Microsoft Azure

Exam DP-100
Designing and Implementing an Azure Data Science Solution

Dry Run session

Case Study 1

Complete the Case Study

- Solution Evaluation
- Question 1
- Question 2
- Question 3

Instructions

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

Note: You cannot go back or review questions of this type on the actual certification exam.

Case Study 1

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

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

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

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

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

Solution Case Study 1

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

Q1

```
from interpret.ext.blackbox import TabularExplainer

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

Does the solution meet the goal?

- Yes
- No

Q2

```
from interpret.ext.blackbox import MimicExplainer

explainer = MimicExplainer(model, x_train,
                           LGBMExplainableModel,
                           augment_data=True,
                           max_num_of_augmentations=10,
                           features=breast_cancer_data.feature_names,
                           classes=classes)
```

Does the solution meet the goal?

- Yes
- No

Q3

```
from interpret.ext.blackbox import PFIExplainer

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

Does the solution meet the goal?

- Yes
- No

Solution Case Study 1

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

Q1

```
from interpret.ext.blackbox import TabularExplainer

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

Does the solution meet the goal?

- (●) Yes
- ( ) No

Q2

```
from interpret.ext.blackbox import MimicExplainer

explainer = MimicExplainer(model, x_train,
                           LGBMExplainableModel,
                           augment_data=True,
                           max_num_of_augmentations=10,
                           features=breast_cancer_data.feature_names,
                           classes=classes)
```

Does the solution meet the goal?

- ( ) Yes
- (●) No

Q3

```
from interpret.ext.blackbox import PFIExplainer

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

Does the solution meet the goal?

- ( ) Yes
- (●) No

Explanation Case Study 1

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

Q1

```
from interpret.ext.blackbox import TabularExplainer

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

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

References

Use Python to interpret & explain models (preview) - Azure Machine Learning | Microsoft Learn

Welcome to the SHAP documentation — SHAP latest documentation

SHAP Values Explained Exactly How You Wished Someone Explained to You | by Samuele Mazzanti | Towards Data Science

Case Study 2

Complete the Case Study

- Solution Evaluation
- Question 1
- Question 2
- Question 3

Instructions

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

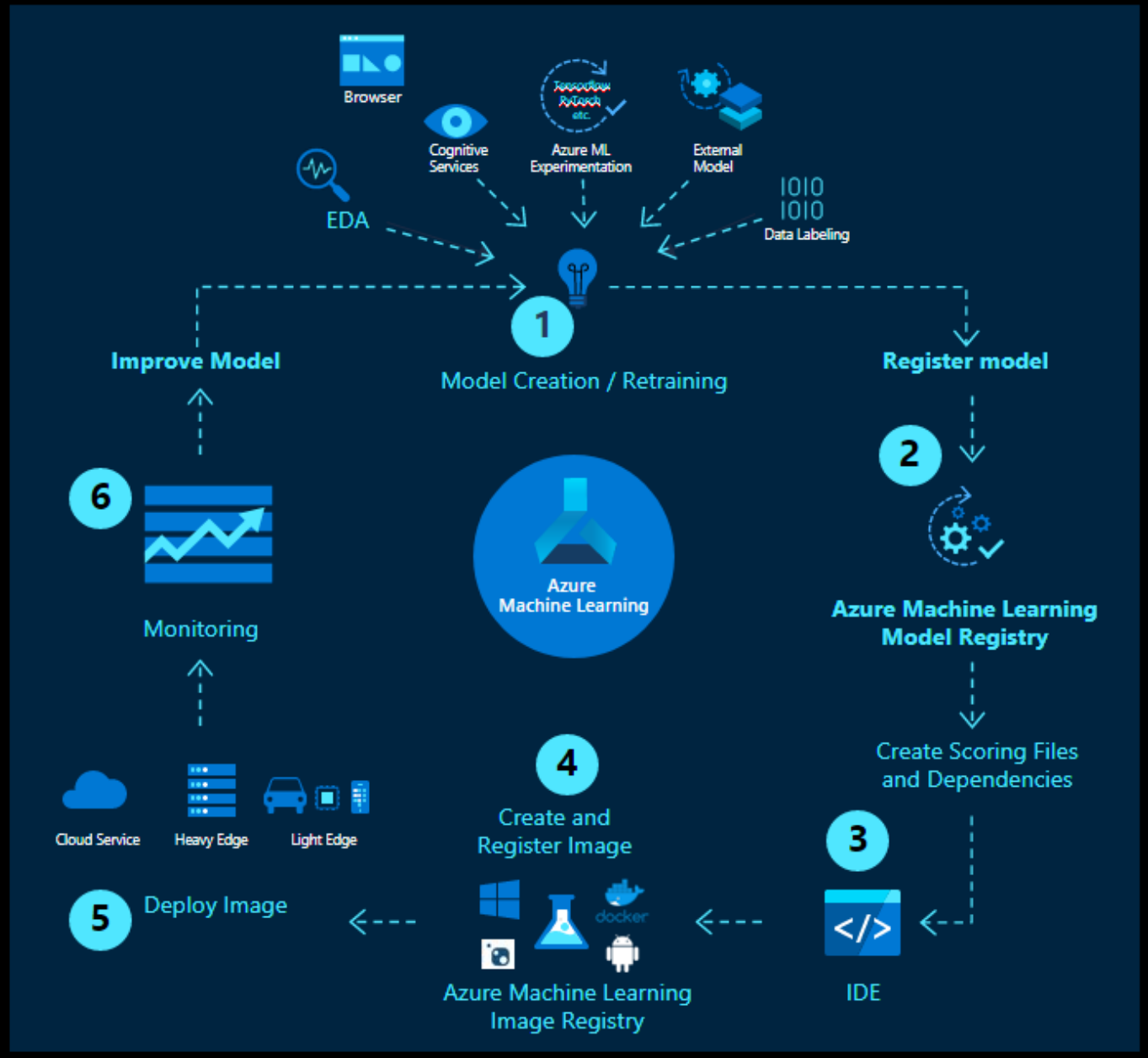
Note: You cannot go back or review questions of this type on the actual certification exam.

Case Study 2

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

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

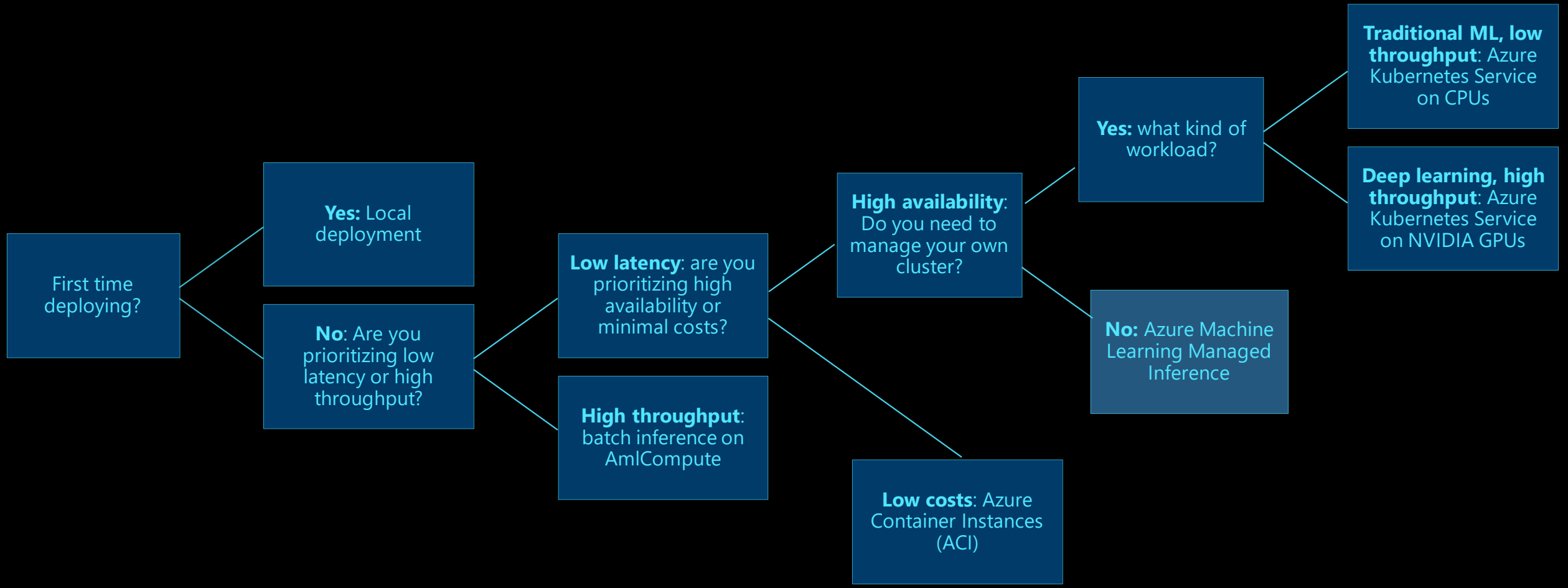
Azure ML Service can implement the end-to-end ML lifecycle



Workflow Steps

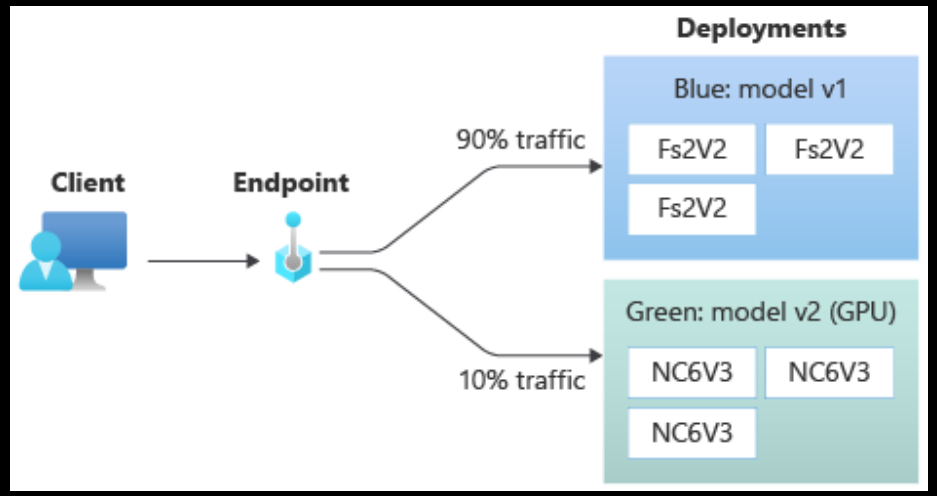
1. a) Develop machine learning training scripts in Python, using AutoML, Designer, Notebooks etc.
   b) Create and configure a compute target
   c) Submit the scripts to the configured compute target to run in that environment. During training, the compute target stores run records to a datastore. There the records are saved to an experiment.
   b) Review the experiment for logged metrics from the current and past runs. If the metrics do not indicate a desired outcome, loop back to step a) and iterate on your scripts.
2. Once a satisfactory run is found, register the persisted model in the model registry.
3. Develop a scoring script.
4. Create an Image and register it in the image registry.
5. Deploy the image as a web service in Azure, or on the edge
6. Monitor the model in production and identify when further improvements / adjustments are needed including model and data drift.

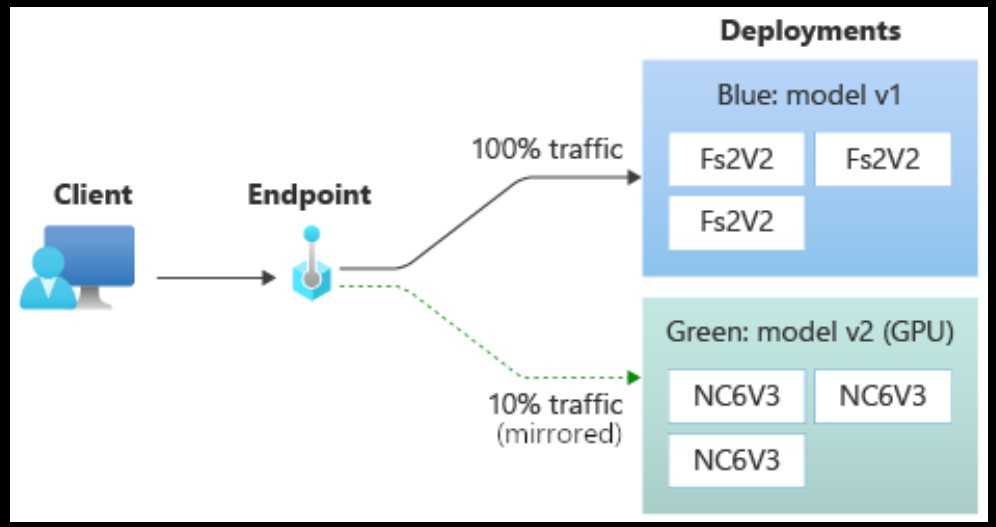
Choosing an inferencing target



Managed Endpoints - GA

•Azure Machine Learning **managed endpoints**, now generally available, help developers and data scientists more easily deploy large-scale machine learning models for both real-time and batch inferencing.





https://docs.microsoft.com/en-us/azure/machine-learning/concept-endpoints

Solution Case Study 2

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

Q1 Execute AksWebservice.deploy_configuration(token_auth_enabled=True, auth_enabled=False)

Does the solution meet the goal?

- Yes
- No

Q2 Execute AksWebservice.deploy_configuration(auth_enabled=True)

Does the solution meet the goal?

- Yes
- No

Q3 Execute AksWebservice.deploy_configuration()

Does the solution meet the goal?

- Yes
- No

Solution Case Study 2

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

Q1 Execute AksWebservice.deploy_configuration(token_auth_enabled=True, auth_enabled=False)

Does the solution meet the goal?

- (●) Yes
- ( ) No

Q2 Execute AksWebservice.deploy_configuration(auth_enabled=True)

Does the solution meet the goal?

- ( ) Yes
- (●) No

Q3 Execute AksWebservice.deploy_configuration()

Does the solution meet the goal?

- ( ) Yes
- (●) No

Explanation Case Study 2

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

Q1 Execute AksWebservice.deploy_configuration(token_auth_enabled=True, auth_enabled=False)

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

References

Deploy ML models to Kubernetes Service with v1 - Azure Machine Learning | Microsoft Learn

azureml.core.webservice.AksWebservice class - Azure Machine Learning Python | Microsoft Learn

Q3

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

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

You need to ensure that your experiment functions properly.

What should you do?

Choose the correct answer

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

Q3

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

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

You need to ensure that your experiment functions properly.

What should you do?

Choose the correct answer

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

Explanation Q3

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

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

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

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

References

What is automated ML? AutoML - Azure Machine Learning | Microsoft Learn

Set up AutoML with Python (v2) - Azure Machine Learning | Microsoft Learn

azureml.train.automl.automlconfig.AutoMLConfig class - Azure Machine Learning Python | Microsoft Learn

Cross Validate Model: Component reference - Azure Machine Learning | Microsoft Learn

Q4

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

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

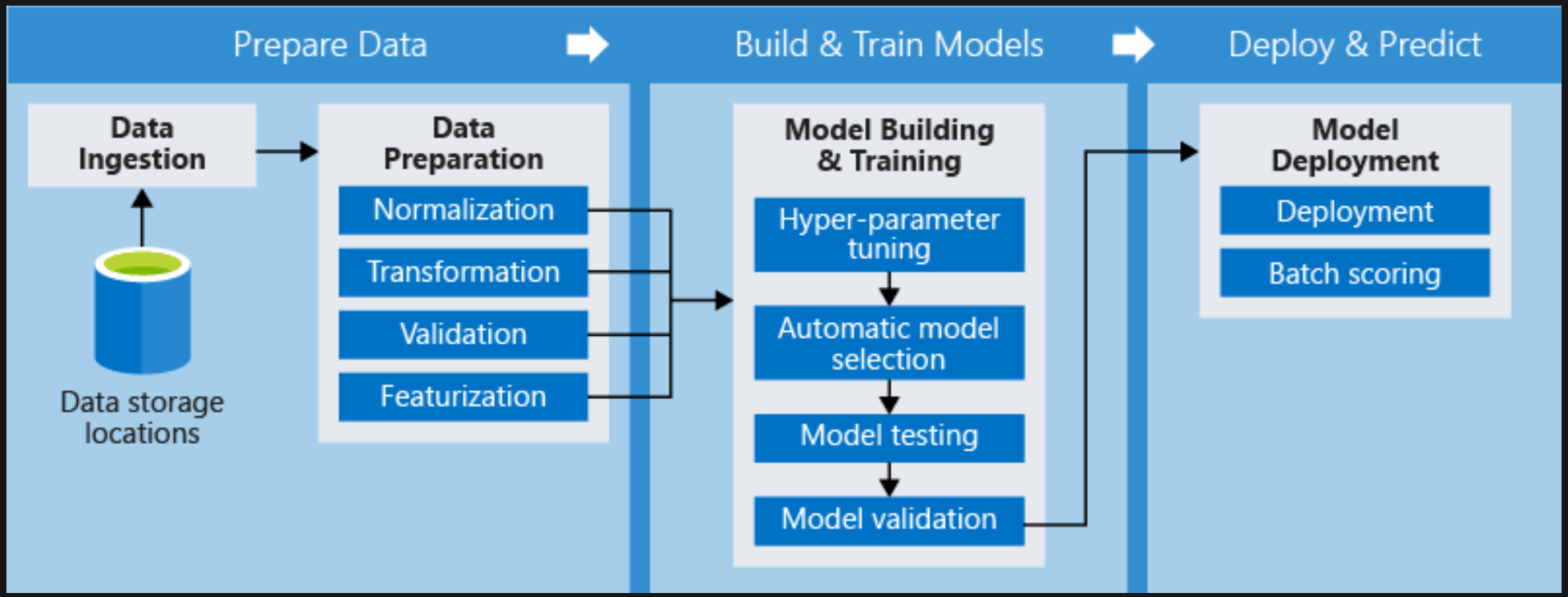
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of possible actions to the answer area and arrange them in the correct order.

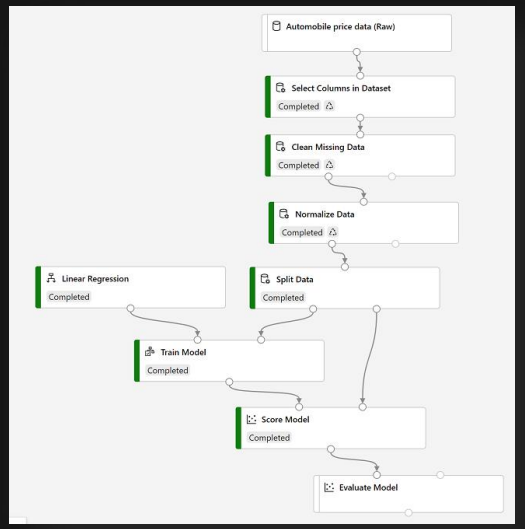
Create a list in the correct order

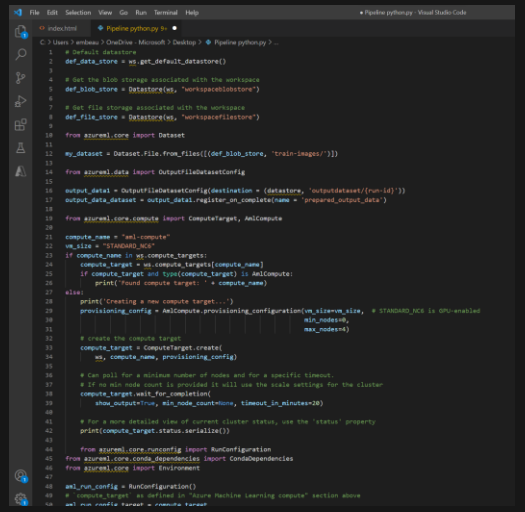
Possible actions	Actions in order
Specify a PipelineData object for data output.	
Register a new Azure Storage file container datastore.	
Create a PipelineData object. Specify a name and output datastore.	
Retrieve the default datastore from the current workspace.	
Register a new dataset version for each pipeline pass.	



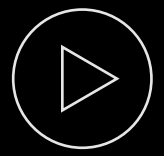

Pipeline artifact







Advantages of Azure ML Pipelines

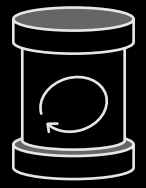


Unattended runs

Schedule a few steps to run in parallel or in sequence to focus on other tasks while your pipeline runs

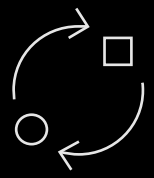
Tracking and versioning

Name and version your data sources, inputs and outputs with the pipelines SDK



Reusability

Create templates of pipelines for specific scenarios such as retraining and batch scoring



Mixed and diverse compute

Use multiple pipelines that are reliably coordinated across heterogeneous and scalable computes and storages

Q4

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

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

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

Create a list in the correct order

Possible actions

- Retrieve the default datastore from the current workspace.
- Register a new dataset version for each pipeline pass.

Actions in order

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


Explanation Q4

You should perform the following steps in order:

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

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

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

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

my_datastore = Datastore.get(default-datastore)

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

References

[Pass data between pipeline steps - Training | Microsoft Learn]

Q5

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

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

What should you do?

Choose the correct answer

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

Q5

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

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

What should you do?

Choose the correct answer

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

Explanation Q5

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

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

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

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

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

References

[Detect data drift on datasets (preview) - Azure Machine Learning | Microsoft Learn]

Q6

You manage an Azure Machine Learning workspace.

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

What should you do?

Choose the correct answer

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

Q6

You manage an Azure Machine Learning workspace.

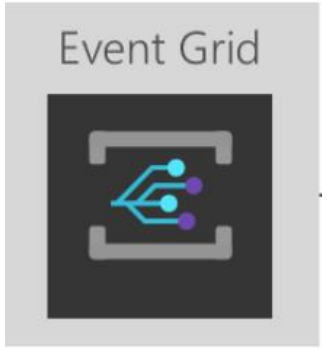
You need to send an HTTP push notification to an external system when a machine learning model is registered or deployed in the workspace.

What should you do?

Choose the correct answer

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

Notify, Automate, Alert with Azure EventGrid



https://learn.microsoft.com/en-us/azure/machine-learning/how-to-use-event-grid

Notify, Automate, Alert with Azure EventGrid (Preview)

Microsoft.MachineLearningServices.**RunCompleted**

- Raised when a machine learning experiment run is completed

Microsoft.MachineLearningServices.**ModelRegistered**

- Raised when a machine learning model is registered in the workspace

Microsoft.MachineLearningServices.**ModelDeployed**

- Raised when a deployment of inference service with one or more models is completed

Microsoft.MachineLearningServices.**DatasetDriftDetected**

- Raised when a data drift detection job for two datasets is completed

Microsoft.MachineLearningServices.**RunStatusChanged**

- Raised when a run status changed, currently only raised when a run status is 'failed'

Explanation Q6

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

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

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

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

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

References

[What is Azure Event Grid? - Azure Event Grid | Microsoft Learn]

[Tutorial: Designer - deploy no-code models - Azure Machine Learning | Microsoft Learn]

[Set up authentication - Azure Machine Learning | Microsoft Learn]

[Event Hubs—Real-Time Data Ingestion | Microsoft Azure]

Q7

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

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

Q7

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

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

Explanation Q7

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 - Azure Machine Learning | Microsoft Learn

Q8

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

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

Q8

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

- [ ] 8 GB
- [ ] 10 GB
- [x] 20 GB
- [ ] 2 GB

Explanation Q8

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 Data Assets - Azure Machine Learning | Microsoft Learn

Q9

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?

```
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_auto_approval=True,
                 sku='enterprise')
```

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

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

Q9

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?

○
```
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_auto_approval=True,
                 sku='enterprise')
```

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

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

Explanation Q9

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

azureml.core.workspace.Workspace class - Azure Machine Learning Python | Microsoft Learn

Q10

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

You need to write a batch inference script.

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

Choose the correct answers

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

Q10

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

You need to write a batch inference script.

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

Choose the correct answers

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

Explanation Q10

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

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

References

Tutorial: ML pipelines with Python SDK v2 - Azure Machine Learning | Microsoft Learn

Q11

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

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

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

Choose the correct answers

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

Q11

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

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

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

Choose the correct answers

- [x] Configure accuracy as the primary metric.
- [x] Set the primary metric goal to MAXIMIZE.
- [ ] Configure Spearman correlation as the primary metric.
- [ ] Set the primary metric goal to MINIMIZE.

Explanation Q11

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

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

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

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

References

Hyperparameter tuning a model (v2) - Azure Machine Learning | Microsoft Learn

Evaluate AutoML experiment results - Azure Machine Learning | Microsoft Learn

Q12

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 = Select your answer (
    {
        '--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=' Select your answer ',
                          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 = Select your (workspace = ws, name = 'infection_training_hyperdrive')
run = experiment.Select you (config=hyperdrive)
```

params =
- Select your answer
- BayesianParameterSampling
- GridParameterSampling
- RandomParameterSampling

entry_script='
- Select your answer
- infection_training.py
- infection_ds

experiment =
- Select your answer
- Experiment
- Workspace

run = experiment.
- Select your answer
- submit
- get_runs

Q12

```
params = GridParameterSampling     (
    {
        '--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='infection_training.py   ',
                          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 = Experiment   (workspace = ws, name = 'infection_training_hyperdrive')
run = experiment.submit   (config=hyperdrive)
```

Explanation Q12

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

Hyperparameter tuning a model (v2) - Azure Machine Learning | Microsoft Learn

Q13

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

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

Q13

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

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

Explanation Q13

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 - Azure Machine Learning | Microsoft Learn

Convert to Dataset: Component reference - Azure Machine Learning | Microsoft Learn

Q14

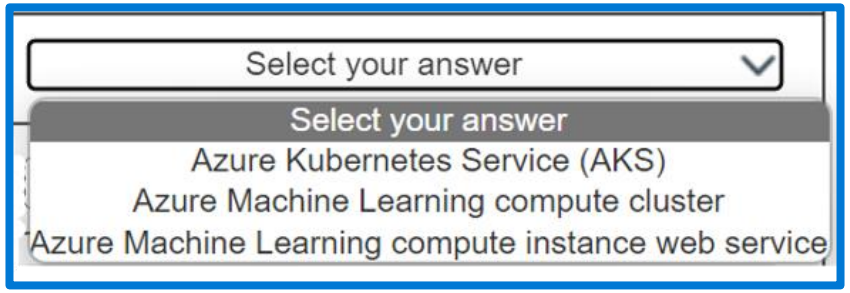
You are preparing to register multiple trained models using Azure Machine Learning.

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

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

Choose the correct options

Model Usage Scenario	Deployment Target
Batch inference	Select your answer
Real-time inference	Select your answer
Testing and debugging	Select your answer



Q14

Model Usage Scenario	Deployment Target
Batch inference	Azure Machine Learning compute cluster
Real-time inference	Azure Kubernetes Service (AKS)
Testing and debugging	Azure Machine Learning compute instance web s

Explanation Q14

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

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

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

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

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

References

Deploy machine learning models to online endpoints - Azure Machine Learning | Microsoft Learn

What are compute targets - Azure Machine Learning | Microsoft Learn

Q15

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

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

Q15

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

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

Explanation Q15

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 - Training | Microsoft Learn

Use datastores - Azure Machine Learning | Microsoft Learn

Q16

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
- Tensorflow
- Chainer
- PyTorch

Q16

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
- ( ) Tensorflow
- ( ) Chainer
- ( ) PyTorch

Explanation Q16

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

Train scikit-learn machine learning models (v2) - Azure Machine Learning | Microsoft Learn
Train and deploy a TensorFlow model (SDK v2) - Azure Machine Learning | Microsoft Learn
Train deep learning Keras models (SDK v2) - Azure Machine Learning | Microsoft Learn
Train deep learning PyTorch models (SDK v2) - Azure Machine Learning | Microsoft Learn

Q17

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

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

What should you do?

Choose the correct answer

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

Q17

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

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

What should you do?

Choose the correct answer

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

Explanation Q17

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

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

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

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

References

Train and deploy a reinforcement learning model (preview) - Azure Machine Learning | Microsoft Learn

Train with machine learning datasets - Azure Machine Learning | Microsoft Learn

Configure a training job - Azure Machine Learning | Microsoft Learn

Train scikit-learn machine learning models (v2) - Azure Machine Learning | Microsoft Learn

Q18

You create Azure Machine Learning workspaces.

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

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

Create a list in the correct order

Possible actions

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

Actions in order

Q18

Possible actions

- Use Azure Machine Learning SDK to create an Azure Machine Learning workspace. Set the SKU to Basic.
- Create a multi-label image classification project.

Actions in order

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

Explanation Q18

You should complete the following actions in order:

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

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

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

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

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

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

References

Set up image labeling project - Azure Machine Learning | Microsoft Learn

What is a workspace? - Azure Machine Learning | Microsoft Learn

Q19

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.	○	○

Q19

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.	○	○

Q19

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 Q19

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 on your production models - Azure Machine Learning | Microsoft Learn

azureml.datadrift.datadriftdetector.DataDriftDetector class - Azure Machine Learning Python | Microsoft Learn

Q20

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

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

Q20

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

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

Explanation Q20

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

Configure a training job - Azure Machine Learning | Microsoft Learn

What is an Azure Machine Learning compute instance? - Azure Machine Learning | Microsoft Learn

Data ingestion with Azure Data Factory - Azure Machine Learning | Microsoft Learn

Q21

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

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

Q21

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

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

Explanation Q21

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: Component reference - Azure Machine Learning | Microsoft Learn

azureml.core.Run class - Azure Machine Learning Python | Microsoft Learn

Q22

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

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

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

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

Which Azure Machine Learning module should you use?

Choose the correct answer

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

Q22

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

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

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

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

Which Azure Machine Learning module should you use?

Choose the correct answer

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

Explanation Q22

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

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

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

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

References

Clean Missing Data: Component Reference - Azure Machine Learning | Microsoft Learn

Select Columns in Dataset: Component Reference - Azure Machine Learning | Microsoft Learn

Normalize Data: Component Reference - Azure Machine Learning | Microsoft Learn

Partition and Sample: Component reference - Azure Machine Learning | Microsoft Learn

Q23

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

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

Q23

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

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

Explanation Q23

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

Microsoft Azure Stack Edge Pro FPGA overview | Microsoft Learn

Deploy ML models to Kubernetes Service with v1 - Azure Machine Learning | Microsoft Learn

Deploy machine learning models to online endpoints - Azure Machine Learning | Microsoft Learn

Q24

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

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

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

Which line of code should you add to your script?

Choose the correct answer

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

Q24

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

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

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

Which line of code should you add to your script?

Choose the correct answer

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

Explanation Q24

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

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

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

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

References

Where to save & write experiment files - Azure Machine Learning | Microsoft Learn

Introduction to datastores - Training | Microsoft Learn

azureml.core.Run class - Azure Machine Learning Python | Microsoft Learn

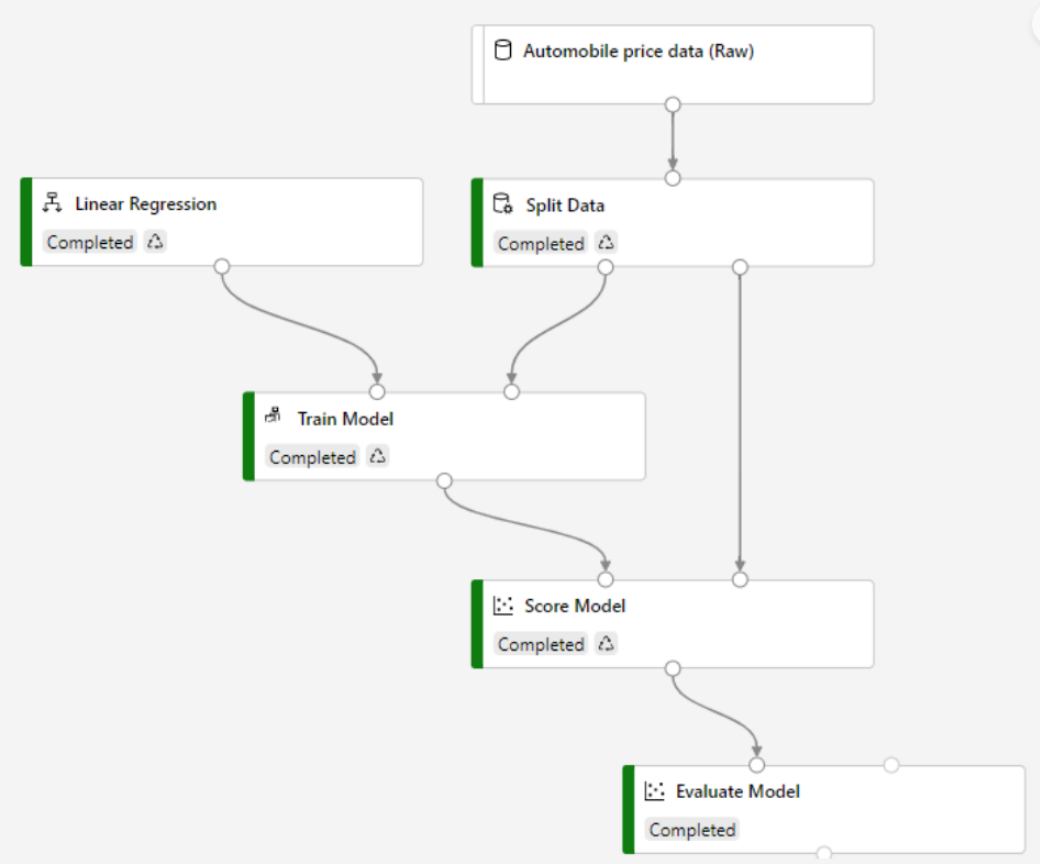
Q25

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

You need to submit the run and understand the model.

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

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



Q25

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

You need to submit the run and understand the model.

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

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

Explanation Q25

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

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

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

References

What is the Azure Machine Learning designer? - Azure Machine Learning | Microsoft Learn

Tutorial: Designer - train a no-code regression model - Azure Machine Learning | Microsoft Learn

Q26

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.	○	○

Q26

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 Q26

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 the Azure Machine Learning designer? - Azure Machine Learning | Microsoft Learn]

Q27

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.
- Use Aciwebservice.deployconfiguration to specify the number of CPU cores.
- Set auth_enabled to True in your deployment configuration.
- Install Docker on your local machine.

Q27

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

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

Explanation Q27

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

Set up authentication - Azure Machine Learning | Microsoft Learn

Q28

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

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

Q28

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

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

Explanation Q28

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

Monitor and manage Hadoop with Ambari REST API - Azure HDInsight | Microsoft Learn

What is Microsoft Sentinel? | Microsoft Learn

Azure HDInsight SDK for .NET - Azure for .NET Developers | Microsoft Learn

Q29

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

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

Which termination policy should you use?

Choose the correct answer

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

Q29

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

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

Which termination policy should you use?

Choose the correct answer

- [ ] Bandit policy
- [ ] No termination policy
- [x] Truncation selection policy
- [ ] Median stopping policy

Explanation Q29

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

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

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

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

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

References

Hyperparameter tuning a model (v2) - Azure Machine Learning | Microsoft Learn

Q30

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

Q30

Your organization provides you with a set of images for a learning set for an im
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 b

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

What should you do?

Choose the correct answer

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


```
"images": [
    {
        "license": 4,
        "file_name": "000000397133.jpg",
        "coco_url": "http://images.cocodataset.org/val2017/000000397133.jpg",
        "height": 427,
        "width": 640,
        "date_captured": "2013-11-14 17:02:52",
        "flickr_url": "http://farm7.staticflickr.com/6116/6255196340_da26cf2c9e_z.jpg",
        "id": 397133
    },
    {
        "license": 1,
        "file_name": "000000037777.jpg",
        "coco_url": "http://images.cocodataset.org/val2017/000000037777.jpg",
        "height": 230,
        "width": 352,
        "date_captured": "2013-11-14 20:55:31",
        "flickr_url": "http://farm9.staticflickr.com/8429/7839199426_f6d48aa585_z.jpg",
        "id": 37777
    },
    ...
]
```

Explanation Q30

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

Labeling images and text documents - Azure Machine Learning | Microsoft Learn

Create and explore datasets with labels - Azure Machine Learning | Microsoft Learn

Q31

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

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

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

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

Create a list in the correct order

Possible commands

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

Commands in order

Q31

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

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

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

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

Create a list in the correct order

Possible commands

- jupyter notebook
- conda activate AzureML

Commands in order

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

Explanation Q31

You should execute the commands in the following order:

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

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

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

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

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

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

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

References

Set up Python development environment - Azure Machine Learning | Microsoft Learn

Q32

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

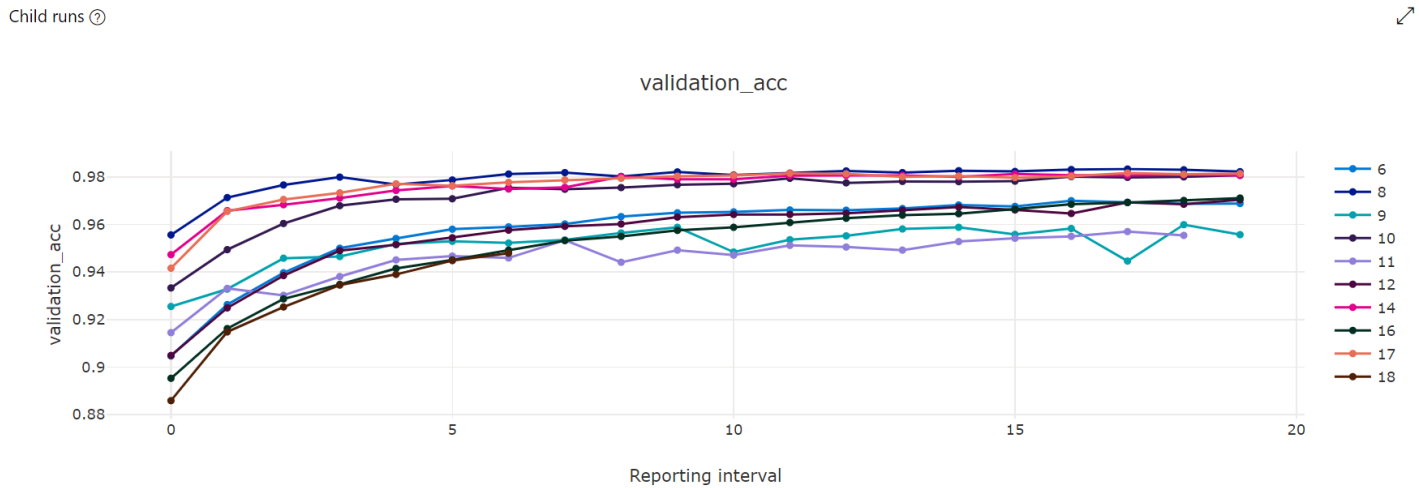
```
from azureml.train.hyperdrive import GridParameterSampling
from azureml.train.hyperdrive import choice

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

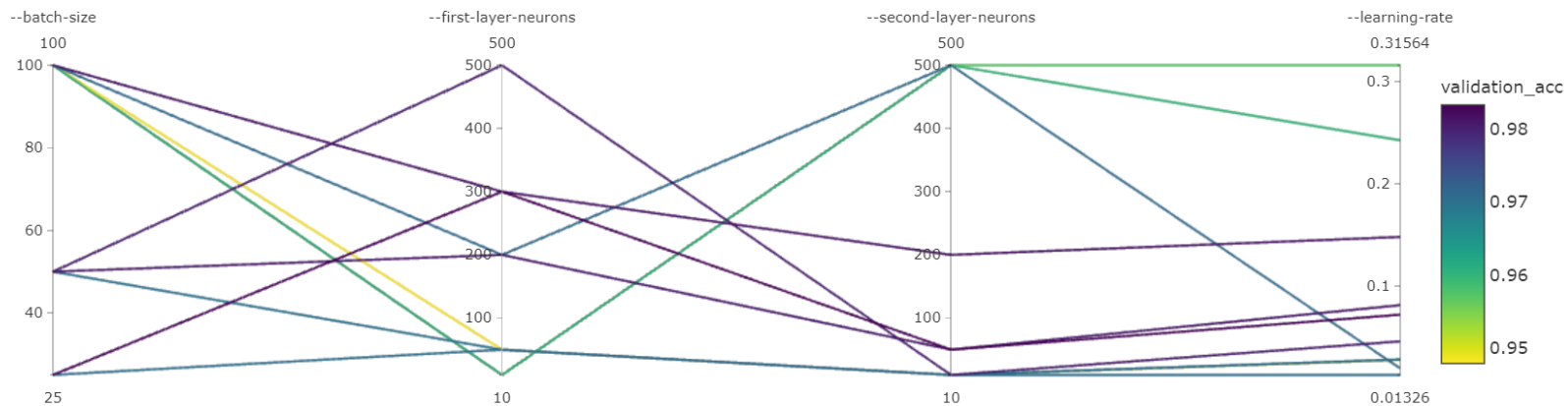
You need to configure the hyperparameter values when training the model.

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.	○	○

1. Define the parameter search space
2. Specify a primary metric to optimize
3. Specify early termination policy for low-performing runs
4. Create and assign resources
5. Launch an experiment with the defined configuration
6. Visualize the training runs
7. Select the best configuration for your model





Q32

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 Q32

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

Hyperparameter tuning a model (v2) - Azure Machine Learning | Microsoft Learn

azureml.train.hyperdrive.GridParameterSampling class - Azure Machine Learning Python | Microsoft Learn

Q33

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

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

Which step should you use?

Choose the correct answer

- PythonScriptStep
- ParallelRunConfig
- AdlaStep
- ParallelRunStep

Q33

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

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

Which step should you use?

Choose the correct answer

- PythonScriptStep
- ParallelRunConfig
- AdlaStep
- ParallelRunStep

Explanation Q33

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

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

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

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

References

Tutorial: ML pipelines with Python SDK v2 - Azure Machine Learning | Microsoft Learn

azureml.pipeline.steps.adla_step.AdlaStep class - Azure Machine Learning Python | Microsoft Learn

azureml.pipeline.steps.PythonScriptStep class - Azure Machine Learning Python | Microsoft Learn

azureml.pipeline.steps.parallel_run_step.ParallelRunStep class - Azure Machine Learning Python | Microsoft Learn

Q34

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

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

Q34

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

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

Explanation Q34

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

Troubleshooting online endpoints deployment - Azure Machine Learning | Microsoft Learn

503 Service Unavailable - HTTP | MDN (mozilla.org)

Q35

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.

Drag and drop the answers

Requirement	Compute Target
Tune hyperparameters using Azure Machine Learning (ML) designer.	
Use your own virtual machine, attached to your virtual network, for hyperparameter tuning.	
Use Apache Spark to train your model.	
Auto scale instances for models based on compute requirements.	

Azure ML compute cluster | **Azure HDInsight** | **Remote VM**

Q35

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.

Drag and drop the answers

Requirement	Compute Target
Tune hyperparameters using Azure Machine Learning (ML) designer.	Azure ML compute cluster
Use your own virtual machine, attached to your virtual network, for hyperparameter tuning.	Remote VM
Use Apache Spark to train your model.	Azure HDInsight
Auto scale instances for models based on compute requirements.	Azure ML compute cluster

Azure ML compute cluster | Azure HDInsight | Remote VM

Explanation Q35

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

Configure a training job - Azure Machine Learning | Microsoft Learn

Q36

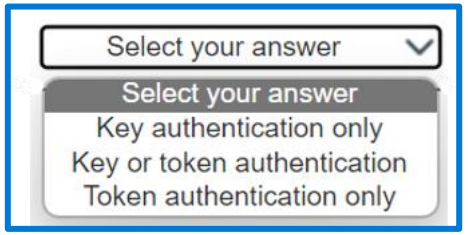
You use Azure Machine Learning to create machine learning models. You plan to deploy your models as web services using various compute targets.

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)	Select your answer
Azure Kubernetes Service (AKS)	Select your answer



Q36

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)	Key authentication only
Azure Kubernetes Service (AKS)	Key or token authentication

Explanation Q36

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 - Azure Machine Learning | Microsoft Learn

Serverless containers in Azure - Azure Container Instances | Microsoft Learn

Introduction to Azure Kubernetes Service - Azure Kubernetes Service | Microsoft Learn

Q37

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

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

Which code segment should you use?

Choose the correct answer

```
from azureml.core.webservice import Webservice

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

```
from azureml.core import Experiment

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

```
from azureml.core.compute import ComputeTarget

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

```
from azureml.core import Experiment

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

Q37

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

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

Which code segment should you use?

Choose the correct answer

- from azureml.core.webservice import Webservice

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

- from azureml.core import Experiment

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

- from azureml.core.compute import ComputeTarget

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

- from azureml.core import Experiment

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

Explanation Q37

You should use:

```
from azureml.core import Experiment

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

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

You should not use:

```
from azureml.core import Experiment

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

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

You should not use:

```
from azureml.core.compute import ComputeTarget

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

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

You should not use:

```
from azureml.core.webservice import Webservice

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

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

References

Log metrics, parameters and files with MLflow - Azure Machine Learning | Microsoft Learn

Monitor and analyze jobs in studio - Azure Machine Learning | Microsoft Learn

Q38

Your model training script includes the following code:

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

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

What should you do?

Choose the correct answer

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

Q38

Your model training script includes the following code:

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

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

What should you do?

Choose the correct answer

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

Explanation Q38

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

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

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

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

References

Log metrics, parameters and files with MLflow - Azure Machine Learning | Microsoft Learn

azureml.core.Run class - Azure Machine Learning Python | Microsoft Learn

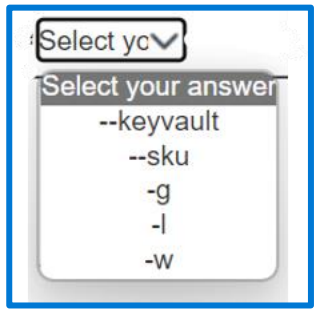
Q39

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

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

Choose the correct options

az ml workspace create [Select yc] AML-Workspace [Select yc] AML-LearningResources



Q39

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

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

Choose the correct options

az ml workspace create [-w] AML-Workspace [-g] AML-LearningResources

Q39

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

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

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

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

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

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

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

References

Create workspaces with Azure CLI - Azure Machine Learning | Microsoft Learn

az ml workspace | Microsoft Learn

Azure Machine Learning workspaces - Training | Microsoft Learn

Q40

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

You need to determine a deployment target compute resource that can be used for testing and debugging while incurring minimal cost.

What two deployment targets should you select? Each correct answer presents a complete solution.

Choose the correct answers

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

Q40

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

You need to determine a deployment target compute resource that can be used for testing and debugging while incurring minimal cost.

What two deployment targets should you select? Each correct answer presents a complete solution.

Choose the correct answers

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

Explanation Q40

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

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

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

References

[What are compute targets - Azure Machine Learning | Microsoft Learn]

Microsoft Azure