. A version is a bookmark of the data's state and is useful in cases where new data needs to be used for retraining. By creating versions, you can return to a specific version of the dataset if necessary. To register a dataset version, you can set the create_new_version parameter of the register method to True.

You should not use the get_by_name method from the Dataset class and increment the version parameter. You can use the get_by_name method to retrieve a specific version of a dataset. If a version is not specified, the latest dataset version registered with the workspace is returned.

You should not create and register a tabular dataset and specify the new version name. You create a tabular dataset to reference tabular data - such as that found in a comma-separated-value (CSV) file. You use the Dataset.Tabular.from_delimited_files method to register tabular data in a dataset.

You should not create and register a new file dataset and use the from_files method to specify the bookmarked data. You create a file dataset to reference the file or files you want to use in your machine learning experiments. The from_files method is used to identify the path and file specification that will be used when the dataset is created.

**References**

Version and track datasets in experiments

Dataset class

Introduction to datasets

Question 50

You use Azure Machine Learning to create a machine learning pipeline.

You need to ensure that files can be passed between pipeline steps using a named datastore.

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

Create a list in the correct order

**Possible actions**

**Actions in order**

- Register a new Azure Storage file container datastore.
- Create a PipelineData object. Specify a name and output datastore.
- Specify a PipelineData object for data output.

**Explanation**

You should perform the following steps in order:

1. Register a new Azure Storage file container datastore.
2. Create a PipelineData object. Specify a name and output datastore.
3. Specify a PipelineData object for data output.

A PipelineData object can be used to pass data between steps in a pipeline. When you create a PipelineData object, you must specify a name. If you do not provide a data store reference, the default workspace datastore will be used. In this scenario, you are required to use a named datastore, so the first action requires registering a new file container datastore.

Once a PipelineData object has been created, you can specify the output of a pipeline step to be written to this object in the same way as you would to a local folder. In successive pipeline steps, you can then read data by referencing the PipelineData object created in an earlier step.

You should not retrieve the default datastore from the current workspace. You can use the get() method to retrieve a datastore by name. The following code retrieves a datastore named default-datastore:

```
my_datastore = Datastore.get(default-datastore)
```

You should not register a new dataset version for each pipeline pass. Azure Machine Learning allows you to register a new dataset using an existing dataset name using versioning. A version is a bookmark of the data's state and is useful in cases where new data needs to be used for retraining. By creating versions, you can return to a specific version of the dataset if necessary.

**References**

Pass data between pipeline steps

Introduction to datastores

Version and track datasets in experiments

Question 51

You are developing a financial markets prediction model. You create a tabular timeseries dataset named time_series_ds with a partition timestamp by date.

You are leveraging Azure ML SDK to query the dataset. You want to view 100 records from the dataset for January 2020.

You need to write the code to view the records in a Pandas dataframe.

Which code should you execute?

Choose the correct answer

- [x]
```
jan_ds = time_series_ds.time_between(
                                      start_time=datetime(2019, 12, 31),
                                      end_time=datetime(2020, 2, 1),
                                      include_boundary=False)
jan_ds.take(100).to_pandas_dataframe()
```

- [ ]
```
jan_ds = time_series_ds.time_between(
                                     start_time=datetime(2020, 1, 1),
                                     end_time=datetime(2020, 1, 31))
jan_ds.to_pandas_dataframe().take(100)
```

- [ ]
```
jan_ds = time_series_ds.time_between(
                                     start_time=datetime(2020, 1, 1),
                                     end_time=datetime(2020, 1, 31),
                                     include_boundary=False)
jan_ds.to_pandas_dataframe().take(100)
```

- [ ]
```
jan_ds = time_series_ds.time_between(
                                      start_time=datetime(2019, 12, 31),
                                      end_time=datetime(2020, 2, 1))
jan_ds.take(100).to_pandas_dataframe()
```

**Explanation**

You should execute:

```
jan_ds = time_series_ds.time_between(
start_time=datetime(2019, 12, 31),
end_time=datetime(2020, 2, 1),
include_boundary=False)
jan_ds.take(100).to_pandas_dataframe()
```

The time_between method retrieves records based on the start_time and end_time provided. Specifying include_boundary to False excludes the boundary dates. In this case, it would exclude 12/31/2019 and 2/1/2020. Also, providing take(100) will return 100 rows and show them in the dataframe.

You should not execute:

```
jan_ds = time_series_ds.time_between(
start_time=datetime(2020, 1, 1),
end_time=datetime(2020, 1, 31),
include_boundary=False)
jan_ds.to_pandas_dataframe().take(100)
```

The time_between method retrieves records based on the start_time and end_time. However, specifying include_boundary to False, will exclude 1/1/2020 and 1/31/2020 from the results returned. Also, take() is a tabular dataset method and not a method applied to the pandas_dataframe.

You should not execute:

```
jan_ds = time_series_ds.time_between(
start_time=datetime(2019, 12, 31),
end_time=datetime(2020, 2, 1))
jan_ds.take(100).to_pandas_dataframe()
```

The time between method retrieves records based on the start_time and end_time. However, not specifying include_boundary sets it to True. This will include 12/31/2019 and 2/1/2020 in the results returned. The requirement is only to get the records for January, 2020.

You should not execute:

```
jan_ds = time_series_ds.time_between(
start_time=datetime(2020, 1, 1),
end_time=datetime(2020, 1, 31))
jan_ds.to_pandas_dataframe().take(100)
```

The time_between method retrieves records based on the start_time and end_time, and by default, it includes the boundary dates specified, which correctly returns the records for January 2020. However take() is a tabular dataset method and not a method applied to the pandas_dataframe.

### References

[Tabular Time Series Related API Demo with NOAA Weather Data]

[TabularDataset class]

Question 52

You use Azure Machine Learning to generate models that will be used to identify faces in images.

You need to ensure that model training is optimized for this task.

Which two actions should you perform? Each correct answer presents part of the solution.

Choose the correct answers

- ☐ Set the primary metric goal to MAXIMIZE.
- ☐ Set the primary metric goal to MINIMIZE.
- ☐ Configure Spearman correlation as the primary metric.
- ☐ Configure accuracy as the primary metric.

**Explanation**

You should configure accuracy as the primary metric and set the primary metric goal to MAXIMIZE. Automated machine learning in Azure Machine Learning uses the primary metric you define to optimize model training. The metrics you can configure are dependent on the machine learning task type, such as regression or classification.

In this question, you will use image classification to identify faces in images. The accuracy metric can be used for classification tasks, and it calculates the proportion of instances that have been correctly classified. As you want your model to be as accurate as possible, you should set the primary metric goal to MAXIMIZE, meaning Azure Machine Learning will attempt to maximize the model's classification accuracy.

You should not configure Spearman correlation as the primary metric. Spearman correlation calculates the monotonic relationship between two values. For example, two stock tickers may have a monotonic relationship where stock A's price decreases when stock B's price increases.

You should not set the primary metric goal to MINIMIZE. This metric goal is useful when you are tracking experiment errors and you want to minimize the number of errors a model reports.

**References**

[Tune hyperparameters for your model with Azure Machine Learning]

[Understand automated machine learning results]

Question 53

You want to reduce the cost for the training compute clusters created in your Azure Machine Learning workspace. The model training will be executed intermittently on text files, and execution time will not be a factor. You will use a Standard_D1 virtual machine size.

You need to ensure minimal cost for the compute environment during idle time.

How should you configure your compute cluster? To answer, drag the appropriate configuration to each setting. A configuration item may be used once, more than once, or not at all.

Drag and drop the answers

**CPU (Central Processing Unit)**

**Low priority**

**0**

**Explanation**

You should select CPU (Central Processing Unit) for your virtual machine type. The virtual machine (VM) used for the compute environment will process text files. CPUs are used to parse text files. In cases where you have to process media files (images and videos), you should leverage a Graphic Processing Unit (GPU) which would have higher cost for the instance. CPU allows you to process text files and minimize the cost.

You should select Low priority for the Virtual machine priority. Selecting Low priority will let you use unutilized capacity of Azure managed VMs. These allocations are preemptible, but come at a reduced price compared to Dedicated VMs.

You should set the minimum number of nodes to 0. Even though there will be startup time associated with the execution of the training session for models, when there are no training sessions being conducted, Azure will ensure that no capacity is allocated.

**References**

Plan and manage costs for Azure Machine Learning

GPUs vs CPUs for deployment of deep learning models

Question 54

You use Azure Machine Learning to train models on data collected from Internet of Things (IoT) devices.

You need to monitor and analyze drift in your data as new information is collected from your IoT devices.

What should you do?

Choose the correct answer

- Define a dataset monitor and configure a target dataset with a timeseries trait.
- Stream Azure Machine Learning metric information to Azure Event Hub.
- Add logging functions to your pipeline with the Execute Python Script module.
- Use ScriptRunConfig to add logging functions to your training scripts.

**Explanation**

You should define a dataset monitor and configure a target dataset with a timeseries trait. Data drift is the phenomenon where the data used to train a model diverges from later model input data. This can occur for a variety of reasons, and the concern is that, if left unchecked, data drift can lead to model performance degradation over time.

You can define a dataset monitor if you want to monitor for statistical changes and data drift in your datasets. Each dataset monitor requires a baseline dataset, which is typically the dataset that was used to initially train the model. You must also specify a target dataset, which is where new data is stored and compared with the baseline dataset. This target dataset must have the timeseries trait set, which is typically done by adding a timestamp column. Once the dataset monitor is created and configured, you can view drift analysis information in the Azure Machine Learning portal.

You should not stream Azure Machine Learning metric information to Azure Event Hub. Azure Event Hub is a stand-alone platform that, like Azure Monitor, can ingest logging and other information from a variety of Azure Services. However, Event Hub is focused on data analysis to discover actionable insights, sometimes referred to as business intelligence.

You should not use ScriptRunConfig to add logging functions to your training scripts. The ScriptRunConfig class is used to create an object that contains both training environment configuration information, as well as a training script. This ScriptRunConfig object can be used to initiate a fully configured training run as part of a machine learning experiment.

You should not add logging functions to your pipeline with the Execute Python Script module. This module can be added to a drag-and-drop designer pipeline to run Python code. This is useful in cases where an existing Azure Machine Learning designer module does not provide the functionality you need for your experiments.

**References**

Detect data drift (preview) on datasets

ScriptRunConfig class

Execute Python Script module

Question 55

You plan to create an experiment using Azure Machine Learning workspace to explore the effects of global warming on temperature variations over the years. You define a variable named ws that points to your workspace and a variable named experiment_name that refers to the name of the experiment.

You need to enable logging the application state during the training process.

Which code segment should you use?

Choose the correct answer

- from azureml.core.compute import ComputeTarget

```
compute_target = ComputeTarget.attach(
   workspace=ws, name="example", attach_configuration=config)
compute.wait_for_completion(show_output=True)
```

- from azureml.core.webservice import Webservice

```
service = Webservice(name="service-name", workspace=ws)
logs = service.get_logs()
```

- from azureml.core import Experiment

```
exp = Experiment(workspace=ws, name='global_warming_experiment')
run = exp.start_logging()
run.log("temprature", 78)
```

- from azureml.core import Experiment

```
experiment = Experiment(ws, experiment_name)
run = experiment.submit(config=run_config_object, show_output=True)
```

**Explanation**

You should use:

```
from azureml.core import Experiment


experiment = Experiment(ws, experiment_name)
run = experiment.submit(config=run_config_object,
show_output=True)
```

The Experiment.submit method can be configured with the show_output parameter to enable local logging during the training process.

You should not use:

```
from azureml.core import Experiment

exp = Experiment(workspace=ws,
name='global_warming_experiment')
run = exp.start_logging()
run.log("temprature", 78)
```

The Experiment.start_logging() enables logging for run-related data within the experiment. This will not show the logs generated during the training process.

You should not use:

```
from azureml.core.compute import ComputeTarget

compute_target = ComputeTarget.attach(
    workspace=ws, name="example",
attach_configuration=config)
compute.wait_for_completion(show_output=True)
```

The ComputeTarget.wait_for_completion method configures logging during a compute target creation. This will not show the logs generated during the training process.

You should not use:

```
from azureml.core.webservice import Webservice

service = Webservice(name="service-name", workspace=ws)
logs = service.get_logs()
```

The services.get_logs() enables logs to be retrieved for a previously deployed web service. The logs may contain detailed information about a past run, but they do not show the logs generated during the training process.

**References**

[Enable logging in Azure ML training runs]

[Start, monitor, and cancel training runs in Python]

Question 56

You create a batch inference pipeline that leverages ParallelRunStep for multi-node processing.

You want to run the pipeline, and you write the following code:

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

pipeline = Pipeline(workspace=ws,
steps=[parallelrun_step])
experiment = Experiment(ws, 'digit_identification')
pipeline_run = experiment.submit(pipeline)
```

You need to add code to stream run logs and monitor the run status.

What line of code should you use to view the logs?

Choose the correct answer

- pipeline_run.get_status()
- pipeline_run.wait_for_completion()
- pipeline_run.get_graph()
- pipeline_run.get_steps()

**Explanation**

You should use pipeline_run.wait_for_completion() and add it at the end of your code. The wait_for_completion method is used to monitor the status and by default, it will stream the logs to the output determined by the sys.stdout configuration.

You should not use pipeline_run.get_status(). The get_status method retrieves the status of the run, however, it will not provide streaming logs while the pipeline is being executed.

You should not use pipeline_run.get_graph(). The get_graph renders the graph of the pipeline run. It will not provide streaming logs while the pipeline is being executed.

You should not use pipeline_run.get_steps(). The get_steps method lists the generated step runs. The method provides a list of all the pipeline steps that have completed or have started running.

**References**

Creating a batch inference pipeline

Run batch inference on large amounts of data by using Azure Machine Learning

Question 57

You tune hyperparameters on your Hyperdrive Experiment based on Random sampling. You want to terminate 30 percent of the lowest performing runs at each evaluation interval, based on their performance of the primary metric.

You need to associate an early termination policy to your Hyperdrive Experiment.

Which termination policy should you use?

Choose the correct answer

- Bandit policy
- No termination policy
- Truncation selection policy
- Median stopping policy

**Explanation**

You should use the Truncation selection policy. The Truncation selection policy will cancel a percentage of runs with low performance on the primary metric for a given evaluation interval. You can configure a truncation policy providing the following parameters:

- truncation_percentage: the percentage of lowest performing runs to terminate for a given evaluation interval. Values should be between 1 and 99.
- evaluation_interval: the interval at which the running models are evaluated.
- delay_evaluation: the initial evaluation delay interval.

You should not use the Bandit policy. It does not meet your requirements. This policy uses slack factor or slack amount specified to terminate a run at a given evaluation interval. Similar to the Truncation selection policy, Bandit is a termination policy that uses the primary metric to determine the termination at a given interval. However, it does not base it on a percentage of the lowest performing models.

You should not use the Median stopping policy. It does not meet your requirements. This policy uses averages of primary metrics to determine the termination of the run. You cannot specify a percentage of runs to terminate based on the performance of the primary metric.

You should not use the No termination policy. The No termination policy will not terminate any hyperparameter tuning runs. All runs will complete even if the performance of the hyperparameter is not optimal.

### References

[Tune hyperparameters for your model with Azure Machine Learning]

Question 58

You create a deep learning classification model for processing large volume of image files. Your deployment compute target should support real-time inferencing. You want to ensure you are able to conduct GPU-based inferencing for your model.
You need to configure your deployment target for your inferencing model.

Which deployment compute target should you use?

Choose the correct answer

- Azure Kubernetes Service (AKS)
- Azure Machine Learning compute clusters
- Azure Container Instances (ACI)
- Azure Machine Learning compute instances

**Explanation**

You should select Azure Kubernetes Service (AKS). AKS supports creating GPU-based compute instances for deploying real-time inference models. It is recommended for production deployments.

You should not select Azure Machine Learning compute instances. Azure Machine Learning compute instances deploy real-time services. However, they do not provide GPU-based instance configuration.

You should not select Azure Container Instances (ACI). ACI are well suited for testing/debugging workloads and do not provide GPU-based instance configuration. They only support low-scale CPU-based workloads.

You should not select Azure Machine Learning compute clusters. Although they support GPU-based instances, they are used for batch inference models. They are not suitable for real-time inferencing.
**References**
[What are compute targets in Azure Machine Learning?]

Question 59

You have a set of files in the default data store in Azure Machine Learning studio. You plan to create a tabular dataset named global-infections. You want to train a script using this data.

You write the code to get a reference to the dataset as follows:

```
from azureml.core import Workspace, Dataset

subscription_id = '12345678-xdad-yde5-8888-999999999999'
resource_group = 'mlResourceGroup'
workspace_name = 'my-ml-workspace-01'


ws= Workspace(subscription_id, resource_group,
workspace_name)
default_ds = ws.get_default_datastore()
tab_data_set =
Dataset.Tabular.from_delimited_files(path=(default_ds,
'mydatacontainer/*.csv'))


try:
    tab_data_set = tab_data_set.register(workspace=ws,
                                         name='global-
infections',
                                         description='infe
ctions data',
                                         tags = {'format':
'CSV'},
                                         create_new_versio
n=True)
except Exception as ex:
    print(ex)
```

You plan to submit an estimator as part of your experiment run.

You need to configure an estimator and set the correct property to provide access to the dataset.

Which property should you set?

Choose the correct answer

- inputs = [tab_data_set.as_named_input('tab_data_set')]
- script_params = {'--training_ds': training_ds}
- source_directory = tab_data_set
- environment_definition = {'training_data': training_ds}

**Explanation**

You should set the inputs property. The inputs property can be used to pass the dataset as a string. The inputs property can be set to a list of data references or with a dataset consumption configuration.

You should not use the source_directory property. The source_directory property specifies where the source script is located. It does not provide a reference to the tab_data_set dataset.

You should not use the script_params property. The script_params property consists of the dictionary of command-line arguments to pass to the training script specified in the entry_script property.

You should not use the environment_definition property. The environment definition for an experiment includes PythonSection, DockerSection, and environment variables. Any environment option not directly exposed through other parameters to the Estimator construction can be set using the environment_definition parameter.

**References**

SKLearn class

Question 60

You create, train, and test a linear regression model.

You plan to deploy the model as a web service endpoint that users can use to get real-time outcomes based on the features they select.

You need to create the compute resource. You want to ensure that the cluster that runs the pipeline is managed by Azure.

Which compute option should you use?

Choose the correct answer

- Inference cluster
- Compute instance
- Compute cluster
- Attached compute

**Explanation**

You should create an Inference cluster. In order to host a real-time web service endpoint, you would have to create a cluster consisting of an Azure Kubernetes Service (AKS) instance or Azure Container Instance (ACI) managed by Azure. AKS is recommended for production use, while ACI is used primarily for small models or for development/testing purposes. This is the only type of cluster that is supported for deploying an inference pipeline as a hosted real-time service.

Compute instance, compute cluster, and attached compute are not supported to run inference pipelines as a real-time service. These clusters are more suited to host training and batch pipelines.

**References**


Tutorial: Deploy a machine learning model with the designer (preview)

Set up and use compute targets for model training