Point 1:

You are planning to host practical training to acquaint staff with Docker for Windows. Staff devices must support the installation of Docker.

Docker Requirements

1. 4 GB of system RAM
2. BIOS-enabled virtualization
3. Windows 10 64-bit

Point 2:

GPU is required for a Deep Learning Virtual Machine (DLVM) to support Compute Unified Device Architecture (CUDA) computations.

Point 3:

You need to implement a Data Science Virtual Machine (DVSM) that supports the Caffe2 Deep Learning (DL) framework.

Ubuntu 16.04 supports Caffe2 deep learning framework.

Point 4:

You have been tasked with employing a Machine Learning (ML) model, which makes use of a PostgreSQL database and needs GPU processing, to forecast prices. You are preparing to create a VM that has the necessary tools built into it.

Geo AI Data Science Virtual Machine (Geo-DSVM) Windows edition cannot be used for forecasting prices.

The Azure Geo AI Data Science VM (Geo-DSVM) delivers geospatial analytics capabilities from Microsoft’s Data Science VM.

Point 5:

You have been tasked with employing a ML model, which makes use of a PostgreSQL database and needs GPU processing, to forecast prices. You are preparing to create a VM that has the necessary tools built into it.

Deep Learning Virtual Machine (DLVM) Windows edition can be used for forecasting prices.

Point 6:

You have been tasked with employing a ML model, which makes use of a PostgreSQL database and needs GPU processing, to forecast prices. You are preparing to create a VM that has the necessary tools built into it.

Data Science Virtual Machine (DSVM) Windows edition cannot be used to forecasting prices.

Point 7:

To move data into Azure Blob Storage for the purpose of supporting Azure ML, you can utilise

1. AzCopy
2. SSIS (SQL Server Integration Services)
3. Azure Storage Explorer

Point 8:

To move a large dataset from Azure ML Studio to a Weka environment, the data must be converted to the ARFF (Attribute-Relation File Format) format.

Point 9:

You have been tasked with designing a DL model, which accommodates the most recent edition of python, to recognize language.

TensorFlow is a suitable DL framework for designing a DL model in the Data Science Virtual Machine (DSVM).

Point 10:

You have been tasked with evaluating your model on a partial data sample via k-fold cross-validation. You have already configured a k-parameter as the number of splits. You now have to configure the k parameter for the cross-validation with the usual value choice.

Recommendation: You configure the use of the value k=3

The requirements will not be satisfied.

Point 11:

You have been tasked with evaluating your model on a partial data sample via k-fold cross-validation. You have already configured a k-parameter as the number of splits. You now have to configure the k parameter for the cross-validation with the usual value choice.

Recommendation: You configure the use of the value k=10

The requirements will be satisfied.

Leave One Out (LOO) cross-validation. LOO CV is sometimes useful but typically doesn’t shake up the data enough. The estimates from each fold are highly correlated and hence their average can have high variance. This is why the usual choice is K=5 or 10. It provides a good compromise for the bias-variance trade-off.

Point 12:

Split Data Module can split date into two separate datasets while constructing a ML experiment via Azure ML Studio.

Point 13:

You have been tasked with creating a new Azure pipeline via the ML designer. You have to make sure that the pipeline trains a model using data in a comma-separated values (CSV) file that is published on a website. A dataset for the file does not exist.

Import Data object to the pipeline can be used to ingest data from the CSV file into the designer pipeline with the least amount of administrative effort.

Point 14.

You are in process of creating a ML model. Your dataset includes rows with null and missing values. You plan to make use of the Clean Missing Data module in Azure ML Studio to detect and fix the null and missing values in the dataset.

Recommendation: You make use of the Replace with median option.

Yes. It can be used to detect and fix null and missing values in the dataset using the Clean Missing Data module.

Point 15.

You are in process of creating a ML model. Your dataset includes rows with null and missing values. You plan to make use of the Clean Missing Data module in Azure ML Studio to detect and fix the null and missing values in the dataset.

Recommendation: You make use of the Custom substitution value option.

Yes. It can be used to detect and fix null and missing values in the dataset using the Clean Missing Data module.

Point 16.

You are in process of creating a ML model. Your dataset includes rows with null and missing values. You plan to make use of the Clean Missing Data module in Azure ML Studio to detect and fix the null and missing values in the dataset.

Recommendation: You make use of the Remove entire row option.

Yes. It can be used to detect and fix null and missing values in the dataset using the Clean Missing Data module.

Point 17.

To transform a categorical feature into a binary indicator, you should make use of the Convert to Indicator Values module in Azure ML Studio. The purpose of this module is to convert columns that contain categorical values into a series of binary indicator columns that can more easily be used as features in a ML model.

Point 18.

To improve the amount of low incidence cases in a dataset, you should make use of the SMOTE module. Use the SMOTE module in Azure ML Studio to increase the number of underrepresented cases in a dataset used for ML.

Point 19.

Box Plot visualization can be used to reveal outliers in your data.

Point 20.

You are planning to host practical training to acquaint learners with data visualization creation using python. Learner devices are able to connect to the internet. Learner devices are currently NOT configured for python development. Also, learners are unable to install software on their devices as they lack administrator permissions. Furthermore, they are unable to access Azure subscriptions. It is imperative that learners are able to execute python-based data visualization code.

Consider configuring to the use of Azure Notebooks.

Point 21.

Probabilistic PCA is a data cleaning option of the Clean Missing Data module that does not require predictors for each column.

Point 22.

Binary classification confusion matrix visualization allows for precision to be used as the measurement for assessing binary classification ML model.

Point 23.

You have been tasked with evaluating your model on a partial data sample via k-fold cross-validation. You have already configured a k-parameter as the number of splits. You now have to configure the k parameter for the cross-validation with the usual value choice. Recommendation: You configure the use of the value k=1

The requirements will not be satisfied.

Point 24.

You are in the process of constructing a regression model. You would like to make it a Poisson regression model.

Poisson regression model requires feature values to be

1. Whole numbers
2. Positive value

A Poisson distribution is a discrete distribution; therefore, it is not meaningful to use this method with non-whole numbers. Counts cannot be negative. The method will fail outright if you attempt to use it with negative labels.

Point 25.

You are in the process of carrying out feature engineering on a dataset. You want to add a feature to the dataset and fill the column value.

Recommendation: You must use Group Categorical Values Azure ML studio

The requirements will not be satisfied.

Point 26.

You are in the process of carrying out feature engineering on a dataset. You want to add a feature to the dataset and fill the column value.

Recommendation: You must use Join Data Azure ML studio

The requirements will not be satisfied.

Point 27.

You are in the process of carrying out feature engineering on a dataset. You want to add a feature to the dataset and fill the column value.

Recommendation: You must use Edit Metadata Azure ML studio

The requirements will not be satisfied.

Typical metadata changes might include marking columns as features.

Point 28.

To ascertain if two sets of data differ considerably, we perform a paired t-test.

Conditions that must apply to use a paired test

1. A matched pairs of scores
2. Sampling distribution of data is normal

Point 29.

You want to train a classification model using data located in a CSV file. The classification model will be trained via the Automated ML interface using the Classification task type. You have been informed that only linear models need to be assessed by the Automated ML.

You should Disabling Deep Learning.

Point 30.

Suppose you have to train a regression model via automated ML. The data available to you has features with missing values, as well as categorical features with little discrete values. You want to make sure that automated ML is configured as follows:

1. Missing values must be automatically imputed.
2. Categorical features must be encoded as part of the training task

Make use of the featurization parameter with the ‘auto’ value pair.

Point 31.

Suppose you want to develop a linear regression model using Azure ML Studio. You perform an experiment to assess various algorithms.

Linear Regression is the best algorithm that reduces the variances between actual and predicted values.

Point 32.

Suppose you have been tasked with constructing a ML model that translates language text into a different language text. The ML model must be constructed and trained to learn the sequence.

Convolutional Neural Networks (CNNs) cannot be used.

Point 33.

Suppose you have been tasked with constructing a ML model that translates language text into a different language text. The ML model must be constructed and trained to learn the sequence.

Generative Adversarial Networks (GANs) cannot be used.

Point 34.

Suppose you have been tasked with constructing a ML model that translates language text into a different language text. The ML model must be constructed and trained to learn the sequence.

Recurrent Neural Networks (GANs) can be used.

RNNs are designed to take sequences of text as inputs or return sequences of text as outputs, or both.

Point 35.

Precision and Accuracy are the two-evaluation metrics to evaluate the performance of a binary classification model.

The evaluation metrics available for binary classification models are: Accuracy, Precision, Recall, F1 score, and AUC.

Point 36.

Suppose you are building a binary classification model using Azure ML studio Two-Class Neural Network module. You are preparing to configure the Tune Model Hyperparameters module for the purpose of tuning accuracy for the model. The valid parameters for the Two-Class Neural Network module are:

1. Random number seed
2. The initial learning weights diameter
3. Number of learning iterations.

Point 37.

Suppose you create a binary classification model. You are preparing to carry out a parameter sweep of the model to tune hyperparameters. You have to make sure that the sweep allows for every possible combination of hyperparameters to be iterated. Also, the computing resources needed to carry out the sweep must be reduced.

Consider making use of the Random grid sweep mode.

Point 38.

Suppose you are in the process of constructing a deep convolutional neural network (CNN). The CNN will be used for image classification. But the CNN model is overfitting. To reduce the overfitting and to converge the model for an optimal fit,

Add L1/L2 regularization, and make use of training data augmentation.

Point 39.

Suppose you are planning to make an Azure ML designer to train models and you need to choose a suitable compute type.

Attached compute will not be a suitable choice.

Point 40.

Suppose you are planning to make an Azure ML designer to train models and you need to choose a suitable compute type.

Inference cluster will not be a suitable choice.

Point 41.

Suppose you are planning to make an Azure ML designer to train models and you need to choose a suitable compute type.

Compute cluster will be a suitable choice.

Point 42.

Suppose you are making use of the Azure ML to designer construct an experiment. After dividing a dataset into training and testing sets, you configure the algorithm to be Two-Class Boosted Decision Tree. You are preparing to ascertain the Area Under the Curve (AUC).

Train, Score, Evaluate is a sequential combination of the models required to achieve this goal.

Point 43.

Suppose you created an Azure ML workspace. You must implement dedicated compute for model training in the workspace by using Azure Synapse compute resources. The solution must attach the dedicated compute and start an Azure Synapse session. You need to implement the compute resources. Actions performed in sequence

1. Create an Azure Synapse workspace by using the Azure portal
2. Create an Apache Spark pool by using the Azure portal.
3. Create a linked service by using Azure ML studio.

Point 44.

You are developing a hands-on workshop to introduce Docker for Windows to attendees. You need to ensure that workshop attendees can install Docker on their devices.

Prerequisite components to install Docker on devices include

1. BIOS-enabled virtualization
2. Windows 10 64-bit Professional

Point 45.

Suppose your team is building a data engineering and data science development environment. The environment must support the following requirements:

1. Support Python and Scala
2. Compose data storage, movement, and processing services into automated data pipelines
3. The same tool should be used for the orchestration of data engineering and data science.
4. Support workload isolation and interactive workloads
5. Enable scaling across a cluster of machines.

Build the environment in Azure Databricks and use Azure Data Factory for orchestration.

Point 46.

Suppose you are building an intelligent solution using ML models. The environment must support the following requirements:

1. Data scientist must build notebooks in a cloud environment.
2. Data scientist must use automatic feature engineering and model building in ML pipelines.
3. Notebooks must be deployed to retrain using Spark instances with dynamic worker allocation.
4. Notebooks must be exportable to be version controlled locally.

Four actions you should perform in sequence:

1. Create an Azure HDInsight cluster to include the Apache Spark MLib library.
2. Install Microsoft ML for Apache Spark.
3. Create and execute the zeppelin notebooks on the cluster.
4. When the cluster is ready, export Zeppelin notebooks to a local environment.

Point 47.

Suppose you plan to build a team data science environment. You have the following requirements:

1. Models must be built using Caffe2 or Chainer frameworks.
2. Data Scientists must be able to use a data science environment to build the ML pipelines and train models on their personal devices in both connected and disconnected network environments.

Personal devices must support updating ML pipelines when connected to a network.

A data science environment that can be used is Azure ML Service. Caffe2 and Chainer are supported by DSVM. DSVM integrates with Azure ML.

Point 48.

Suppose you are implementing a ML model to predict stock prices. The model uses a PostgreSQL database and requires GPU processing.

To create a VM that is pre-configured with the required tools, we can create a Deep Learning Virtual Machine (DLVM) Linux edition.

Point 49.

Suppose you are developing a DL model to analyze semi-structured, unstructured, and structured data types. You have the following data available for model building:

1. Video recording of sporting events.
2. Transcripts of radio commentary about events
3. Logs from related social media feeds captured during sporting events.

An environment for creating the model can be Azure Cognitive Services.

Point 50.

You must store data in Azure Blob Storage to support Azure ML. To move data to and from Azure Blob storage use:

1. Azure Storage-Explorer
2. AzCopy
3. Python script

Point 51.

Suppose you are moving a large dataset from Azure ML Studio to a Weka environment.

Convert to ARFF module is used to format the data for the Weka environment. The ARFF data specification for Weka supports multiple ML tasks, including data preprocessing, classification, and feature selection.

Point 52.

TensorFlow can be used to create a speech recognition DL model.

TensorFlow can train and run deep neural networks for handwritten digit classification, image recognition, word embeddings, recurrent neural networks, sequence-to-sequence models for machine translation, natural language processing, and Partial Differential equation-based simulations.

Point 53.

Suppose you plan to use a DLVM to train DL models using Compute Unified Device Architecture (CUDA) computations. You need to configure the DLVM to support CUDA.

A GPU is used to implement it. A DLVM is a pre-configured environment for DL using GPU instances.

Point 54.

Suppose you plan to use a Data Science Virtual Machine (DVSM) with the open-source DL frameworks Caffe2 and PyTorch.

Data Science Virtual Machine for Linux (Ubuntu) framework can be used.

Caffe2 and PyTorch is supported by DSVM for Linux. Microsoft offers Linux editions of the DSVM on Ubuntu 16.04 LTS and CentOS 7.4

Point 55.

You are performing sentiment analysis using a CSV file that includes 12,000 customer reviews written in a short sentence format. You add the CSV file to Azure ML Studio and configure it as the starting point dataset of an experiment. You add the Extract N-Gram features from Text module to the experiment to extract key phrases from the customer review column in the dataset.

You must create a new n-gram dictionary from the customer review text and set the maximum n-gram size to trigrams.

Vocabulary mode – Create

N-grams size – 3

Point 56.

Suppose you are developing a data science workspace that uses an Azure ML service. Azure Container Services can be used to deploy the workspace.

Azure Container Instances can be used as compute target for testing or development. Use of low-scale CPU-based workloads that require less than 48 GB of RAM.

Point 57.

Suppose you have an imbalanced dataset for classification task. Synthetic Minority Oversampling Technique (SMOTE) is used to improve the classification accuracy.

Point 58.

Suppose you have to configure DLVM for windows. You need to recommend tools and frameworks to perform the following

1. Build deep neural network (DNN) models – Vowpal Wabbit
2. Perform interactive data exploration and visualization – PowerBI Desktop

Point 59.

Partition and Sample module is used to divide data into two distinct datasets in Azure ML Studio to build a ML experiment.

Partition and Sample with the Stratified split option outputs multiple datasets, partitioned using the rules specified.

Point 60.

Suppose there is a high degree of missing values in the data and you need to prepare the data for analysis.

Modules to run in sequence for producing the experiment

1. Import Data
2. Clean Missing Data
3. Partition and Sample

Point 61.

Suppose you are retrieving data from a large datastore by using Azure ML Studio. You must create a subset of the data for testing purposes using a random sampling seed based on the system clock. You add the Partition and Sample module to your experiment. You need to select the properties for the module.

Partition or Sample mode – Sampling

Random seed for sampling – 0

Point 62.

Suppose you are creating a ML model, but the dataset has null rows. You need to use the Clean Missing Data module in Azure ML Studio to identify and resolve the null and missing data in the dataset.

Remove entire row parameter can be used. This is useful if the missing value can be considered randomly missing.

Point 63.

Suppose you have to train a model using data in an Azure Storage blob container named finance-data. You need to register the container as a datastore in an Azure ML workspace and ensure that an error will be raised if the container does not exist. The code is like

```

datastore = Datastore.register\_azure\_blob\_container(workspace = ws, datastore\_name = ‘finance\_datastore’, container\_name = ‘finance-data’, account\_name = ‘fintrainingdatastorage’, account\_key = ‘xxxwrt..’, create\_if\_not\_exists = False)

‘’’

Point 64.

Suppose you plan to provision an Azure ML Basic edition workspace for a data science project. The tasks that can be performed in the workspace.

1. Create a compute instance and use it to run code in Jupyter notebooks
2. Create an Azure Kubernetes Service (AKS) inference cluster.
3. Create a tabular dataset that supports versioning.

Point 65.

Suppose a coworker registers a datastore in a ML services workspace by using the code:

Datastore.register\_azure\_blob\_container(workspace=ws, datastore\_name=’demo\_datastore’, container\_name = ‘demo\_datacontainer’, account\_name=’demo\_account’, account\_key=’0A0A-0A0A-0AOA’, create\_if\_no\_exists = True’)

To access the datastore from a notebook

|  |
| --- |
| import azureml.core |
| from azureml.core import Workspace, Datastore |
| ws = workspace.from\_config() |
| datastore = Datastore.get(ws, ‘demo\_datastore’) |

Point 67.

|  |  |  |
| --- | --- | --- |
| Environment name | Compute type | Scenario |
| nb\_server | Compute Instance | Run an Azure ML Designer training pipeline. |
| mlc\_cluster | Machine Learning Compute | Deploying a web service from the Azure ML designer. |

Point 68.

Suppose you created an Azure ML compute target named ComputeOne by using the STANDARD\_D1 VM image. ComputeOne is currently idle and has zero active nodes. You define a python variable named ‘ws’ that references the Azure ML workspace. You run the following python code:

|  |
| --- |
| from azureml.core.compute import ComputeTarget, AmlCompute |
| from azureml.core.compute\_target import ComputeTargetException |
| the\_cluster\_name = ‘ComputeOne’ |
| try: |
| the\_cluster = ComputeTarget(workspace=ws, name=the\_cluster\_name) |
| print(‘Step 1’) |
| except ComputeTargetException: |
| config = AmlCompute.provisioning\_configuration(vm\_size=’STANDARD\_DS12\_v2’, max\_nodes=4) |
| the\_cluster = ComputeTarget.create(ws, the\_cluster\_name, config) |
| print(‘Step2’) |

A new ML compute resource is created with a virtual machine size of STANDARD\_DS12\_v2 and a maximum of four nodes.

Any experiments configured to use ‘the\_cluster’ will run on ComputeOne.

The text Step2 will be printed to the screen.

Point 69.

For developing a deep learning model by TensorFlow using an Azure ML Compute Instance, where CUDA-based model training is required, we can use

1. STANDARD\_NC12 with 2 GPUs
2. STANDARD\_NC24 with 4 GPUs

Point 70.

Suppose you are analysing a raw dataset that requires cleaning, and you must perform transformations and manipulations by using Azure ML Studio.

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

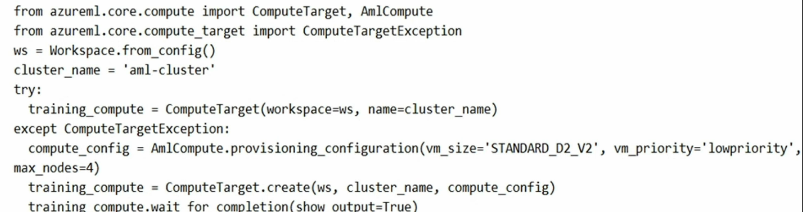
Point 71.

Suppose you are using Azure ML Studio to perform feature engineering on a dataset. You need to normalize values to produce a feature column grouped into bins.

Applying an Entropy Minimum Description Length (MDL) binning mode will not work.

Point 72.

Suppose you are preparing to use the Azure ML SDK to run an experiment and need to create compute. You run the following code:



|  |  |
| --- | --- |
| If a compute cluster named aml-cluster already exists in the workspace, it will be deleted and replaced | No |
| The wait\_for\_completion() method will not return until the aml-cluster compute has four active nodes. | Yes |
| If the code creates a new aml-cluster compute target, it may be pre-empted due to capacity constraints. | Yes |
| The aml-cluster compute target is deleted from the workspace after the training experiment completes | No |

Point 73.

Suppose you are using Azure ML Studio to normalize values to produce an output column into bins to predict a target column.

Applying a Quantiles normalisation with Quantile Index normalization will work.

Point 74.

Suppose you are using Azure ML Studio; one class has a much smaller number of observations than the other classes in the training set. You need to select an appropriate data sampling strategy to compensate for the class imbalance.

Applying a Scale and Reduce sampling mode will not work.

Instead use SMOTE sampling mode.

Point 75.

Suppose you are analyzing a dataset by Azure ML Studio;and need ot generate a statistical summary that contains the p-value and the unique count for feature column.

Export Count Table and Summarise Data can be used for this task.

Point 76.

Suppose you are analysing a numerical dataset which contains missing values in several columns. You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set. You need to analyze a full dataset to include all values.

Last Observation Carried Forward (LOCF) method to impute the missing data points cannot be used.

Instead use the Multiple Imputation by Chained Equations (MICE) method.

Point 77.

Suppose you are creating a ML model in python. The provided dataset contains several numerical columns and one text column. The text column represents a product’s category. The product category will always be one of the following:

Bikes, Cars, Vans, Boats

You are building a regression model using the scikit-learn python package.

You need to transform the text data to be compatible with the scikit-learn python package. The code for this:

|  |
| --- |
| from sklearn import linear\_model |
| import pandas as df |
| dataset = df.read\_csv(“data\\ProductSales.csv”) |
| ProductCategoryMapping = {“Bikes”: 1, “Cars”: 2, “Boats”: 3, “Vans”: 4} |
| dataset[‘ProductCategoryMapping’] = dataset[‘ProductCategory’].transpose[ProductCategoryMapping] |
| reqr = linear\_model.LinearRegression() |

Point 78.

Suppose you plan to deliver a hands-on workshop to several students. The workshop will focus on creating data visualisations using python. Each student will use a device that has internet access. Student devices are not configured for python development. Students do not have administrator access to install software on their devices. Azure subscriptions are not available for students. You need to ensure that students can run python-based data visualization code.

Azure Notebooks tool can be used.

Point 79.

Suppose you are analysing a numerical dataset which contains missing values in several columns. You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set. You need to analyze a full dataset to include all values.

Replacing each missing value with Multiple Imputation by Chained Equations (MICE) method will work.

With a MICE method, each variable with missing data is modelled conditionally using the other variables in the data before filling in the missing values.

Point 80.

Suppose you are analysing a numerical dataset which contains missing values in several columns. You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set. You need to analyze a full dataset to include all values.

Removing the entire column that contains the missing data point will not work.

Point 81.

Suppose you are creating a new experiment in Azure ML Studio. You have a small dataset that has missing values in many columns. The data does not require the application of predictors for each column. You plan to use the Clean Missing Data.

Replace using Probabilistic PCA method will work.

MICE method requires the application of predictors for each column.

Point 82.

Split Data module can be used to divide data into two distinct datasets.

Point 83.

Suppose you want to track the health and migration of birds. You create a multi-class image classification DL model that uses a set of labelled bird photographs collected by experts. You have 100,000 photographs of birds. All photographs use the JPG format and are stored in an Azure blob container in an Azure subscription. You need to access the bird photograph files in the Azure blob container from the Azure ML service workspace that will be used for DL model training. You must minimize data movement.

Register the Azure blob storage containing the bird photographs as a datastore in Azure ML service.

Point 84.

Suppose you are analysing a numerical dataset which contains missing values in several columns. You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set. You need to analyze a full dataset to include all values.

Calculating the column median value and use the median value as the replacement for any missing value in the column will work.

Point 86.

You are a data scientist using Azure ML Studio. You need to normalize values to produce an output column into bins to predict a target column.

Applying an Equal Width with Custom Start and Stop binning mode cannot solve this.

Use the Entropy MDL binning mode which has a target column.

Point 87.

You are a data scientist using Azure ML Studio. You need to normalize values to produce an output column into bins to predict a target column.

Applying a Quantiles binning mode with a PQuantile normalization can solve this.

Point 88.

Suppose you are evaluating a Python NumPy array that contains six data points defined as follows: data = [10, 20, 30, 40, 50, 60]

You must generate the following output by using the k-fold algorithm implantation in the python scikit-learn ML library:

train: [10, 40, 50, 60], test: [20, 30]

train: [20, 30, 40, 60], test: [10, 50]

train: [10, 20, 30, 50], test: [40, 60]

You need to implement a cross-validation to generate the output.

|  |
| --- |
| from numpy import array |
| from sklearn.model\_selection import k-fold |
| data = array([10, 20, 30, 40, 50, 60]) |
| kfold = Kfold(n\_splits = 3, shuffle = True, random\_State = 1) |
| for train, test in kFold, split(data): |
| print(‘train: %s, test: %s’ %(data[train], data[test]) |

Print 89.

Suppose you have a time series dataset in Azure ML Studio. You need to split your dataset into training and testing subsets by using the Split Data module.

Relative Expression Split mode can be used.

Print 90.

Suppose you are preparing to build a DL CNN for image classification. You create a script to train the model using CUDA devices.

You must submit an experiment that runs this script in the Azure ML workspace. The compute resources to be used for running the code to submit the experiment, and for running the script in order to minimize model training time

|  |  |
| --- | --- |
| Resources type | Option |
| Run code to submit the experiment | the ds-workstation compute instance |
| Run the training script | the gpu-cluster compute target |

Point 91.

Suppose you create an Azure ML workspace. You are preparing a local python environment on a laptop computer. You want to use the laptop to connect to the workspace and run experiments. You create the following ‘config.json’ file.

{

“workspace\_name”: “ml-workspace”

}

You must use the Azure ML SDK to interact with data and experiments in the workspace. You need to configure the config.json file to connect to the workspace from the python environment.

Additional parameters that must be added to the config.json file in order to connect to the workspace

1. resource\_group
2. subscription\_id

Point 92.

Suppose you are performing a classification task in Azure ML. You must prepare balanced testing and training samples based on a provided data set. You need to split the data with a 0.75:0.25 ratio

|  |  |
| --- | --- |
| Parameter | Value |
| Splitting mode | Split rows |
| Fraction of rows in the first output dataset | 0.75 |
| Randomized split | True |
| Stratified split | False |

Point 93.

Suppose you create an Azure ML compute resource to train models. The compute resource is configured as follows:

Minimum nodes: 2, Maximum nodes: 4

You must decrease the minimum number of nodes and increase the maximum number of nodes to the following values:

Minimum nodes: 0, Maximum nodes: 8

To configure the compute resources, you can use

1. Azure ML studio
2. Run the update method of the AmlCompute class in the python SDK
3. Azure portal

Point 94.

Suppose you have a dataset that contains 2000 rows. You are building a ML classification model by using Azure Learning Studio. You add a Partition and Sample module to the experiment. You need to configure the module. You must meet the following requirments:

1. Divide the data into subsets
2. Assign the rows into folds using a round-robin method.
3. Allow rows in the dataset to be reused.

Choose

1. Assign to Folds
2. Use replacement in the partitioning

Point 95.

Suppose you have created a new Azure subscription. No resources are provisioned in the subscription. You need to create an Azure ML workspace.

Three ways in which it is possible

1. Run python code that uses the Azure ML SDK library and calls the workspace.create method with name, subscription\_id, and resource\_group parameters.
2. Navigate to Azure ML studio and create a workspace.
3. Use the Azure Command Line Interface (CLI) with Azure ML extension to call the az group, create function with –name and –location parameters

Point 96.

Suppose you create an Azure ML workspace and set up a development environment. You plan to train a deep neural network (DNN) by using the tensorflow framework and by using estimators to submit training scripts. You must optimize computation speed for training runs. You need to choose the appropriate estimator to use as well as the appropriate training compute target configuration.

|  |  |
| --- | --- |
| Parameter | Value |
| Estimator | Tensorflow |
| Training compute | 12 vCPU, 112 GB memory, 680 GB SSD, 2 GPU, 24 GB GPU memory |

Point 97.

Suppose you have an Azure ML workspace named workspace1 that is accessible from a public endpoint. The workspace contains an Azure Blob storage datastore named store1 that represents a blob container in an Azure storage account named account1. You configure workspace1 and account1 to be accessible by using private endpoints in the same virtual network.

You must be able to access the contents of store1 by using the Azure ML SDK for python. You must be able to preview the contents of store1 by using Azure ML studio.

|  |  |
| --- | --- |
| Requirement | Action |
| Access the contents of store1 by using the Azure ML SDK for python | Regenerate the keys of account1 |
| Preview the contents of store1 by using Azure ML studio | Update authentication for store1 |

Point 98.

Suppose you are using an Azure ML workspace. You set up an environment for model testing and an environment for production. The compute target for testing must minimize cost and deployment efforts. The compute target for production must provide fast response time, autoscaling of the deployed service, and support real-time inferencing. You need to configure compute targets for model testing and production.

|  |  |
| --- | --- |
| Environment | Compute Target |
| Testing | Local web service |
| Production | Azure Kubernetes Service (AKS) |

Point 99.

Suppose you are using a git repository to track work in an Azure ML workspace. You need to authenticate a Git account by using SSH. Which three actions should you perform in sequence.

1. Generate a public/private key pair
2. Add the public key to the Git account
3. Clone the Git repository by using an SSH repository URL.

Point 100.

Suppose you use Azure ML to train a model based on a dataset named dataset1. You define a dataset monitor and create a dataset named dataset2 that contains new data. You need to compare dataset1 and dataset2 by using the Azure ML SDK for python.

Backfill method of Data Drift Detector class could be used.

Point 101.

Suppose you use an Azure ML workspace. You have a trained model that must be deployed as a web service. Users must authenticate by using Azure Active Directory.

Deploy the model to Azure Kubernetes Service (AKS). During deployment, set the token\_auth\_enabled parameter of the target configuration object to true.

To control token authentication, use the toke\_auth\_enabled parameter when you create or update a deployment. Token authentication is diabled by default when you deploy to AKS.

Point 102.

Notes

1. When you use Azure ML, Azure Databricks, or Azure Synapse Analytics for model training, there are three common options for storing data:

Azure Blob Storage – unstructured data

Azure Data Lake Storage (Gen 2) – advanced version of blob storage

Azure SQL Database