Question 61

You use Azure Machine Learning to deploy a machine learning model to an Azure Kubernetes Service (AKS) cluster. During testing, you receive a large number of HTTP 503 errors.

You need to reduce the incidence of HTTP 503 errors.

Which three actions should you perform? Each correct answer presents a complete solution.

Choose the correct answers

- ☐ Change the utilization level at which containers autoscale up.
- ☐ Increase the request timeout threshold.
- ☐ Change the minimum number of replicas.
- ☐ Modify the autoscale_max_replicas parameter.
- ☐ Migrate the service to an Azure Machine Learning compute cluster.

Explanation

You can change the utilization level at which containers autoscale up. AKS clusters are designed for heavy, real-time production workloads. One of the primary benefits of deploying to AKS is its support for autoscaling. This means that as workload increases or decreases, an AKS cluster can add or terminate cluster nodes.

An HTTP 503 Service Unavailable error indicates that the service is operational but it is unable to respond to requests. This often indicates that the server is overloaded and does not have the resources to process the request.

By default, AKS scales up when cluster utilization exceeds 70 percent. If there is a sudden increase in requests, the cluster may not be able to add nodes quickly enough to handle the requests. By reducing this threshold, you allow the cluster to scale up under lighter loads.

You can also modify the autoscale_max_replicas parameter. By default, an AKS cluster can scale up to 10 containers (nodes). By increasing this parameter, you can ensure that your cluster can handle a higher number of simultaneous requests.

You can also change the minimum number of replicas. This parameter defines the minimum number of nodes that should be online in an AKS cluster. The default value is 1. By increasing this parameter, you can ensure that the cluster always has enough resources to deal with spikes in requests.

You should not increase the request timeout threshold. An HTTP 504 error indicates that a request has timed out. You can increase the request timeout threshold from one minute if requests are timing out too quickly. However, this will not reduce the incidence of HTTP 503 errors.

You should not migrate the service to an Azure Machine Learning compute cluster. Azure Machine Learning compute clusters are scalable machine learning platforms that consists 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.

References

Troubleshoot Docker deployment of models with Azure Kubernetes Service and Azure Container Instances

503 Service Unavailable

Deploy models with Azure Machine Learning

Question 62

You need to determine the appropriate compute specifications for a training workload.

Your solution must support automated machine learning (AutoML), machine learning pipelines, and Azure Machine Learning designer.

What should you do?

Choose the correct answer

- Deploy an Azure Machine Learning compute instance.
- Create and deploy a remote virtual machine.
- Deploy an Azure Machine Learning compute cluster.
- Install the Azure Machine Learning SDK on your local computer.

Explanation

You should deploy an Azure Machine Learning compute cluster. Azure Machine Learning compute clusters are scalable machine learning platforms that consist of one or more CPU or GPU nodes. Cluster resources can be shared with other users in the machine learning workspace. Compute clusters support:

- AutoML, which is used to automate the process of training and tuning machine learning models.
- Machine learning pipelines, which are machine learning workflows.
- Azure Machine Learning designer, which facilitates graphical, drag-and-drop creation of machine learning models.

You should not install the Azure Machine Learning SDK on your local computer. Using your local computer allows you to preform machine learning testing without incurring cloud compute costs. Once you have refined your machine learning scripts, you can deploy them on scalable Azure cloud compute resources. Local computer training targets support AutoML, but they do not support machine learning pipelines or Azure

Machine Learning designer.

You should not deploy an Azure Machine Learning compute instance. Azure Machine Learning compute instances are highly scalable cloud compute resources.

You should not create and deploy a remote virtual machine. Like using your local computer, a remote virtual machine allows you to leverage remote compute resources located outside of the Azure cloud. However, these systems may or may not be scalable, and they do not support Azure Machine Learning designer.

References

What are compute targets in Azure Machine Learning?

Set up and use compute targets for model training

Question 63

You register an Azure Machine Learning (ML) model. You enable model data collection to collect data from your AKS deployment of the model, and you see data being collected in model data blob container.

You write the following script:

```
from azureml.core import Experiment, Run, RunDetails
from azureml.datadrift import DataDriftDetector,
AlertConfiguration

alert_config =
AlertConfiguration('your_email@abcinc.com')

datadrift = DataDriftDetector.create(ws, model.name,
model.version, services, frequency="Day",
alert_config=alert_config, drift_threshold=0.3)

target_date = datetime.today()
run = datadrift.run(target_date, services,
feature_list=feature_list, create_compute_target=True)
```

For each of the following statements about the drift configuration in the code, select Yes if the statement is true. Otherwise, select No.

Statement	**Yes**	**No**
DataDriftDectector would send an email if the drift_coefficient is 0.4 when evaluated.	○	○
After the configuration, this script will run drift detection on the day the script is executed.	○	○

The first run of drift detection will execute on an existing compute cluster in your Azure ML environment.	○	○

Explanation

DataDriftDectector would send an email if the drift_coefficient is 0.4 when evaluated. Drift_threshold is configured to be 0.3 in the DataDriftDectector.create method. This configuration along with the alert_config parameter ensures that when a drift_coefficient of 0.3 or higher is detected, an email is triggered. A drift coefficient of 0 indicates no drift, and 1 indicates maximum drift.

After the configuration, this script will run drift detection on the day the script is executed. Datadrift.run is triggered passing datetime.today(), which triggers a run of the datadrift evaluation the day the script is executed.

The first run of drift detection will not execute on an existing compute cluster in your Azure ML environment based on the parameter settings. The datadrift.run methods has a parameter create_compute_target set to True. Setting the value to True creates a new compute cluster for the execution of the run.

References

[Collect data from models in production]

[DataDriftDetector class]

Question 64

You plan to monitor and debug pipelines by routing Azure Machine Learning logs to Azure Application Insights using the OpenCensus Python library.

You have several query requirements and need to add fields to a Custom Dictionary to meet each of them.

Which fields should you add in each case? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

Query Requirements	Field
Differentiate between training and scoring runs.	
Focus on a specific issue.	
View logs for all steps over time.	

Select your answ er

Select your answ er

Select your answ er

Explanation

You can use the OpenCensus Python library to forward Azure Machine Learning logs to Application Insights to provide advanced pipeline monitoring, tracing, and metrics tracking. To do this, you define a Custom Dimensions dictionary which provides field names for your logging data. These fields can then be specified in Application Insight queries.

You should specify the run_type field to differentiate between training and scoring runs. In machine learning, you train a model by applying an algorithm to input data. Scoring is the process of using a trained model to predict future values.

You should specify the step_id field to focus on a specific issue. This allows you to view details for a specific step. For example, you could query the logs for all entries with the same step_id.

You should specify the parent_run_id field to view logs for all steps over time. Pipelines are machine learning workflows composed of individual steps. By querying the parent_run_id, you can view information for all steps that are part of a run.

References

Collect machine learning pipeline log files in Application Insights for alerts and debugging

What is Application Insights?

Python

Question 65

Your organization uses field-based sensors to monitor electricity usage throughout its manufacturing facilities.

You need to apply a machine learning model to data collected by your devices before it is shipped to the cloud.

What should you do?

Choose the correct answer

- Register your model in an Azure Machine Learning workspace. Deploy the model to Azure Kubernetes Service (AKS).
- Install the Azure Machine Learning SDK preview package. Deploy your machine learning model as an app in Azure Functions.
- Install Docker on your local machine. Deploy your model using a local web service.
- Create an IoT Edge module. Deploy the module and machine learning model using Azure Stack Edge.

Explanation

You should create an IoT Edge module and deploy the module and machine learning model using Azure Stack Edge. Azure Stack Edge is a hardware-as-a-service platform offered by Microsoft. As part of an Azure Stack Edge subscription, Microsoft provides a hardware-accelerated compute device that can process and analyze data collected by IoT devices before it is sent to the cloud. In addition to applying your machine learning models to IoT data, Stack Edge can filter, aggregate, and optimize data in order to reduce bandwidth requirements.

You should not register your model in an Azure Machine Learning workspace and deploy it to AKS. AKS supports highly scalable compute options for Azure Machine Learning experiments. In addition to supporting multiple-node clusters, AKS can be used for experiments that require hardware acceleration via GPU or Field-Programmable Gate Arrays (FPGA).

You should not install the Azure Machine Learning SDK preview package and deploy your machine learning model as an app in Azure Functions. Once machine learning training is complete, you can create a Docker image based on the trained model. This model can then be deployed as an Azure Functions as containerized code.

You should not install Docker on your local machine and deploy your experiments using a local web service. This option is used for limited testing and troubleshooting. Depending on the capabilities of your local machine and the libraries installed in your machine learning container, hardware acceleration may or may not be supported. When using GPU for inference, local web service is not supported.

References

What is Azure Stack Edge?

Deploy a model to an Azure Kubernetes Service cluster

Deploy a machine learning model to Azure Functions (preview)

Deploy models with Azure Machine Learning

Question 66

You plan to deploy a batch inferencing pipeline, and you want to use ParallelRunStep. In order to create the ParallelRunStep, you define the ParallelRunConfig as follows:

```
parallel_run_config = ParallelRunConfig(
    source_directory=experiment_folder,
    entry_script="batch_infections.py",
    mini_batch_size="5",
    error_threshold=10,
    output_action="append_row",
    environment=batch_env,
    compute_target=inference_cluster,
    node_count=2)
```

You create a parallel run step and submit the pipeline to execute for an experiment as follows. The input dataset consumption configuration is declared as input_infections_ds_consumption.

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



parallelrun_step = ParallelRunStep(
name="predict-infections",
parallel_run_config=parallel_run_config,
inputs=[input_infections_ds_consumption],
output=output_dir,
allow_reuse=True
)

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

You need to download the output files that are generated by the experiment to a local folder.

Which script should you execute?

Choose the correct answer

- import tempfile

```
batch_run = pipeline_run.find_step_run(parallelrun_step.name)[0]
batch_output = batch_run.get_output_data(pipeline.name)

target_dir = tempfile.mkdtemp()
batch_output.download(local_path=target_dir)
```

- import tempfile

```
batch_run = pipeline_run.find_step_run(parallelrun_step.name)
batch_output = batch_run.get_output_data(output_dir.name)

target_dir = tempfile.mkdtemp()
batch_output.download(local_path=target_dir)
```

- import tempfile

```
batch_run = pipeline_run.find_step_run(parallel_run_config.name)[0]
batch_output = batch_run.get_output_data(output_dir.name)

target_dir = tempfile.mkdtemp()
batch_output.download(local_path=target_dir)
```

- import tempfile

```
batch_run = pipeline_run.find_step_run(parallelrun_step.name)[0]
batch_output = batch_run.get_output_data(output_dir.name)

target_dir = tempfile.mkdtemp()
batch_output.download(local_path=target_dir)
```

Explanation

You should execute the following script:

```
import tempfile


batch_run =
pipeline_run.find_step_run(parallelrun_step.name)[0]
batch_output = batch_run.get_output_data(output_dir.name)


target_dir = tempfile.mkdtemp()
batch_output.download(local_path=target_dir)
```

You should first get the last execution of the parallel run step from the pipeline_run object. You would also read the output data in a batch_output object. You can then create a local target directory and call the download of the batch_output object to download the output files.

You should not execute:

```
import tempfile


batch_run =
pipeline_run.find_step_run(parallelrun_step.name)
batch_output = batch_run.get_output_data(output_dir.name)


target_dir = tempfile.mkdtemp()
batch_output.download(local_path=target_dir)
```

The batch_run should only reference the last execution and not the array of all the executions. You need to specify the index of 0 to the results of the find_step_run method in order to get the last execution.

You should not execute:

```
import tempfile


batch_run =
pipeline_run.find_step_run(parallelrun_step.name)[0]
batch_output = batch_run.get_output_data(pipeline.name)


target_dir = tempfile.mkdtemp()
batch_output.download(local_path=target_dir)
```

You need to specify the output_dir name in the get_output_data method. Specifying the pipeline name will not provide the path where the output data is located.

You should not execute:

```
import tempfile


batch_run =
pipeline_run.find_step_run(parallel_run_config.name)[0]
batch_output = batch_run.get_output_data(output_dir.name)


target_dir = tempfile.mkdtemp()
batch_output.download(local_path=target_dir)
```

You need to provide the parallelrun_step name in the find_step_run method. parallel_run_config will not provide the step object executed by the experiment.

References

[ParallelRunConfig class]

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

Question 67

You want to create a datastore in Azure Machine Learning (ML) studio. Before that, you need to create an instance of the storage where you want your relational data to reside. You navigate to the Datastore option in Azure ML studio.

You need to select the storage option for your data.

Which source storage type should you select?

Choose the correct answer

- Azure Blob Container
- Azure Data Lake Gen2
- Databricks File System
- Azure Database for PostgreSQL

Explanation

You should select Azure Database for PostgreSQL. Azure Database for PostgreSQL is one of the supported cloud-based storage services in Azure for Azure Machine Learning to consume data as a datastore. It also provides support for storing relational data.

The other supported storage services are:

- Azure Blob Container
- Azure File Share
- Azure Data Lake
- Azure Data Lake Gen2
- Azure SQL Database
- Databricks File System
- Azure Database for MySQL

You should not select Azure Blob Container, Azure Data Lake Gen2 or Databricks File System. They do not provide support for storing relational data.

References

[Secure data access in Azure Machine Learning]

Question 68

You create a Machine Learning (ML) model using Azure Machine Learning designer. You click train the model, click the Publish button on the designer canvas, select the Create new radio button, and click Publish.

You need to determine the effect publishing has on the service.

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

Statement	Yes	No
The Publish button will deploy the model as a web service endpoint.	○	◉
The Publish button publishes the pipeline with a REST endpoint.	◉	○
The Publish button runs the pipeline against the test data provided.	○	◉

Explanation

The Publish button will not deploy the model as a web service endpoint. To deploy the model as a web service endpoint, you need to navigate to the Models menu on ML studio, select the model you want to deploy, and click the Deploy button.

The Publish button creates a REST endpoint to the pipeline that other users/developers/data scientists can make calls to. It provides an endpoint with a key-based authentication.

The Publish button does not run the pipeline. To run the pipeline against the test dataset, which is part of the model, you need to click Submit on the designer canvas.

References

What is Azure Machine Learning designer?

Question 69

You prepare the following code to register an Azure blob container as a datastore:

```
my_blob_datastore_name='azure-blob-datastore'
my_container_name=os.getenv("BLOB_CONTAINER", "azure-
blob-container")
my_account_name=os.getenv("BLOB_ACCOUNTNAME", "azure-
account")
my_account_key=os.getenv("BLOB_ACCOUNT_KEY", "12345")

my_blob_datastore =
Datastore.register_azure_blob_container(
workspace=ws,
datastore_name=my_blob_datastore_name,
container_name=my_container_name,
account_name=my_account_name,
account_key=my_account_key)
```

You need to ensure that azure-blog-datastore, which is located on a virtual network, is accessible.

What should you do?

Choose the correct answer

- Add the skip_validation=True parameter to the register_azure_blob_container method.
- Add the filesystem= ‘virtual’ parameter to the register_azure_data_lake_gen2 method.
- Add the following line of code: blob_store = Datastore.get(ws, datastore_name='virtual')
- Create an Azure file share datastore named azure-blob-datastore.

Explanation

You should add the skip_validation=True parameter to the register_azure_blob_container method. Azure Storage provides a layered security model that limits access to data based on network types and IP addressing information. If your blob container is in a virtual network, adding the skip_validation parameter will bypass validation of storage keys.

You should not create an Azure file share datastore named azure-blob-datastore. Azure Machine Learning supports different Azure storage services when creating datastores, such as blob containers, file shares, and databases.

You should not add the filesystem= ‘virtual’ parameter to the register_azure_data_lake_gen2 method. The filesystem parameter allows you to define the name of a Data Lake Gen2 filesystem name.

You should not add the following line of code: blob_store = Datastore.get(ws, datastore_name='virtual'). This code will retrieve the datastore named ‘virtual’ and store it as a reference in the blob_store variable.

References

Connect to Azure storage services

Register an Azure blob container as a datastore

Question 70

You use Azure Machine Learning designer to create and publish machine learning pipelines as web services.

You need to configure the correct machine learning task for each compute target.

Based on the compute target, which machine learning task should you perform? To answer, select the most appropriate machine learning task from the drop-down menus.

Choose the correct options

Compute Target	Machine Learning Task
Azure Machine Learning compute cluster	
Azure Machine Learning compute instance	
Azure Kubernetes Service (AKS)	

Select your answ er

Select your answ er

Select your answ er

Explanation

You should perform training tasks only on an Azure Machine Learning compute cluster. 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. Azure Machine Learning compute clusters can be used for training pipelines, but not for pipeline deployment because this functionality is not supported.

You should perform training tasks only on an Azure Machine Learning compute instance. 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. Unlike a compute cluster, a compute instance cannot scale down to zero, meaning that usage charges accrue unless you power off the VM. Azure Machine Learning compute instances can be used for training pipelines, but not for pipeline deployment because this functionality is not supported.

You should perform deployment only on AKS. AKS clusters are designed for heavy, real-time production workloads. One of the primary benefits of deploying to AKS is its support for autoscaling. This means that as workload increases or decreases, an AKS cluster can add or terminate cluster nodes. In addition to supporting multiple-node clusters, AKS can be used for experiments that require hardware acceleration via GPU or Field-Programmable Gate Arrays (FPGAs).

References

[What is Azure Machine Learning designer (preview)?]

[What are compute targets in Azure Machine Learning?]

Question 71

You use Azure Machine Learning Studio to create a drag-and-drop designer pipeline.

You need to ensure that a dictionary can be logged with each experiment run.

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

Choose the correct answers

- [ ] Drag the Execute Python Script module onto the designer canvas.
- [ ] Submit the experiment with a ScriptRunConfig object.
- [ ] Define a custom script and add the run.start_logging method.
- [ ] Add the run.log_table method under the azureml_main entry point function.

Explanation

You should drag the Execute Python Script module onto the designer canvas. 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.

You should also add the run.log_table method to the code editor. 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. 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 it allows you to store key values along with associated data.

You should not define a custom script and add the run.start_logging method. The run.start_logging method is used for interactive runs, such as those executed from a Jupyter notebook. This method logs metrics to the experiment's run record.

You should not submit the experiment with a ScriptRunConfig object. 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.

References

Execute Python Script module

Run class

ScriptRunConfig class

Question 72

You are using Azure Machine Learning (ML) designer to train a binary classification model.

You add a historical dataset that has features and outcome labels. You create the pipeline in the designer that has the Two-Class Decision Forest module for binary classification, a train module that takes inputs from the dataset, and the Two-Class Decision Forest module.

You need to use a random subset of the data as training data to know the score and evaluate the model to determine the accuracy of inferences.

Which other module should you use to implement this requirement?

Choose the correct answer

- Clip Values
- Normalize Data
- Join Data
- Split Data

Explanation

You should use the Split Data module. This module is used when you need to separate data into training and testing sets. You can customize the way data is divided. The module also provides options to randomize data selection.

You should not use the Join Data module. This module is used to merge two datasets using a database-style join operation.

You should not use the Normalize Data module. Normalization is a technique that is often applied as part of data preparation for machine learning. You would change the numeric data in the columns to use a common scale without distorting differences in the range of values or losing information.

You should not use the Clip Values module. This module allows you to replace data values that are above or below a specified threshold with a mean, a constant, or other substitute value.

References

Split Data module

Join Data

Normalize Data module

Clip Values

Question 73

You are responsible for training machine learning (ML) models for your organization. You have various requirements to train and evaluate models.

You need to decide the optimal compute target for your requirements with minimal configuration.

What compute target should you select? To answer, drag the appropriate compute resource to each requirement. A compute resource may be used once, more than once, or not at all.

Explanation

You should use Azure ML compute cluster to tune hyperparameters using Azure ML designer. This is the only training target that is supported by Azure ML designer for Automated ML jobs running hyperparameter tuning.

You should select Remote VM when using your own virtual machine (VM) for hyperparameter tuning. Azure ML supports bringing in a VM that is reachable by Azure ML. You can have the VM attached to your virtual network. This provides the benefit of leveraging environments like conda or Python as well as running training in a containerized environment.

You should select Azure HDInsight to use Apache Spark to train your models. Azure HDInsight provides a pre-configured environment with Apache Spark.

You should select Azure ML compute cluster to auto scale instances for models based on compute requirements. Azure ML compute cluster can be configured to scale up when training jobs are submitted.

References

Set up and use compute targets for model training

Question 74

Your organization is developing a model for the classification of images into multiple classes. You write the following code:

```
from azureml.train.hyperdrive import
GridParameterSampling
from azureml.train.hyperdrive import choice


param_sampling = GridParameterSampling( {
        "num_hidden_layers": choice(1, 2, 3),
        "batch_size": choice(16, 32, 64)
    }
)
```

You need to configure the hyperparameter values when training the model.

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

Statement	Yes	No
The parameter sampling has six samples.	○	○
If you want to increase the sampling size, you can use uniform(0.05, 0.1) for batch_size.	○	○
You can create discrete hyperparameters using a range object for grid sampling.	○	○

Explanation

The parameter sampling will have 9 possible samples based on 3 choices for num_hidden_layers, multiplied by 3 values for batch size.

If you want to increase the sampling size, you cannot use uniform(0.05, 0.1) for batch_size. GridParameterSampling does a grid search over all possible values defined in the search space. The hyperparameters have to be specified using a choice function. Specifying uniform() does not allow grid sampling.

You can create discrete hyperparameters using a range object for grid sampling. GridParameterSampling allows use of a range within a choice function to create a list of discrete values for the search space. For example, you could set batch_size to choice(range(1, 5)) for a sampling of 1, 2, 3, and 4.

References

Tune hyperparameters for your model with Azure Machine Learning

GridParameterSampling class

Question 75

You design and log a monitoring solution for a Linux-based Azure HDInsight solution that is used to analyze large volumes of data. Various operations will execute continuously on the cluster.

You need the logging solution to be able to monitor the performance of the cluster. You want to minimize the configuration effort associated with implementing the monitoring solution.

What should you use?

Choose the correct answer

- Azure Sentinel
- Azure Log Analytics
- HDInsight .Net SDK
- Apache Ambari

Explanation

You should use Apache Ambari. Apache Ambari simplifies the management and monitoring of Hadoop clusters by providing an easy-to-use web UI backed by its REST APIs. Apache Ambari is provided by default with a Linux-based HDInsight cluster, so configuration efforts are reduced.

You should not use Azure Log Analytics. Azure Log Analytics is suited for querying diagnostics logs for resources created in Azure. You can also implement customized logging to surface logs in a Log Analytics workspace, but that would increase the effort associated with the configuration.

You should not use Azure Sentinel. Azure Sentinel is a cloud security information event management (SIEM) solution. Azure Sentinel would not provide insights into the performance metrics within a cluster.

You should not use HDInsight .Net SDK. The effort associated with creating a logging framework would be higher than by using Ambari.

References

Manage HDInsight clusters by using the Apache Ambari REST API

What is Azure Sentinel?

Azure HDInsight SDK for .NET

Question 76

You are asked to write a Python script to register a dataset using Azure Machine Learning libraries. The data for the dataset resides in multiple CSV files located in the default datastore storage account in mydatacontainer.

You need to register the dataset using the name global-infections.

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

Choose the correct options

```
from azureml.core import Workspace, Dataset

subscription_id = '12345678-xdad-yde5-8888-999999999999'
resource_group = 'mlResourceGroup'
workspace_name = 'my-ml-workspace-01'
ws= Workspace(                    ,
                                   ,
                                   )
default_ds =                    .get_default_datastore()
tab_data_set = Dataset.Tabular.from_delimited_files(
                  path=(                ,
                  'mydatacontainer/*.csv'))
try:
    tab_data_set = tab_data_set.register(
                        workspace=                  ,
                        name='global-infections',
                        description='infections data',
                        tags = {'format':' csv '},
                        create_new_version=True)
except Exception as ex:
    print(ex)
```

Select your answer

Select your answer

Select your answer
Select your answer
Select your answer
Select your answer

Explanation

You should complete the code 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='infections
data',

                               tags = {'format':'csv'},
                               create_new_version=True)
except Exception as ex:
    print(ex)
```

The workspace object takes subscription id, resource group, and workspace name as parameters to get a reference to the workspace.

The workspace has a method named get_default_datastore that returns the default datastore. To get the reference to the datastore, you need to use the workspace reference ws to call the get_default_datastore method.

To create the tabular dataset, you should use Dataset.Tabular.from_delimited_files and pass it the reference to the default_ds datastore.

Dataset.register is used to registers the datasets. You have to provide the workspace reference, ws, using the workspace parameter and specify the format.

References

[Dataset class]

Question 77

You use Azure Machine Learning to transform images and perform object detection within each image. The experiment output appears to be incorrect and you plan to use logs to troubleshoot experiment run errors.

You need to upload each transformed PNG image to the run record along with an array of RGB values from your transformation model. The images must be visible in the run record.

Which logging method should you use to generate the required logging information? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

Required Logging Information	Logging Method
RGB array values	
Transformed PNG image	

Select your answ er

Select your answ er

Explanation

You should use the run.log_list method to log an array of RGB values to the run record. 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.

The run.log_list method is useful when you need to log lists of values, arrays, dictionaries, vectors, or any other non-scalar value. Vectors appear in an experiment's run details as single-variable line charts.

You should use the run.log_image method to log transformed PNG images to the run record. These images appear in the experiment's run details. You can also use this method to log matplotlib plots to the run record. Plots are useful for visualizing experiment output.

You should not use the run.log method. This method can be used to log scalar string or numerical values.

You should not use the run.upload_file method. This method can be used to upload files that are stored with the run record as artifacts.

References

[Azure / MachineLearningNotebooks]

[Run class]

Question 78

You are writing code to tune your hyperparameters through a Hyperdrive experiment. The sampling method selected should allow you to specify discrete values.

You complete the necessary initialization of your workspace variables named ws, register datasets named "infection dataset", and create a script named infection_training.py for your estimator that has the init and run methods. You also include the necessary libraries.

You need to complete the script for your Hyperdrive experiment. Complete the code below. To answer, select the appropriate options from the drop-down menus.

Choose the correct options

```
params =                              (
    {
        '--regularization': choice(0.001, 0.005, 0.01, 0.05, 0.1, 1.0)
    }
)

infection_ds = ws.datasets.get("infection dataset")

hyper_estimator = SKLearn(source_directory=experiment_folder,
                          inputs=[infection_ds.as_named_input('infection')],
                          pip_packages=['azureml-sdk'],
                          entry_script='                    ',
                          compute_target = training_cluster)

hyperdrive = HyperDriveConfig(estimator=hyper_estimator,
                          hyperparameter_sampling=params,
                          policy=None,
                          primary_metric_name='AUC',
                          primary_metric_goal=PrimaryMetricGoal.MAXIMIZE,
                          max_total_runs=6,
                          max_concurrent_runs=4)

experiment =             (workspace = ws, name = 'infection_training_hyperdrive')
run = experiment.            (config=hyperdrive)
```

Select your answ er

Select your answ er

Select your answ er

Select your answ er

Explanation

You should select GridParameterSampling for the sampling method. GridParameterSampling allows you to set discrete values for your hyperparameters.

You should not select BayesianParameterSampling since it will try to intelligently pick values for hyperparameters based on a provided parameter space. It will not allow you to specify specific discrete values.

You should not select RandomParameterSampling since the tuning parameter values are selected randomly over a parameter space provided. It will not allow you to specify specific discrete values.

You should specify the file name infection_training.py where the init and run methods are present for the entry script. The entry script looks for these 2 methods. You cannot specify a dataset infection_ds as an entry point script.

You should select the Experiment class from the Azure ML SDK in order to create the experiment. You should specify the workspace and name as parameters.

You should not select the Workspace class. The Workspace class does not allow the creation of an experiment.

You should submit the experiment specifying the hyperdrive config to execute the tuning.

You should not use the get_runs method. The get_runs method will not execute the experiment. It will only get a list of runs that have already been executed by the experiment.

References

[Tune hyperparameters for your model with Azure Machine Learning]

Question 79

You create a model that will forecast the sales for your organization based on number of features and labels provided by historical data.

You need to programmatically create a pipeline that runs a processing script to load data from the datastore and feed the data to a Machine Learning model training script. The Pipeline, PipelineData, and PythonScriptStep libraries are already imported.

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

Choose the correct options

```
datastore = ws.get_default_datastore()

data_output = PipelineData("processed_data", datastore=datastore)


process_step = PythonScriptStep(script_name="process.py",
                    arguments=["--data_for_train",              ],
                    outputs=[              ], compute_target=aml_compute,
                    source_directory=process_directory)


train_step = PythonScriptStep(script_name="train.py",
                    arguments=["--data_for_train",               ],
                    inputs=[              ], compute_target=aml_compute,
                    source_directory=train_directory)

pipeline = Pipeline(workspace=ws, steps=[process_step, train_step])
```

Select your answer ▼

Select your answer ▼

Select your answer ▼

Select your answer ▼

Explanation

You should complete the code as follows

```
datastore = ws.get_default_datastore()

data_output = PipelineData("processed_data", datastore
=datastore)

process_step = PythonScriptStep(script_name="process.py",
arguments=["--data_for_train", data_output],
outputs=[data_output], compute_target=aml_compute,
Source_directory=process_directory)

train_step = PythonScriptStep(script_name="train.py",
arguments=["--data_for_train", data_output],
inputs=[data_output], compute_target=aml_compute,
source_directory=train_directory)
pipeline = Pipeline(workspace=ws, steps=[process_step,
train_step])
```

In the process step, you pass the dataset named data_output as the data to train parameter. You can provide the same dataset as the output parameter to store the output of the process step to generate the dataset that will be used in the train step.

In the train step, you can use data_output as the input for the training module.

References

[PipelineData class]

Question 80

You are planning to create a compute target named cpu-cluster using the STANDARD_D2_V2 virtual machine (VM) size.

You write the following Python script to create the compute target. You have a reference to the workspace in the ws variable in the script.

```
from azureml.core.compute import ComputeTarget,
AmlCompute
from azureml.core.compute_target import
ComputeTargetException
cpu_cluster_name = "cpu-cluster"
try:
       cpu_cluster = ComputeTarget(workspace=ws,
name=cpu_cluster_name
       print('Found existing cluster, use it.')
except ComputeTargetException:
       compute_config =
AmlCompute.provisioning_configuration(vm_size='STANDARD_D
2_V2',

     max_nodes=4)
       cpu_cluster = ComputeTarget.create(ws,
cpu_cluster_name, compute_config)
cpu_cluster.wait_for_completion(show_output=True)
```

You execute the script.

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

Statement	Yes	No
The new Azure Machine Learning compute instance is created with a maximum of four nodes.	○	○

The call ComputeTarget(workspace=ws, name=cpu_cluster_name) returns a reference to the compute target.	○	○
Logs are printed after execution is completed.	○	○

Explanation

The new Azure Machine Learning compute is created with a maximum of four nodes. The max_nodes parameter passed to provisioning_configuration configures the config to have a maximum of four nodes. The create method uses this configuration to create the compute target.

The call ComputeTarget(workspace=ws, name=cpu_cluster_name) does not return a reference to the compute target. The compute target does not exist. If you call this method before you create the compute target, it will fail and throw an exception.

Logs are printed after execution is completed. The method call wait_for_completion with the attribute show_output=True configures logging to only show logs after the script execution is completed.

References

[ComputeTarget class]

[What is the Azure Machine Learning SDK for Python?]

Question 81

You use the Azure Machine Learning SDK to create an automated machine learning pipeline. Your source data contains salary information, and some values may be missing.

You need to ensure the pipeline can load your data and run without errors.

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

Choose the correct answers

- ☐ Create a FileDataset.
- ☐ Set the featurization experiment parameter to auto.
- ☐ Create a TabularDataset.
- ☐ Store source data files in the ./logs directory.
- ☐ Configure a local compute target.

Explanation

You should Create a TabularDataset and set the featurization experiment parameter to auto. Automated machine learning requires that input data be stored in tabular form. 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.

When configuring a pipeline using Azure Machine Learning SDK, you can define automated machine learning settings that control how the experiment is run. These settings are typically defined in dictionary format and passed to the AutoMLConfig class. If the featurization parameter is set to auto, input data will automatically be preprocessed and missing values will be handled.

You should not configure a local compute target. Machine learning pipelines require remote compute and cannot be run on a local compute target. Remote compute targets can reside inside or outside of the Azure cloud.

You should not store source data files in the ./logs directory. Azure Machine Learning supports several locations for storing experiment output. Files can be saved to storage on the local compute instance; however, 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 create a FileDataset. You create a file dataset to reference the unstructured file or files you want to use in your machine learning experiments.

References

Configure automated ML experiments in Python

Tutorial: Use automated machine learning to predict taxi fares

Dataset class

What are compute targets in Azure Machine Learning?

Where to save and write files for Azure Machine Learning experiments

Question 82

You plan to a create a compute environment for your Image Classification Deep Learning model leveraging Compute Unified Device Architecture (CUDA) computations. You plan to use the Data Science Virtual Machine (DSVM) for this purpose.

You need to configure your VM to support CUDA.

Which series should you configure your VM size?

Choose the correct answer

- DSv2
- Fs
- Lsv2
- NCv2

Explanation

You should configure your VM size to the NCv2 series. In order for your model to run image classification deep learning model leveraging CUDA, you need to configure the VM that provides support for GPUs.

You should not configure your VM to the DSv2 series. DSv2 series VMs are general purpose VMs and don't support GPUs. For running image classification deep learning model leveraging CUDA, GPU-based graphic processing is recommended.

You should not configure your VM to the Fs series. Fs series VMs are compute optimized VMs and don't support GPUs. These are good for CPU intensive workloads. For running image classification deep learning model leveraging CUDA, GPU-based graphic processing is recommended.

You should not configure your VM to the Lsv2. Lsv2 series VMs are storaged optimized VMs. This type of VMs is good for high disk IO operations like hosting databases. For running image classification deep learning model leveraging CUDA, GPU- based graphic processing is recommended.

References

5 Steps to More Interactive Deep Learning

CUDA Toolkit

Deep learning and AI frameworks for the Azure Data Science VM

Sizes for virtual machines in Azure

Question 83

You use Azure Machine Learning studio to manage compute targets.

You need to deploy the appropriate compute target for different machine learning tasks.

Which compute target should you create for each machine learning task? To answer, select the appropriate compute targets from the drop-down menus.

Choose the correct options

Machine Learning Task	Compute Target
Deploy trained models to provide real-time predictive services at scale.	
Support scalable, on-demand processing using low-priority virtual machines (VMs).	
Use Azure Databricks clusters.	

Select your answ er

Select your answ er

Select your answ er

Explanation

You should create an inference cluster to deploy trained models that will provide real-time predictive services at scale. 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 Kubernetes Service (AKS) and are sometimes referred to as AKS clusters.

Compute clusters cannot be used with models that offer predictive services because they use low-priority VMs and may not scale to the load required. Additionally, inference clusters are used for real-time predictive services and computer cluster VM availability is not guaranteed, thus the provided service is not real-time. Similarly, attached compute resources cannot be used for real-time inference because they do not scale to the workloads involved in real-time inference.

You should create a compute cluster to support scalable, on-demand processing using low-priority VMs. 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 VMs, which do not have guaranteed availability. Using low-priority VMs can help reduce machine learning costs.

AKS target could theoretically be used to Support scalable, on-demand processing. However, AKS is designed for compute-intensive operations at scale and is best suited for real-time predictive services. Attached compute should not be used to support scalable, on-demand processing because it does not provide scalability.

You should use attached compute to use Azure Databricks clusters. Attached compute supports linking to on-cloud Azure compute resources, including VMs and Azure Databricks clusters. Azure Databricks is a cloud-based Apache Spark-based analytics platform. Databricks clusters are often used to run machine learning pipelines. You might choose Azure Databricks if you are collaborating with other machine learning teams, or because it is based on the open-source Apache Spark ecosystem. Neither inference cluster nor compute clusters can use Azure Databricks clusters.

References
[Create compute resources]
[What are compute targets in Azure Machine Learning?]

[Set up and use compute targets for model training]

[Clusters]

Question 84

You have the following code that creates a schedule:

```
from azureml.pipeline.core import Schedule
from azureml.core.datastore import Datastore

datastore = Datastore(workspace=ws,
name="workspaceblobstore")
schedule = Schedule.create(workspace,
name="TestSchedule", pipeline_id="pipeline_id"
                              experiment_name="helloworld
", datastore=datastore,
                              polling_interval=5,
path_on_datastore="file/path")
```

You are trying to understand the schedule configuration.

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

Statement	Yes	No
The pipeline is scheduled to run on a time interval.	○	○
Every time the data in the datastore changes, the helloworld experiment is executed.	○	○
The helloworld experiment will be executed at a polling_interval of 5 minutes.	○	○

Explanation

The pipeline is not scheduled to run on a time interval. Schedule.create configures the runs dependent on the datastore changes, and not based on a fixed time interval.

Every time the data in the datastore changes, the helloworld experiment is executed. Schedule.create configures the runs dependent on the datastore changes using the datastore and the path_on_datastore parameter. The datastore is polled every 5 minutes to check for any changes.

The helloworld experiment will not be executed at a polling_interval of 5 minutes. The polling_interval configures the frequency at which the datastore is checked for any changes. It does not set an experiment run interval.

References

Schedule class

Question 85

You have an imbalanced dataset for which you want Azure Automated Machine Learning (ML) to extract features.

You have the following requirements based on the dataset:

- Auto reduce linear dimensionality.
- Features should be scaled by their maximum absolute value.
- Standardize features extracted to ensure that data is normally distributed.

You need to determine whether these requirements would be met by the Automated ML experiment.

What three scaling/normalization techniques are required by Automated ML to meet the requirements? Each correct answer presents part of the solution.

Choose the correct answers

- [ ] Encoding
- [ ] MaxAbsScaler
- [ ] Transformation
- [ ] StandardScaleWrapper
- [ ] PCA

Explanation

The MaxAbsScaler, PCA, and StandardScaleWrapper normalization or scaling techniques need to be applied to the dataset for the algorithms to perform well.

MaxAbsScaler ensures that each feature is scaled by its maximum absolute value.

PCA ensure linear dimensionality reduction. It uses singular value decompositions of the data and reduces the dimensional space.

StandardScaleWrapper ensures that the features are standardized by altering the dataset, removing the mean, and scaling to unit variance.

You should not use Encoding and Transformation. You can add additional feature engineering techniques, but these are applied when only automatic featurization is enabled.

References

What is automated machine learning (AutoML)?

Question 86

You create and train an Azure Machine Learning (ML) model using Azure ML studio. The model takes in comma-separated (CSV) data as the source.

You need to identify the steps to deploy the model as a real-time service and ensure that the endpoints are accessible to users of the service.

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

Create a list in any order

Possible actions

Explanation

You should perform the following actions:

- Create a real-time inference pipeline.
- Create an inferencing cluster.
- Deploy the real-time endpoint.
- Test the real-time endpoint.

You should create a real-time inference pipeline from the trained model. You need to create the inference pipeline to deploy the model as a web service.

You should also create an inferencing cluster where the pipeline will be hosted. You can perform these first two steps in any order.

Then, you need to select the Deploy button in the designer view for the inference model to deploy the services. Once deployed, you should go to the test tab to ensure that the real-time endpoints are accessible by providing a data row, and click the Test button.

You should not create a batch inference pipeline. Batch inference pipelines are created for running large amounts of data in a batch rather than providing a real-time end point.

You should not create a compute cluster. A compute cluster can be used for training the model. To deploy an inference pipeline, you need to create an inference cluster.

References

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

Question 87

You create compute targets for Azure Machine Learning training experiments.

You need to ensure that each target supports GPU for training.

What should you do?

Choose the correct answer

Install Docker on your local machine. Deploy your experiments using a local web service.

Provision Azure Internet of Things (IoT) Hub and register IoT devices. Deploy machine learning models on your IoT devices.

Register the model in a workspace. Deploy the model to Azure Kubernetes Service (AKS).

Install the Azure Machine Learning SDK preview package. Deploy your machine learning model as an app in Azure Functions.

Explanation

You should register the model in a workspace and deploy your experiments to AKS. AKS supports highly scalable compute options for Azure Machine Learning experiments. In addition to supporting multiple-node clusters, AKS can be used for experiments that require hardware acceleration via GPU or field-programmable gate arrays (FPGA). Finally, AKS can dynamically scale compute availability based on workload.

You should not install Docker on your local machine and deploy your experiments using a local web service. This option is used for limited testing and troubleshooting. Depending on the capabilities of your local machine and the libraries installed in your machine learning container, hardware acceleration may or may not be supported. When using GPU for inference, local web service is not supported.

You should not install the Azure Machine Learning SDK preview package and deploy your machine learning model as an app in Azure Functions. Once machine learning training is complete, you can create a Docker image based on the trained model. This model can then be deployed into Azure Functions as containerized code.

You should not provision Azure IoT Hub, register IoT devices, and deploy machine learning models on your IoT devices. This deployment option is known as Azure Data Box Edge. This option is useful when you want to apply machine learning models to data collected by IoT edge devices before the data is shipped to the cloud.

References

Deploy models with Azure Machine Learning

Deploy a model to an Azure Kubernetes Service cluster

Deploy a machine learning model to Azure Functions (preview)

What is Azure Stack Edge?

Question 88

You want to run machine learning tasks on image files stored in data/files/images.

You need to create and register a dataset for a workspace named new_project. Only image files should be processed.

Which code block should you execute?

Choose the correct answer

- from azureml.core import Dataset
blob_ds = ws.get_default_datastore()
file_ds = Dataset.File.from_files(path=(blob_ds, 'data/files/images/*.png'))
file_ds = file_ds.register(workspace=ws, name='new_project_one')

- from azureml.core import Dataset
blob_ds = ws.get_default_datastore()
file_ds = Dataset.File.from_files(path=(blob_ds, 'data/files/images/1.png'))
file_ds = file_ds.register(workspace=ws, name='new_project')

- from azureml.core import Dataset
blob_ds = ws.get_default_datastore()
file_ds = Dataset.File.from_files(path=(blob_ds, 'data/files/images/*.png'))
file_ds = file_ds.register(workspace=ws, name='new_project')

- from azureml.core import Dataset
blob_ds = ws.get_default_datastore()
file_ds = Dataset.File.from_files(path=(blob_ds, 'data/files/images/*'))
file_ds = file_ds.register(workspace=ws, name='new_project')

Explanation

You should execute the following code block:

```
from azureml.core import Dataset
blob_ds = ws.get_default_datastore()
file_ds = Dataset.File.from_files(path=(blob_ds, 'data/files/images/*.png'))
file_ds = file_ds.register(workspace=ws, name='new_project')
```

This block creates a dataset based on the required path and uses the wildcard (*) symbol to indicate all files that end with the Portable Network Graphics (PNG) extension.

You should not execute the following code block:

```
from azureml.core import Dataset
blob_ds = ws.get_default_datastore()
file_ds = Dataset.File.from_files(path=(blob_ds, 'data/files/images/1.png'))
file_ds = file_ds.register(workspace=ws, name='new_project')
```

This block creates a dataset based on the required path. However, only a single file, 1.png, is included in the dataset.

You should not execute the following code block:

```
from azureml.core import Dataset
blob_ds = ws.get_default_datastore()
file_ds = Dataset.File.from_files(path=(blob_ds, 'data/files/images/*'))
file_ds = file_ds.register(workspace=ws, name='new_project')
```

This block creates a dataset based on the required path. However, the wildcard (*) symbol is used to indicate all files - regardless of file type - are included in the dataset.

You should not execute the following code block:

```
from azureml.core import Dataset
blob_ds = ws.get_default_datastore()
file_ds = Dataset.File.from_files(path=(blob_ds, 'data/files/images/*.png'))
file_ds = file_ds.register(workspace=ws, name='new_project_one')
```

This block creates a dataset based on the required path. However, the workspace name is set to new_project_one, which is incorrect.

References

Create Azure Machine Learning datasets to access data

Introduction to datasets

Question 89

You are asked to create a tabular dataset from a storage account data store in which you have weather files stored for 2018 and 2019.

You want to use python to automate creating the tabular dataset.

You have the following requirements for the dataset:

- Load your workspace using the config JSON file.
- Use a comma separated file for the dataset file.
- Create a dataset for files 11.csv and 12.csv in 2018 and all 2019 csv files.

You need to complete the code based on the requirements.

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

Choose the correct options

```
from azureml.core import Workspace, Datastore, Dataset
datastore_name = 'datastoreName'

workspace = Workspace.from_config()

datastore = Datastore.get(workspace, datastore_name)

datastore_paths = [(datastore, '                    '),
                   (datastore, '                    '),
                   (datastore, '                    ')]

weather_ds = Dataset.Tabular.from_delimited_files(path=datastore_paths)
```

Select your answ er

Select your answ er

Select your answ er

Explanation

You should complete the code as follows:

```
from azureml.core import Workspace, Datastore, Dataset
datastore_name = 'datastoreName'
workspace = Workspace.from_config()
datastore = Datastore.get(workspace, datastore_name)
datastore_paths = [(datastore, 'weather/2018/11.csv'),
                   (datastore, 'weather/2018/12.csv'),
                   (datastore, 'weather/2019/*.csv')]
weather_ds =
Dataset.Tabular.from_delimited_files(path=datastore_paths
)
```

You have to specify the dataset weather/2018/11.csv and weather/2018/12.csv to ensure that only those files from 2018 are selected. For 2019, you want all files to be selected. You will have to select weather/2019/*.csv for 2019.

References

[Create Azure Machine Learning datasets]

Question 90

You are working on an Azure Machine Learning experiment that will perform image classification.

You need to conserve resources by terminating the lowest 25 percent performing runs.

What should you do?

Choose the correct answer

- Create a BanditPolicy object.
- Configure Bayesian sampling.
- Configure truncation selection.
- Define a median stopping policy.

Explanation

You should configure truncation selection. A truncation selection policy is an early termination policy that cancels a percentage of low-performing runs. Each run's performance is evaluated using a primary metric, and the percentage of jobs to cancel is specified when you define the policy. As part of its calculations, a truncation selection policy considers previous performance for runs with less run time.

You should not define a median stopping policy. A median stopping policy calculates running averages across training runs that cancels any runs whose performance falls below the median of the running average.

You should not configure Bayesian sampling. This sampling method selects hyperparameters based on the performance of previous runs. The Bayesian sampling method does not support early termination.

You should not create a BanditPolicy object. A BanditPolicy object allows you to create an early termination policy which will terminate training jobs that are not likely to result in an accurate machine learning model. As part of a

BanditPolicy configuration, you can define the amount of slack between the best performing job and the job being evaluated.

References

Tune hyperparameters for your model with Azure Machine Learning

BanditPolicy class

Question 91

You create a script to classify images to build a deep learning neural network. You use scikit-learn in your script file to train the model and elastic cloud compute resources to leverage Azure Machine Learning to scale out the open-source training jobs.

You need to select an estimator library to run the script. You do not want additional Python libraries to be installed in the environment for the estimator.

Which estimator should you use?

Choose the correct answer

- SKLearn
- Chainer
- Tensorflow
- PyTorch

Explanation

You should use the SKLearn estimator. SKLearn estimators are well suited for deep learning image classification models. Since you do not want to include additional libraries, while keeping the time to train the model to a minimum, you should use SKLearn as the estimator when using scikit-learn.

You should not use Tensorflow, Chainer, or PyTorch. Although Azure provides the capability to use any estimator framework, Tensorflow, Keras, and PyTorch would require associated learn module libraries to be included.

References

Build scikit-learn models at scale with Azure Machine Learning
Build a TensorFlow deep learning model at scale with Azure Machine Learning
Train and register a Keras classification model with Azure Machine Learning

Train Pytorch deep learning models at scale with Azure Machine Learning

Question 92

You use the Azure Machine Learning SDK to create a machine learning experiment named experiment_1. You load data and train a model.

You need to view the training results in Azure Machine Learning Studio.

What should you do?

Choose the correct answer

- Call the experiment.start_logging method.
- Call the complete method from the Run class. Set the status to True.
- Call the run.upload_file method. Specify the model name and path.
- Call the experiment_1 variable in Python.

Explanation

You should call the experiment_1 variable in Python. You can use the Azure Machine Learning SDK to create and manage machine learning workspaces, create and manage experiments, load data, and train machine learning models. If you want to use Azure Machine Learning Studio to view associated experiment graphs or individual runs, you can call the experiment variable to generate an Azure Machine Learning Studio link. Following this link takes you to the Azure Machine Learning studio page for your experiment.

You should not call the complete method from the Run class and set the status to True. The run.complete method is called after each iteration of an experiment. As the name indicates, this method marks the run as completed.

You should not call the run.upload_file method and specify the model name and path. This method is used to serialize and upload the model to Azure Machine Learning Studio. You can then use Azure Machine Learning SDK to download the run at a future date.

You should not call the experiment.start_logging method. This method is used to initiate an interactive logging session for an experiment.

References

Tutorial: Train your first ML model

Run class

Experiment class

Question 93

You identify that the storage account used by your Azure Machine Learning (ML) workspace is compromised. In order to prevent any further data breach, you regenerate your storage account keys.

Workspace users report problems using the workspace immediately after you change the storage account keys.

You need to update the storage keys in the workspace.

What CLI command should you run?

Choose the correct answer

- az ml workspace share
- az ml workspace sync-keys
- az ml workspace create
- az ml workspace update

Explanation

You should run the az ml workspace sync-keys command. Right after you change the keys of a storage account that is associated to an Azure ML workspace, it could take up to an hour before the workspace is updated with the new keys. You can use this CLI command to force the update and get the workspace operational.

You should not run the az ml workspace share command. This command lets you share the workspace with users if you need to provide access to the workspace. It will not fix the problem with the storage keys not being updated in the workspace.

You should not run the az ml workspace update command. This command lets you modify the workspace attributes. It will not fix the problem with the storage keys not being updated in the workspace.

You should not run the az ml workspace create command. This command lets you create a new workspace. You do not need to create a new workspace for the users, instead just refresh the storage access keys.

References

Create a workspace for Azure Machine Learning with Azure CLI

Question 94

You create an Azure Machine Learning (ML) workspace. You want this workspace to use the Azure Machine Learning designer to create models and deploy endpoints for others to use. For network security reasons, you want to associate a private endpoint with the workspace and ensure that the private endpoint gets approved on creation.

You decide to use Azure ML Python SDK to create the workspace. You define the following variables:

- subscriptionId: stores the subscription id in which the workspace is being created.
- resourceGroup: stores the name of the resource group to be created.
- workspaceName: the name of the new workspace being created.
- privateEndpointConfig: stores a reference to the private endpoint configuration.

You need to write the code to create the workspace. Workspace class has already been imported.

Which code should you execute?

Choose the correct answer

```
Workspace.create(name=workspaceName,
                 subscription_id=subscriptionId,
                 resource_group = resourceGroup,
                 private_endpoint_config=privateEndpointConfig,
                 private_endpoint_auto_approval=True,
                 sku='basic')
```

```
Workspace.create(name=workspaceName,
                 subscription_id=subscriptionId,
                 resource_group = resourceGroup,
                 private_endpoint_config=privateEndpointConfig,
                 sku='basic')
```

- [x] Workspace.create(name=workspaceName,
    subscription_id=subscriptionId,
    resource_group = resourceGroup,
    private_endpoint_config=privateEndpointConfig,
    sku='enterprise')

- [ ] Workspace.create(name=workspaceName,
    subscription_id=subscriptionId,
    resource_group = resourceGroup,
    private_endpoint_auto_approval=True,
    sku='enterprise')

Explanation

You should execute:

Workspace.create(name=workspaceName,
subscription_id=subscriptionId,
resource_group = resourceGroup,
private_endpoint_config=privateEndpointConfig,
sku='enterprise')

The workspace create method takes the private_endpoint_config parameter to pass the configuration details of the private endpoint. The private_endpoint_auto_approval is an optional parameter with a default value of True. Not specifying this parameter will also ensure that the private endpoint is auto approved. You should also set the sku value to enterprise. The Azure Machine Learning designer is only available with the enterprise edition of the workspace.

You should not execute:

Workspace.create(name=workspaceName,
subscription_id=subscriptionId,
resource_group = resourceGroup,
private_endpoint_config=privateEndpointConfig,

private_endpoint_auto_approval=True,
sku='basic')

The create method sets the sku to basic. The basic edition of the workspace does not support the creation of models using the Azure ML designer.

You should not execute:

Workspace.create(name=workspaceName,
subscription_id=subscriptionId,
resource_group = resourceGroup,
private_endpoint_auto_approval=True,
sku='enterprise')

You need to create a private endpoint. You have to specify the private_endpoint_config parameter with the appropriate configuration to create the private endpoint.

You should not execute:

Workspace.create(name=workspaceName,
subscription_id=subscriptionId,
resource_group = resourceGroup,
private_endpoint_config=privateEndpointConfig,
sku='basic')

The create method sets the sku to basic. The basic edition of the workspace does not support creation of models using the Azure ML designer.

References

[Workspace class]()

Question 95

You use the Execute Python Script module in Azure Machine Learning designer to evaluate and score machine learning models. The first lines of your script are listed below:

```
def azureml(dataframe1 = None):

   return metrics,
```

You need to ensure that the completed function will not return errors.

Which action should you take?

Choose the correct answer

- Add a second input dataframe argument.
- Ensure that the function returns a Python dictionary.
- Ensure that dataframe1 contains a valid dataset.
- Change the entry point function name to azureml_main.

Explanation

You should change the entry point function name to azureml_main. The Execute Python Script 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. The Execute Python Script module requires that your script uses an entry-point function named azureml_main.

You should not add a second input dataframe argument. The Execute Python Script module supports two optional dataset inputs. If these inputs are not provided, data can be generated using Python and imported into the module.

You should not ensure that dataframe1 contains a valid dataset. The Execute Python Script module supports two optional dataset inputs. If these inputs are not provided, data can be generated using Python and imported into the module.

You should not ensure that the function returns a Python dictionary. The azureml_main function returns Python sequence consisting of a single Pandas dataframe. Dictionaries are unordered, indexed data collections comparable to arrays in other programming languages. They contain elements known as key-value (key:value) pairs.

References

Execute Python Script

Create Python Model module

Python Dictionaries

Question 95

You use Azure Machine Learning to create and publish a batch inference pipeline.

Your pipeline is used by automated processes.

You need to ensure that these processes can authenticate to your service automatically.

What should you do?

Choose the correct answer

- Instantiate an AbstractAuthentication class object. Specify the name of the cloud where your pipeline resides.
- Supply workspace, subscription, and resource group information to the Workspace constructor.
- Use the from_config Workspace method to initiate the authentication flow.
- Create a Service Principal (SP) and grant it access to your workspace. Configure SP authentication.

Explanation

You should create a SP, grant it access to your workspace, and configure SP authentication. SP authentication can be used in scenarios when prompting a user for credentials is not feasible, such as when automated processes need to interact with a published pipeline. A service principal is any directory object that can be used for authentication. User accounts are considered service principals.

You should not supply workspace, subscription, and resource group information to the Workspace constructor. A Workspace object is used to define and configure a workspace. In this scenario, providing the listed information to the Workspace constructor will trigger user-interface (UI) based authentication. UI-based authentication is also known as interactive authentication.

You should not instantiate an AbstractAuthentication class object and specify the name of the cloud where your pipeline resides. An AbstractAuthentication object can be used to facilitate authentication as well as acquire authentication tokens.

You should not use the from_config Workspace method to initiate the authentication flow. The from_config method is used to load a workspace from a configuration file. When called, this method will trigger the UI-based authentication flow.

References

Set up authentication for Azure Machine Learning resources and workflows

ServicePrincipalAuthentication class

AbstractAuthentication class

Workspace class

Question 97

You have a dataset on which you want to run a classification algorithm. Analyzing the dataset you find out that one class of items has a much smaller number of records.

You need to select appropriate sampling strategy to account for the class imbalance.

Which data transformation module should you select?

Choose the correct answer

- Synthetic Minority Oversampling Technique (SMOTE)
- Partition and Sample
- Split Data
- Normalize Data

Explanation

You should select the SMOTE module. SMOTE is a statistical technique used to append number of cases to your dataset and ensure balance for the new samples created. This technique can be used to remove imbalance of a certain class in your dataset.

You should not select Normalize Data. Normalize Data module is used to normalize columns within the dataset to eliminate bias that might be caused by large differences in the unit of values present in a column. You can normalize column values in such cases to fall between 0 and 1.

You should not use Partition and Sample module. This module helps create partition of the source dataset and maintain the same ratio of values. This is useful to reduce the size of the dataset and not eliminate imbalance of the source data.

You should not use Split Data module. The Split Data module is used to divide the dataset into two distinct sets based on the splitting mode provided. This is useful when training models by creating a training set and a testing set.

References

SMOTE

Normalize Data module

Partition and Sample module

Split Data module

Question 98

You manually define security rules to isolate your Azure Machine Learning training processes within an Azure Virtual Network.

You need to ensure that only required communications from the Internet are allowed.

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

Choose the correct answers

- ☐ Define an inbound security rule.
- ☐ Ensure the source service tag is set to AzureMachineLearning.
- ☐ Create an outbound security rule.
- ☐ Ensure the source service tag is set to BatchNodeManagement.
- ☐ Configure a destination port range of 22.
- ☐ Configure a source port range of 29876-29877.

Explanation

You should define an inbound security rule. Azure Virtual Networks serve as security boundaries between your Azure resources and the Internet. You can control inbound and outbound access to or from an Azure Virtual Network by defining security rules that act like network firewall rules. To ensure that only required communications from the Internet are allowed, you should define an inbound security rule.

You should ensure the source service tag is set to BatchNodeManagement. Security rules include an option to identify groups of Azure cloud resource nodes (IP addresses) using service tags. This allows you to define security rules without having to manually identify the IP addresses for Azure resources that you want to allow to access your virtual network. As Azure Machine Learning relies on other Azure resources and services, you must ensure that a security rule is created using the BatchNodeManagement service tag.

You should configure a source port range of 29876-29877. The BatchNodeMnagement service tag requires a destination port range of 29876 to 29877. This allows core Azure resources required for machine learning processes to communicate with resources in your Azure Virtual Network.

You should not ensure the source service tag is set to AzureMachineLearning. The AzureMachineLearning service tag is only required if you use Azure-based compute targets. If you use your own compute targets, this service tag is not required.

You should not configure a destination port range of 22. Port 22 is used by the Secure Shell (SSH) protocol in order to facilitate remote, command-line management of resources. Creating an inbound security rule to allow SSH communications is optional.

You should not create an outbound security rule. Outbound security rules control resource access initiated from resources within your Azure Virtual Network. In this scenario, you are trying to control inbound access.

References

[Network isolation during training & inference with private virtual networks]

Question 99

You use an Azure Machine Learning workspace to train machine learning models using various compute targets.

You need to create a remote virtual machine (VM) compute target that can be used for Azure Machine Learning training.

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

Explanation

You should complete the following actions in order:

1. Provision a VM that runs Ubuntu.
2. Use the ComputeTarget.attach method to link the VM to a workspace.
3. Create a run configuration for the remote VM compute target.

Azure Machine Learning supports using any remotely accessible Ubuntu-based VM as a compute resource. This resource can be located within Azure, on a third-party cloud provider, or within your own organization. As long as the VM has a publicly accessible IP address, you can attach the resource to an Azure Machine Learning workspace using the VM's address and authentication credentials.

Once the remote VM is set up, you can link the VM to a workspace by using the ComputeTarget.attach method in a Python script. This method takes as its input an attach configuration that consists of a resource ID, Secure Shell (SSH) port, and authentication credentials.

A run configuration contains the configuration needed to submit a training run on a remote VM. Among other details, a run configuration can define any dependencies required for the machine learning process. Run configurations greatly reduce configuration overhead by allowing you to define environment parameters that can be reused across compute targets.

You should not provision a VM that runs Windows 10 Enterprise. Azure Machine Learning does not support VMs running Windows 10. Only Ubuntu-based VMs are supported.

You should not use the register_azure_blob_container method to create and link a datastore to the workspace. Datastores act as links to cloud-based data sources and are used by Azure Machine Learning to facilitate access to data that will be ingested by machine learning experiments as well as providing storage for machine learning output. Azure Machine Learning supports several types of data sources, including Azure SQL databases and Azure blob containers.

References

Set up and use compute targets for model training

Quickstart: Set up the Data Science Virtual Machine for Linux (Ubuntu)

Connect to Azure storage services

Question 100

You finish training a deep learning multi-class image classification model.

You execute the following code:

```
from azureml.core import Run
run = Run.get_context()
```

You need to register the model using Azure Machine Learning (ML) SDK.

Which script should you use?

Choose the correct answer

```
model_file = 'infection_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/infection_model.pkl',
model_name='infection_model',
                tags={'Training context':'Inline Training'},
                properties={'AUC': run.get_metrics()['AUC'], 'Accuracy':
run.get_metrics()['Accuracy']})
```

```
run.register_model(model_name='infection_model',
                tags={'Training context':'Inline Training'},
                properties={'AUC': run.get_metrics()['AUC'], 'Accuracy':
run.get_metrics()['Accuracy']})
```

```
model_file = 'infection_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_name='infection_model',
```

tags={'Training context':'Inline Training'},
properties={'AUC': run.get_metrics()['AUC'], 'Accuracy': run.get_metrics()['Accuracy']})

- model_file = 'infection_model.pkl'
joblib.dump(value=model, filename=model_file)
run.complete()

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

Explanation

You should use the following script to register the model:

```
model_file = 'infection_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/infection_model.pk
l', model_name='infection_model',
                   tags={'Training context':'Inline
Training'},
                   properties={'AUC':
run.get_metrics()['AUC'], 'Accuracy':
run.get_metrics()['Accuracy']})
```

You must save the file for the model you want to register to a .pkl file. You need to specify the .pkl file in the register_model method.

You should not use:

```
run.register_model(model_name='infection_model',
```

```
                   tags={'Training context':'Inline
Training'},
                   properties={'AUC':
run.get_metrics()['AUC'], 'Accuracy':
run.get_metrics()['Accuracy']})
```

The register_model requires the model file to be passed as a parameter to register the model.

You should not use:

```
model_file = 'infection_model.pkl'
joblib.dump(value=model, filename=model_file)
run.complete()

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

Once you use the joblib.dump to extract the model, you must save the model to a file that is referred by the register_model method. You need to call the upload_file method.

You should not use:

```
model_file = 'infection_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_name='infection_model',
                   tags={'Training context':'Inline
Training'},
```

```
                     properties={'AUC':
run.get_metrics()['AUC'], 'Accuracy':
run.get_metrics()['Accuracy']})
```

This script extracts the model and saves it to the infection_model.pkl file. However, the path to the file is not specified in the register_model method. You need to specify the model_path parameter in the register_model method.

References

[Model class]

[Build and Deploy a Machine Learning Model with Azure ML Service]

Question 101

You use Azure Machine Learning to create machine learning models. You plan to deploy your models as web services using various compute targets.

You need to ensure that each deployment is configured to require authentication.

Based on compute target, which authentication method should you configure? To answer, select the appropriate compute targets from the drop-down menus.

Choose the correct options

Compute Target	Authentication Methods
Azure Container Instances (ACI)	
Azure Kubernetes Service (AKS)	

Explanation

You should configure key authentication only for ACI compute targets. ACI allows you to package and deploy your machine learning models using easy-to-manage containers. A container is a virtualized app that includes all the resources it needs to run, including file resources, dependencies, and services. ACI only supports key authentication. This means that if authentication is enabled for a model's web service endpoints, all requests must be authenticated using a predefined key. This is similar to many REST Application Programming Interfaces (APIs) which support API keys for authentication.

You should configure key or token authentication for AKS compute targets. AKS supports highly scalable compute options for Azure Machine Learning experiments. In addition to supporting multiple-node clusters, AKS can be used for experiments that require hardware acceleration via GPU or Field-Programmable Gate Arrays (FPGA). Finally, AKS can dynamically scale compute availability based on workload.

Token-based authentication relies on temporary tokens. Once enabled, users or services that connect to your deployed model must submit an Azure Machine Learning JSON Web Token in order to be allowed access. Each token has a limited lifetime, and expired tokens must be refreshed prior to making new calls.

References

Set up authentication for Azure Machine Learning resources and workflows

What is Azure Container Instances?

Azure Kubernetes Service (AKS)

Question 102

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

Once the pipeline is complete, you need to ensure that other users can run the pipeline using a custom input.

Which four 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

Explanation

You should perform the following actions in order:

1. Define a PipelineParameter object.
2. Specify a default pipeline parameter value.
3. Add the PipelineParameter object as a script argument.
4. Use the publish_pipeline method to publish the pipeline.

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.

You can allow other users to run the pipeline using custom inputs by publishing the pipeline. To do this, you must first create a pipeline parameter object using the PipelineParameter class, which allows you to specify the default value of the pipeline parameter. Once this is complete, you can add your PipelineParameter to any step, including script steps, when constructing a pipeline. Finally, you can use the publish_pipeline method to publish the pipeline. This creates a REST endpoint that can by accessed by external users.

You should not save the pipeline to a YAML file. You use this option to export the pipeline steps and then import them into another system.

References

PipelineParameter class

PipelineRun class

Create and run machine learning pipelines with Azure Machine Learning SDK

Question 103

You run an Automated Machine Learning (ML) experiment using the following code:

```
from azureml.core.experiment import Experiment
from azureml.widgets import RunDetails

print('Submitting Auto ML experiment...')
automl_experiment = Experiment(ws, 'infections_automl')
automl_run = automl_experiment.submit(automl_config)
RunDetails(automl_run).show()
automl_run.wait_for_completion(show_output=True)
```

You need to display the best performing model when the run succeeds.

Which Python code should you add?

Choose the correct answer

- best_run = automl_run.get_runs('infections_automl')
print(best_run)

- best_run, fitted_model = automl_run.get_output()
print(best_run)

- best_run = automl_run.summary()
print(best_run)

- best_run = automl_run.get_metrics()
print(best_run)

Explanation

You should add the following lines to your code to display the best model:

```
best_run, fitted_model = automl_run.get_output()
print(best_run)
```

The get_output method returns the best pipeline based on the primary metric when no parameters are specified.

You should not add:

```
best_run = automl_run.summary()
print(best_run)
```

The summary method gets a table containing a summary of all algorithms attempted, as well as their scores.

You should not add:

```
best_run = automl_run.get_runs('infections_automl')
print(best_run)
```

The get_runs method allows the experiment to retrieve a reference to a given execution. This code will fail when executed with the AutoMLRun class

You should not add:

```
best_run = automl_run.get_metrics()
print(best_run)
```

The get_metrices method returns the metrics logged against the run. It does not provide details on the best run from an experiment.

References

AutoMLRun class

Question 104

You work for an IT services company that provides consulting services to various clients. Your client has an Azure subscription that was configured with the following in a resource group named mlResources:

Azure Storage Account: mlStorage
Azure Application Insights: mlAppInsights
Azure Key Vault: mlKeyVault
Azure Container Registry: mlACR

You are asked to create a script that leverages these resources to create an Azure Machine Learning (ML) workspace.

You need to complete the code to create the workspace named mlWorkspace.

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

Choose the correct options

```
workspaceName = "mlWorkspace"

storageId = az storage account show --name "mlStorage" --query "id"

appInsightsId = az monitor app-insights component show \

                       --app "mlAppInsights" \

                       -g "mlResources" --query "id"

keyVaultId = az keyvault show --name "mlKeyVault" --query "id"

acrId = az acr show --name "mlACR" -g "mlResources" --query "id"

az                                -w

                      -g $mlResources \

                      --container-registry $acrId \

                      --storage-account $storageId \

                      --application-insights $appInsightsId \

                      --keyvault $keyVaultId
```

Explanation

You should complete the code as follows:

```
workspaceName = "mlWorkspace"


storageId = az storage account show --name "mlStorage" --
query "id"


appInsightsId = az monitor app-insights component show \
                       --app "mlAppInsights" \
                       -g "mlResources" --query "id"


keyVaultId = az keyvault show --name "mlKeyVault" --query
"id"


acrId = az acr show --name "mlACR" -g "mlResources" --
query "id"


az ml workspace create -w $workspaceName -g $mlResources
\
                     --container-registry $acrId \
--storage-account $storageId \
--application-insights $appInsightsId \
--keyvault $keyVaultId
```

You should select ml workspace. This allows you to create a Machine Learning workspace.

You should not select ml environment. ml environment provides commands that can be used to manage the Machine Learning environment.

You should not select ml service. ml service provides a handle to an operationalized service.

You should not select workspace. workspace would provide commands that can manage a Log Analytics workspace and not a Machine Learning workspace.

You should select create. This allows you to create a new workspace.

You should not select list. Selecting list will provide a list of available workspaces.

You should not select update. Selecting update lets you update an existing workspace.

You should select workspaceName. workspaceName is the variable that is set equal to the name you want to give to the workspace.

You should not select mlWorkspace. mlWorkspace is the name you would like to give your workspace. However, mlWorkspace is not a parameter defined in the Azure ML code.

References

[Create a workspace for Azure Machine Learning with Azure CLI]

Question 105

You use Azure Machine Learning designer to create an inference pipeline and train a predictive model.

You need to deploy your pipeline as a web service that can autoscale based on workload.

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

Explanation

You should perform the following actions in order:

1. Convert the training pipeline into a real-time inference pipeline.
2. Create an AKS cluster.
3. Deploy a real-time endpoint.

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. Once your pipeline has trained a model, you convert the training pipeline into a real-time inference pipeline. This adds the supporting Web Service Input and Web Service Output modules to your pipeline.

An 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, and publishing an autoscaling inference pipeline is only supported on AKS inference clusters. If you have not created an AKS cluster prior to this step, you will need to define one before you can complete the deployment. Once the real-time endpoint has been deployed, applications and services can access the endpoint as they would any other REST API.

You should not set the pipeline as the default for the endpoint. 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

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

Run batch predictions using Azure Machine Learning designer (preview)

Deploy models with Azure Machine Learning

Question 106

Your organization has Azure Machine Learning workspace configured. Deployment fails when you deploy a real-time web service for your Azure Machine Learning model using the following code:

```
from azureml.core.webservice import AciWebservice

aci_config =
AciWebservice.deploy_configuration(cpu_cores=1,
memory_gb=1)
aci_service = Model.deploy(workspace=ws,
                       name='global-warming-service',
                       models=[model],
                       inference_config=inference_config,
                       deployment_config=aci_config)
aci_service.wait_for_deployment(show_output=True)
```

You need to determine the cause for deployment failure by analyzing actions performed during deployment.

Which line of code should you execute?

Choose the correct answer

- aci_service.get_token()
- aci_service.serialize()
- aci_service.get_logs()
- aci_service.get_keys()

Explanation

You should use the get_logs() method. The get_logs() method allows you to print out detailed Docker engine logs. You can view logs for Azure Container Instance (ACI), Azure Kubernetes Service (AKS), and local deployments.

You should not use the get_keys() method. The get_keys() method retrieves the auth keys for the web service, but not the logs that show errors associated with the deployment.

You should not use the get_token() method. The get_token() method retrieves the auth token for the web service, but not the logs that show errors associated with the deployment.

You should not use the serialize() method. The serialize() method converts the web service object into a JSON serialized dictionary object. It does not show logs associated with the deployment.

References

Webservice class

Troubleshoot Docker deployment of models with Azure Kubernetes Service and Azure Container Instances

Question 107

Your Azure Machine Learning Python script includes the following code block:

```
my_explainer = TabularExplainer(
   my_model,
   x_train,
   features=my_data.feature_names,
   classes=classes
   )
```

You need to identify the processes that will be triggered when this code is executed.

Which two statements describe what this code block does? Each correct answer presents part of the solution.

Choose the correct answers

- ☐ An explainer will be loaded based on a model type.
- ☐ A tabular dataset will be processed.
- ☐ Underperforming models will be canceled.
- ☐ Feature importance data will be generated.

Explanation

With the mentioned code an explainer will be loaded based on a model type and a tabular dataset will be processed. Explainers, also known as interpretability techniques, are used to interpret or explain machine learning models. These explanations are used by data scientists to understand how a machine learning model works. For example, if a model is used to predict which type of person is inclined to commit a crime, its users may want to understand how the model makes that prediction. TabularExplainer is used with tabular datasets.

TabularExplainer is categorized as a meta explainer, which means that it chooses an explainer based on how the referenced model is structured. For example, TabularExplainer will use LinearExplainer when a linear model is being evaluated.

Feature importance data will not be generated. 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 prediction, you should use an interpretability technique that calculates and tracks feature importance.

Underperforming models will not be canceled. To conserve resources and reduce training time by terminating underperforming models, you should define an early termination policy. Early termination is part of Azure Machine Learning's automated hyperparameter tuning process.

References

Model interpretability in Azure Machine Learning

Automated and Interpretable Machine Learning

Tune hyperparameters for your model with Azure Machine Learning

Question 108

You are creating an Automated Machine Learning (ML) experiment to evaluate a regression model.

You plan to export your Auto ML generated model to an Open Neural Network Exchange (ONNX) model.

You need to choose algorithms that can be used by Auto ML models and support exporting to ONNX models.

Which two algorithms can you use to meet your goal? Each correct answer presents a complete solution.

Choose the correct answers

- [ ] Random Forest
- [ ] Linear SVC
- [ ] Auto-ARIMA
- [ ] Decision Tree

Explanation

You can use Decision Tree or Random Forest. Both algorithms can be used to evaluate regression problems. They can also be exported to ONNX models.

You should not use the Auto-ARIMA algorithm. Auto-ARIMA is only suited for time series forecasting models. In addition, Auto-ARIMA cannot be exported to ONNX models.

You should not use the Linear SVC algorithm. Linear SVC based classification models can be exported to ONNX models, but they are only suited for data classification problems.

References

ONNX and Azure Machine Learning: Create and accelerate ML models

Configure automated ML experiments in Python

Question 109

You use Azure Machine Learning to deploy machine learning models.

You need to be notified when a model deployment fails.

What should you do?

Choose the correct answer

- Stream Azure Machine Learning logs to Azure Monitor.
- Stream Azure Machine Learning metric information to Azure Event Hub.
- Query AmlComputeJobEvents using the ExecutionState property.
- Query AmlComputeJobEvents using the ProvisionState property.

Explanation

You should stream Azure Machine Learning logs to Azure Monitor. Azure Monitor is a stand-alone platform that can ingest logging and other information from a variety of Azure resources and services, including Azure Machine Learning. Once you configure Azure Machine Learning to forward logs to Azure Monitor, you can query the logs using SQL-like language. You can also create alerts based on Azure Machine Learning metrics to track occurrences such as compute nodes going offline or a model deployment failing.

You should not query AmlComputeJobEvents using the ExecutionState property. The AmlComputeJobEvent log is just one of the logs that can be streamed to Azure Monitor. Information from this log is stored in the Azure Monitor AmlComputeJobEvents table. You can query the ExecutionState property to determine the state of a job. While you can search for failed jobs using this method, you will not be notified when a model deployment fails.

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 do not create alerts in Azure Event Hub.

You should not query AmlComputeJobEvents using the ProvisionState property. The ProvisionState property, which is stored in the AmlComputeJobEvents table, records information about the state of a job submission, not the state of the job itself.

References

Monitoring Azure Machine Learning

Azure Monitor overview

Azure machine learning monitoring data reference

Azure Event Hubs — A big data streaming platform and event ingestion service

Question 110

You use Azure Machine Learning SDK to train machine learning models.

You need to understand how a local machine learning model makes its predictions.

What should you do?

Choose the correct answer

- Create and register a new file dataset. Use the from_files method to specify the bookmarked data.
- Install the azureml-interpret Python package and create an explainer.
- Create a rule and set the source service tag to BatchNodeManagement.
- Use the az ml folder attach command to create a run configuration.

Explanation

You should install the azureml-interpret Python package and create an explainer. Explainers, also known as interpretability techniques, are used to interpret or explain machine learning models. These explanations are used by data scientists to understand how a machine learning model works. For example, if a model is used to predict which type of person is inclined to commit a crime, its users may want to understand how the model makes that prediction. In order to create an explainer for your local machine learning model, you need to install the azureml-interpret Python package.

You should not create a rule and set the source service tag to BatchNodeManagement. Security rules allow you to isolate and protect your experiments by controlling access to Azure Machine Learning resources. If you are going to implement security rules, you must ensure that a security rule is created using the BatchNodeManagement service tag. This allows Azure Machine Learning to interact with other Azure services.

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.

You should not use the az ml folder attach command to create a run configuration. Azure Machine Learning provides the capability to run experiments on different compute targets without requiring scripts to be rewritten. This is done by creating a run configuration, which serves as a template for a training environment. The easiest way to generate a run configuration is to use the az ml folder attach command.

References

Model interpretability in Azure Machine Learning

Network isolation during training & inference with private virtual networks

Dataset class

az ml folder

Question 111

You use Azure CLI to create an Azure Machine Learning compute cluster.

You need to ensure that costs are not incurred when jobs are not running.

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

Choose the correct answers

- ☐ Create a data factory compute target.
- ☐ Specify an idle seconds scale down.
- ☐ Create an Azure Machine Learning compute target.
- ☐ Specify the minimum number of cluster nodes. Set the minimum number of cluster nodes to 0.
- ☐ Create an Azure Machine Learning compute instance target.

Explanation

You should create an Azure Machine Learning compute target and specify the minimum number of cluster nodes. An Azure Machine Learning compute target is a computing resource where machine learning experiments can be run. Azure supports a variety of compute target types, including your local computer, a remote virtual machine (VM), and Azure Machine Learning compute clusters. You can create an Azure Machine Learning compute target cluster using the az ml computetarget create amlcompute command.

Compute clusters are highly scalable targets that consist of one or more compute nodes, and a cluster can scale up or down dynamically based on workload. You can control the maximum number of nodes in the cluster by using the required --max-nodes parameter. By specifying the minimum number of nodes as 0, you can ensure that all active nodes will be terminated when jobs are not running. This will prevent Azure compute costs from accruing during idle times.

You should not specify an idle seconds scale down. The idle seconds parameter allows you to control how long a compute cluster must be idle before unneeded resources are deprovisioned. This allows you to ensure that intermittent pauses in machine learning jobs do not cause unnecessary waiting time as nodes are provisioned and deprovisioned.

You should not create an Azure Machine Learning compute instance target. A compute instance is a single Azure-based VM used for machine learning experiments. You can use compute instances to support automated machine learning (AutoML) and machine learning pipelines.

You should not create a data factory compute target. An Azure Data Factory (ADF) compute target is used to create machine learning pipelines. ADF facilitates ingestion and batch processing of data in order to provide predictive analytics.

References

Set up and use compute targets for model training

az ml computetarget create

What is an Azure Machine Learning compute instance?

Data ingestion with Azure Data Factory

Question 112

You create a custom role definition file.

You need to ensure a user can submit a training run.

What should you do?

Choose the correct answer

- Add /workspaces/environments/write to the Actions element.
- Remove existing AssignableScopes and add the “/” scope.
- Grant the role to the user with the az ml workspace share command.
- Remove /workspaces/computes/write from the NotActions element.

Explanation

You should add /workspaces/environments/write to the Actions element. If the Azure Machine Learning default roles do not provide the granular permissions you need to control resource access, you can create custom role definition files which are constructed as JavaScript Object Notation (JSON) dictionaries.

The Actions section in a role definition file is used to define the permissions that the role has. Submitting a training run requires a user to be assigned a built-in role, such as contributor, or be assigned a custom role with the required privileges. The /workspaces/environments/write permission will allow a user to submit a training run.

You should not remove existing AssignableScopes and add the “/” scope. The AssignableScopes section of a role definition file can be used to limit where a role’s permissions can be exercised. The “/” scope means all scopes.

You should not grant the role to the user with the az ml workspace share command. A role must first be created before it can be assigned. The az role definition create can be used to create a custom role definition, and takes a JSON role definition file as input.

You should not remove /workspaces/computes/write from the NotActions element. The NotActions element is used to define permissions a role should not be assigned. These are subtracted from permissions defined in the Actions element.

References

Manage access to an Azure Machine Learning workspace

az ml workspace

az role definition

Question 113

You use Azure Machine Learning to train and deploy a machine learning model as a web service. Your web service requires authentication.

You need to test the deployment to ensure that requests to the web service will be successful.

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

Explanation

You should perform the following actions in order:

1. Retrieve the scoring_uri property of a Webservice object.
2. Specify bearer authentication in the header.
3. Issue an HTTP POST request with JSON data.

When a machine learning model is deployed as a web service, a REST interface is defined, which client applications can use to consume the service. You determine the service endpoint by using the Azure Machine Learning SDK to retrieve the scoring_uri property of a Webservice object.

As the published model requires authentication, you will need to define an HTTP request header in your code that specifies bearer authentication. Once your code is complete, you can issue an HTTP POST request to the published service endpoint you discovered using the scoring_uri property.

You should not query AmlComputeJobEvents using the ExecutionState property. The AmlComputeJobEvent log is just one of the logs that can be streamed to Azure Monitor. Information from this log is stored in the

Azure Monitor AmlComputeJobEvents table. You can query the ExecutionState property to determine the state of a job.

References

[Consuming a real-time inferencing service]

[Monitoring Azure Machine Learning]

Question 114

You use Azure Machine Learning to tune hyperparameters for your model.

You need to ensure that underperforming runs are terminated automatically.

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

Choose the correct answers

- ☐ Create a RandomParameterSampling object and specify a parameter_space dictionary.
- ☐ Define a PipelineParameter object and specify a default pipeline parameter value.
- ☐ Use Bayesian sampling to select hyperparameter samples.
- ☐ Create and configure a BanditPolicy object.

Explanation

You should create a RandomParameterSampling object and specify a parameter_space dictionary. The random parameter sampling method allows you to define a parameter search space and Azure Machine Learning will randomly choose hyperparameters from within this space. Random parameter sampling supports early termination.

Hyperparameters are used to control the training process for machine learning models, and these parameters can have a significant impact on how a trained model performs. Azure Machine Learning includes a hyperparameter tuning service which supports three parameter sampling methods, and the method selected determines whether early termination is supported.

You should create and configure a BanditPolicy object. A BanditPolicy object allows you to create an early termination policy which will terminate training jobs that are not likely to result in an accurate machine learning model. As part of a BanditPolicy configuration, you can specify how frequently jobs are evaluated and the amount of slack between the best performing job and the job being evaluated. This can greatly reduce training job runtimes and conserve compute resources.

You should not use Bayesian sampling to select hyperparameter samples. This sampling method selects hyperparameters based on the performance of previous runs. The Bayesian sampling method does not support early termination.

You should not define a PipelineParameter object and specify a default pipeline parameter value. You can allow other users to run the pipeline using custom inputs by publishing the pipeline. To do this, you must first create a pipeline parameter using the PipelineParameter class, which allows to specify the default value of the pipeline parameter.

References

Tune hyperparameters for your model with Azure Machine Learning

BanditPolicy class

PipelineParameter class

Question 115

You have an Azure Machine Learning (ML) workspace. You want to provide a group of data scientists access to the workspace. You create a Data Scientist Custom Role, which you will assign to ML workspace resources.

You have a data scientist with the Azure User Principal Name jdoe@company.com.

You need to assign the Data Scientist Custom Role to this user.

Which three tasks can you perform? Each correct answer presents a complete solution.

Choose the correct answers

☐ Assign the role from the portal using the Access Control (IAM) screen for the workspace.

☐ Execute the az role assignment create CLI command.

☐ Execute the az ml workspace share CLI command.

☐ Assign the role from the portal using the Access Control (IAM) screen for the resource group.

☐ Execute the az ml workspace update CLI command.

Explanation

You can assign a custom role to the Azure ML workspace using the following options:

You should execute the az ml workspace share CLI command. You can install the Azure ML CLI and provide the role along with the users UPN to assign the role to the user. This will allow you to assign a custom role to the Azure ML workspace.

You should assign the role from the portal using the Access Control (IAM) screen for the workspace. You can use the IAM screen for the workspace resource. This will allow you to assign a custom role to the Azure ML workspace.

You should execute the az role assignment create CLI command. You can use Azure CLI to assign a role to the workspace resource.

You should not execute the az ml workspace update CLI command. You cannot assign a role using the Azure ML CLI update command. The update command lets you configure the attributes about the workspace, not about the role.

You should not assign the role from the portal using the Access Control (IAM) screen for the resource group. You cannot assign the role to a resource group since the custom role is scoped for Azure ML workspace resources.

References

Manage access to an Azure Machine Learning workspace

Add or remove Azure role assignments using Azure CLI

Add or remove Azure role assignments using the Azure portal

az ml workspace

Question 116

You are asked to automate the creation of a Azure Machine Learning (ML) workspace using the Azure CLI. You install the Azure CLI runtime and log into Azure.

You select the correct subscription where you want to create the Machine Learning workspace. You want to execute Azure ML CLI commands for the first time.

You need to complete the code to automate workspace creation.

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

Create a list in the correct order

Possible commands

Explanation

You should complete the script using the following commands in sequence:

1. az extension add -n azure-cli-ml
2. az group create --name <resource-group-name> --location <location>
3. az ml workspace create -w <workspace-name> -g <resource-group-name>

The first step should be to register the Machine Learning extension to run the az ml commands.

Then you should create the resource group where the Machine Learning workspace will reside.

Finally, you should use the az ml CLI command to create the workspace in the resource group created in the second step.

You should not run the az ml workspace list command. This command just lists any existing Azure Machine Learning workspaces currently present in the Azure subscription.

You should not run the az ml workspace update command. This command allows updates to an existing workspace and does not create a workspace.

References

[Create a workspace for Azure Machine Learning with Azure CLI]

Question 117

You create an Azure Machine Learning (ML) workspace in a subscription your company owns.

You need to create some sample code for other data scientist to use to connect to your workspace.

Which two code segments can you use to meet your goal? Each correct answer presents a complete solution.

Choose the correct answers

- [ ]
```
from azureml.core import Workspace
ws = Workspace.get(name='my-workspace',
                   subscription_id='1234567-abcde-890-fgh...',
                   resource_group='my-resources')
```

- [ ]
```
from azureml.core import Workspace
ws = Workspace.from_config()
```

- [ ]
```
from azureml.core import Workspace
ws = Workspace.create(name='my-workspace',
                   subscription_id='1234567-abcde-890-fgh...',
                   resource_group='my-resources',
                     create_resource_group=True,
                     location='eastus',
                     sku='enterprise')
```

- [ ]
```
from azureml.core import Workspace
ws = Workspace.write_config(path="./file-path",
file_name="ws_config.json")
```

Explanation

You can use the following code segments to meet your goal:

```
from azureml.core import Workspace
```

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

and

```
from azureml.core import Workspace
ws = Workspace.get(name='my-workspace',
               subscription_id='1234567-abcde-890-fgh...',
               resource_group='my-resources')
```

You can connect to a workspace using the config file that Azure ML already creates on compute creation in the root folder using the from_config() method. You can also explicitly specify the name, subscription_id and resource_group using the get() method of the workspace object.

You should not use the following:

```
from azureml.core import Workspace
ws = Workspace.create(name='my-workspace',
               subscription_id='1234567-abcde-890-fgh...',
               resource_group='my-resources',
                create_resource_group=True,
                location='eastus',
                sku='enterprise')
```

The Workspace.create() method is used to create a new workspace. You cannot connect to an existing workspace.

You should not use the following:

```
from azureml.core import Workspace
ws = Workspace.write_config(path="./file-path",
file_name="ws_config.json")
```

The Workspace.write_config() method is used to create a new config file for a workspace. The method will not connect to an existing workspace using the default config file created.

References

Workspace class

Question 118

You are planning the size of the compute resources required for data provided by your marketing team. Your team will run experiments and create dataframes using pandas.

The marketing team provides a 1 GB CSV file with the data. All processing is required to happen in memory.

You need to recommend the minimum memory (RAM) configuration required to support processing these files.

What should you recommend?

Choose the correct answer

- 2 GB
- 10 GB
- 20 GB
- 8 GB

Explanation

You should recommend 20 GB RAM. The size guidance is based on the fact that a 1 GB CSV data file can become 10 GB in a dataframe. You want to have double that for RAM, which equals to 20 GB RAM.

10 GB, 2 GB, and 8 GB RAM will not support all in-memory operations based on the compute size guidance for frameworks like pandas.

References

[Create Azure Machine Learning datasets]

Question 119

You have access to structured data used in a previous machine learning project. The data is not registered in a dataset.

You need to use Azure Machine Learning SDK to use the data directly in a machine learning script.

What should you do?

Choose the correct answer

- Use the from_files method to create a dataset and register the dataset in your workspace.
- Use the get_context method from the Run class to load the run from the remote environment.
- Use the register method from the Dataset class to create a new dataset version.
- Use Python to create a TabularDataset. Specify the path to the data.

Explanation

You should Use Python to create a TabularDataset and specify the path to the data. TabularDatasets represent collections of structured data, like the ones found in a comma-separated-value (CSV) file. You can create a TabularDataset and reference it directly in a training script without having to register the dataset with an Azure Machine Learning workspace.

You should not use the from_files method to create a dataset and register the dataset in your workspace. You create a file dataset to reference the unstructured 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.

You should no use the get_context method from the Run class to load the run from the remote environment. This method is used to obtain a reference to a service context. This reference can then be used to perform tasks such as uploading files.

You should not use the register method from the Dataset class to create a new dataset version. Azure Machine Learning allows you to register a new dataset using an existing dataset name using versioning.

References

Train with datasets in Azure Machine Learning

Dataset class

Run class

Version and track datasets in experiments

Question 120

You are using the Azure Machine Learning SDK to create experiments that will process sparse data. Your code includes code to create an AutoMLConfig object:

```
automl_config = AutoMLConfig(
  task = 'regression',
  compute_target = my_target,
  training_data = my_data,
  label_column_name = my_label,
  **automl_experiment
  )
```

You need to ensure that your data is not scaled and normalized.

What should you do?

Choose the correct answer

- Add the from azureml.widgets import RunDetails line to your script.
- Define the featurization parameter in the automl_experiment data dictionary.
- Add the primary_metric parameter when creating the AutoMLConfig object.
- Move the task = ‘regression’ line to the automl_experiment data dictionary.

Explanation

You should define the featurization parameter in the automl_experiment data dictionary. When configuring a pipeline using Azure Machine Learning SDK, you can define automated machine learning settings that control how the experiment is run. These settings are typically defined in dictionary format and passed to an AutoMLConfig object. If the featurization parameter is set to auto, input data will automatically be preprocessed, and missing values will be handled. In this case, the featurization parameter should be added to the automl_experiment data diction and configured with a value of off.

You should not add the from azureml.widgets import RunDetails line to your script. The RunDetails class can be used in a Jupyter notebook to view how model training is progressing.

You should not add the primary_metric parameter when creating the AutoMLConfig object. The metrics you can configure are dependent on the machine learning task type, such as regression or classification.

You should not move the task = ‘regression’ line to the automl_experiment data dictionary. Most AutoMLConfig parameters can be included inline with the object’s definition, or they can be defined in a data dictionary which is passed to the AutoMLConfig object at runtime.

References

Configure automated ML experiments in Python

RunDetails class

AutoMLConfig class

Question 121

You use Azure Machine Learning Software Development Kit (SDK) to manage machine learning datastores. You need to add a datastore to your workspace. You run the following code:

```
from azureml.core import Workspace, Datastore
ws = Workspace.from_config()
```

You need to retrieve the workspaceblobstore datastore.

What should you do?

Choose the correct answer

- Use the get_default_datastore method of the Workspace object.
- Construct a Python for loop to enumerate all Workspace datastores.
- Use the Datastore.get method to retrieve the default-datastore datastore.
- Use the Workspace set_default_datastore method to set the default datastore.

Explanation

You should use the get_default_datastore method of your Workspace object. When a new workspace is created, it contains a default datastore: workspaceblobstore. In order to retrieve the default datastore, you can use the get_default_datastore Workspace method. Line #3 in the Python code from the question might read:

my_datastore = ws.get_default_datastore()

You should not use the Datastore.get method to retrieve the default-datastore datastore. This would retrieve a datastore named default-datastore. 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 use the Workspace set_default_datastore method to set the default datastore. This command does not allow you to retrieve a datastore. You can use this command to set a new default datastore.

You should not construct a Python for loop to enumerate all Workspace datastores. You could use this method to list all datastores.

References

[Introduction to datastores]

Question 122

You use Azure Machine Learning designer to create a batch inference pipeline. You plan to publish the pipeline using a web service.

You need to ensure that the pipeline can make predictions on the new data supplied at runtime.

What should you do?

Choose the correct answer

- Connect a different dataset to the pipeline.
- Add the Convert to Dataset module to your pipeline.
- Create a parameter for your dataset.
- Publish the pipeline to a new endpoint.

Explanation

You should create a parameter for your dataset. This option allows consumers to provide a dataset to your pipeline at runtime. This is useful in scenarios where a model is trained on a dataset but is used to formulate predictions on new data. You can parameterize a pipeline by using the dataset module.

You should not connect a different dataset to the pipeline. You connect a dataset to a pipeline when you need to provide data input. This process is manual and does not occur automatically at runtime.

You should not publish the pipeline to a new endpoint. A web service is defined when you publish a pipeline, and an HTTP endpoint is created that external applications and services can consume.

You should not add the Convert to Dataset module to your pipeline. The Convert to Dataset module is used to ensure that data normalization changes can be used in other pipelines. Input data to this module must be tabular.

References

Run batch predictions using Azure Machine Learning designer (preview)

Convert to Dataset

Question 122

You use Azure Machine Learning to build models that predict credit risk for a bank's customers. Your dataset includes fields for age, salary, job title, and years of education.

You need to implement a model-agnostic method of identifying how each field impacts a model's predictions.

What should you do?

Choose the correct answer

- Train a global surrogate model and use mimic explainer to explain the model.
- Implement the SHAP tree explainer interpretability technique.
- Use Azure Machine Learning SDK to generate feature importance.
- Configure random sampling to manage the hyperparameter space.

Explanation

You should Use Azure Machine Learning SDK to generate feature importance. 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. Azure Machine Learning supports the Permutation Feature Importance Explainer (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 the SHAP tree explainer interpretability technique. SHAP is not model-agnostic and is used for tree-based models. SHAP explainers use calculates based on coalitional game theory.

You should not train a global surrogate model and use mimic explainer to explain the 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, mimic explainer can be used to interpret the model.

You should not configure random sampling to manage the hyperparameter space. Random sampling is used to control the hyperparameter space that is used during model training.

References

Model interpretability in Azure Machine Learning

5.10 SHAP (SHapley Additive exPlanations)

Tune hyperparameters for your model with Azure Machine Learning

Question 123

You are determining the type of sampling to use for tuning hyperparameters for your experiment.

You need to use the sampling methods that will let you associate an early termination policy. You want to reduce the turning effort involved with configuring various hyperparameters.

Which two sampling methods can you use to meet the goal? Each correct answer presents a complete solution.

Choose the correct answers

- [ ] Grid
- [ ] Bayesian
- [ ] Random

Explanation

You can use the Random or Grid sampling method. Both Random and Grid sampling configurations will let you associate an early cancellation policy.

You should not use the Bayesian sampling method. Bayesian sampling does not let you associate a cancellation policy. When using Bayesian parameter sampling, you should set early_termination_policy = None, or leave the early_termination_policy parameter out.

References

Tune hyperparameters for your model with Azure Machine Learning

Hyperparameter tuning for machine learning models.

Question 125

You use Azure Machine Learning SDK to create and manage machine learning experiments.

You need to consume data from the default workspace datastore in an experiment.

Which two actions should you perform? Each correct answer presents a complete solution.

Choose the correct answers

- [ ] Set the default datastore by using the Workspace set_default_datastore method.
- [ ] Use the get_default_datastore method of the workspace object.
- [ ] Create a reference to workspaceblobstore using the datastore class.
- [ ] Retrieve the workspace's automatically generated Azure file share.

Explanation

You should use the get_default_datastore method of your workspace object. When a new workspace is created, it contains a default datastore, workspaceblobstore. In order to retrieve the default datastore, you can use the get_default_datastore Workspace method. The workspaceblobstore datastore cannot be removed from the workspace.

You can also create a reference to workspaceblobstore using the datastore class. You can use the get method from the datastore class to retrieve a datastore by name. The following code retrieves a datastore named workspaceblobstore:

my_datastore = Datastore.get(workspaceblobstore)

You should not set the default datastore by using the workspace set_default_datastore method. This command does not allow you to retrieve a datastore. You can use this command to set a new default datastore.

You should not retrieve the workspace's automatically generated Azure file share. All workspaces include an automatically registered blob container and file share. The file share - workspacefilestore - is used to store notebooks.

References

Introduction to datastores

Connect to Azure storage services

Question 126

You use Azure Machine Learning to create a machine learning model. You plan to deploy the model as a real-time web service using your local system.

You need to configure deployment settings.

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

Choose the correct answers

- [ ] Specify the service's endpoint port where requests will be accepted.
- [ ] Install Docker on your local machine.
- [ ] Use Aciwebservice.deployconfiguration to specify the number of CPU cores.
- [ ] Set auth_enabled to True in your deployment configuration.

Explanation

You should install Docker on your local machine and specify the service's endpoint port where requests will be accepted. To deploy a model as a real-time web service using your local system, you must install Docker. Docker is a container creation and management platform that can be deployed on a variety of operating systems.

Models deployed locally as a web service will accept requests on an HTTP endpoint. Once your container has the required dependencies installed, you will need to define the port where the HTTP endpoint will listen for service requests. A container is a virtualized app that includes all the resources it needs to run, including file resources, dependencies, and services.

You should not use Aciwebservice.deployconfiguration to specify the number of CPU cores. The deployconfiguration method is used to manage the number of CPU cores and the amount of memory allocated to an Azure Container Instance (ACI).

You should not set auth_enabled to True in your deployment configuration. This setting allows you to enable key-based authentication, but is an optional configuration setting. If you enable key-based authentication on your web service, all service connect attempts will be required to provide a valid Application Programming Interface (API) key prior to accessing the model.

References

Deploy models with Azure Machine Learning

Set up authentication for Azure Machine Learning resources and workflows

Question 127

You want to use an existing Azure Machine Learning environment for future training exercises.

You need to ensure that your machine learning environment is reusable by specifying any additional Azure Machine Learning packages beyond the default that are installed.

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

Explanation

You should complete the following actions in order:

1. Use the az ml folder attach command to create a run configuration.
2. Create an environment.yml file. List all required packages.
3. Set the condaDependenciesFile parameter.

Azure Machine Learning provides the capability to run experiments on different compute targets without requiring scripts to be rewritten. This is done by creating a run configuration, which serves as a template for a training environment. Once created, a run configuration will ensure that the software and dependencies required to run your experiments will be available on any compute resource. The easiest way to generate a run configuration is to use the az ml folder attach command. This command creates a template run configuration file that you can specify when defining a new compute target.

Environment.yml is a Conda environment file. Conda is an open-source environment management system. Within an environment file, you can list any dependencies that are required for your machine learning experiments. Once an environment file is created and configured, you can include the path to your environment file by setting the condaDependenciesFile parameter in your run configuration.

You should not use the az ml workspace command to retrieve your workspace keys. You can use the az ml workspace sync-keys command to manage workspace keys for Azure resources, such as storage spaces.

References

[Set up and use compute targets for model training]

[Managing environments]

[az ml folder]

[Conda]

[az ml workspace]

[Create & use software environments in Azure Machine Learning]

Question 128

You use an Azure Machine Learning designer to train a binary classification model.

When you review the model's metrics in the Evaluate Model module, you notice that the Area Under the Curve (AUC) score is 0.4.

You need to determine the accuracy of the outcomes based on the model AUC score.

What should you conclude about the trained model?

Choose the correct answer

- The model correctly predicts the outcomes 60 percent of the time.
- The model has a difference of 40 percent between False Positives and True Negatives.
- The model outcomes have an average error rate of 0.4.
- The model correctly predicts the outcomes less than 50 percent of the time.

Explanation

The model correctly predicts the outcomes less than 50 percent of the time. AUC is a measure of the quality of the results, which for binary classifications means a better model if the area is closer to 1. In the binary classification model, the graph that represents the curve for efficient models will have the left curve reach near 1 as early as possible, so the area under the curve will increase.

The model does not predict the outcomes 60 percent of the time. An AUC of 0.4 actually means that the curve falls below the 50 percent line.

The model outcomes do not have an average error rate of 0.4. AUC does not measure the average error rate.

The model does not have a difference of 40 percent between False Positives and True Negatives. The AUC is not a measure of just a difference of percentage between False Positives and True Negatives. AUC accounts for all aspects of the confusion matrix using a measure normalized to provide a value between 0 and 1.

References

[Azure Machine Learning: Model Evaluation and Threshold Manipulation]

[Understand automated machine learning results]

Question 129

You have a real-time inference pipeline, as shown in the exhibit. The pipeline takes coma-separated features related to an automobile and predicts its price.

You need to test the pipeline with minimal costs and effort against some test data, which is located in a text file on your machine.

What module should you use?

Choose the correct answer

- Join Data
- Export Data
- Import Data
- Enter Data Manually

Explanation

You should use the Enter Data Manually module. The Enter Data Manually module provides a way to copy and paste the data from your text file to the module and feeds it to the Select Columns in Dataset module.

You should not use the Import Data module. The Import Data module allows you to import data from a data source, such as storage accounts, azure data lakes, Azure SQL etc. Since your data is in a file in your local machine, using this module will add additional steps to create a storage in the cloud and upload the file. This will also increase the cost associated with storing your file in the cloud.

You should not use the Export Data module. The Export Data module is used to write data out to storage in the cloud. You cannot provide the data out to a Select Columns in Dataset module.

You should use the Join Data module. The Join Data module is used to combine data from multiple data sources. This module cannot be used to feed input data from a single source.

References

[What is Machine Learning Studio (classic)?]

[Tutorial: Predict automobile price with the designer]

[Tutorial: Deploy a machine learning model with the designer]

Question 130

Your organization provides you with a set of images for a learning set for an image classification model. You complete the labeling exercise with the multi-label classification option.

You are asked to provide a tabular view of the image and the class the image belongs to for your users.

You need to retrieve the data to display the tabular view with minimal effort.

What should you do?

Choose the correct answer

- Export data labels to the COCO file format.
- Export data labels to the binary file format.
- Add the file name to the labels dataset.
- Name the files with the label that was associated with it during classification.

Explanation

You should export data labels to the COCO file format. Doing so allows you to capture both the reference to the data and its labels. You can then load your exported labeled datasets into a panda dataframe or Torchivision dataset to use open-source libraries for data exploration, and PyTorch provided libraries for image transformation and training. For example, you can use the to_pandas_dataframe() method to display the labeled dataset in a tabular format.

You should not export data labels to the binary format. This is not an option available in Azure for the data classification project created.

You should not name the files with the label associated with it during classification. This would add additional work. Also, this option does not leverage Azure Machine Learning services.

You should not add the file name to the labels dataset. It is not a good practice to modify datasets created by the machine learning models. Also, this would require the manual work of editing the dataset file.

References

[Tag images in a labeling project]

[Create and explore Azure Machine Learning dataset with labels]