DUMPSARENA

Designing and Implementing a Data Science Solution on Azure

Microsoft DP-100

Version Demo

Total Demo Questions: 20

Total Premium Questions: 502

Buy Premium PDF

https://dumpsarena.com

sales@dumpsarena.com

Topic Break Down

Topic	No. of Questions
Topic 2, New Update	208
Topic 3, Case Study 1	8
Topic 4, Case Study 2	12
Topic 5, Mixed Questions	274
Total	502

QUESTION NO: 1 - (HOTSPOT)

HOTSPOT

You create an Azure Machine Learning compute target named ComputeOne by using the STANDARD_D1 virtual machine image.

You define a Python variable named was that references the Azure Machine Learning 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('Step1')
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')
```

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

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

	Yes	No
A new machine learning 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 **Step1** will be printed to the screen.	○	○

ANSWER:

Answer Area

	Yes	No
A new machine learning 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 **Step1** will be printed to the screen.	○	○

Explanation:

Box 1:Yes

ComputeTargetException class: An exception related to failures when creating, interacting with, or configuring a compute target. This exception is commonly raised for failures attaching a compute target, missing headers, and unsupported configuration values.

Create(workspace, name, provisioning_configuration)

Provision a Compute object by specifying a compute type and related configuration.

This method creates a new compute target rather than attaching an existing one.

Box 2: Yes

Box 3: No The line before print('Step1') will fail.

Reference:

https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.computetarget

QUESTION NO: 2

You are building a regression model for estimating the number of calls during an event.

You need to determine whether the feature values achieve the conditions to build a Poisson regression model.

Which two conditions must the feature set contain? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

A. The label data must be a negative value.

B. The label data must be whole numbers.

C. The label data must be non-discrete.

D. The label data must be a positive value.

E. The label data can be positive or negative.

ANSWER: B D

Explanation:

Poisson regression is intended for use in regression models that are used to predict numeric values, typically counts. Therefore, you should use this module to create your regression model only if the values you are trying to predict fit the following conditions:

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

Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/poisson-regression

QUESTION NO: 3

You are solving a classification task.

The dataset is imbalanced.

You need to select an Azure Machine Learning Studio module to improve the classification accuracy.

Which module should you use?

A. Permutation Feature Importance

B. Filter Based Feature Selection

C. Fisher Linear Discriminant Analysis

D. Synthetic Minority Oversampling Technique (SMOTE)

ANSWER: D

Explanation:

Use the SMOTE module in Azure Machine Learning Studio (classic) to increase the number of underrepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.

You connect the SMOTE module to a dataset that is imbalanced. There are many reasons why a dataset might be imbalanced: the category you are targeting might be very rare in the population, or the data might simply be difficult to collect. Typically, you use SMOTE when the class you want to analyze is under-represented.

Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote

QUESTION NO: 4 - (SIMULATION)

You create an Azure Machine Learning dataset containing automobile price data. The dataset includes 10.000 rows and 10 columns. You use the Azure Machine Learning designer to transform the dataset by using an Execute Python Script component and custom code.

The code must combine three columns to create a new column.

You need to configure the code function.

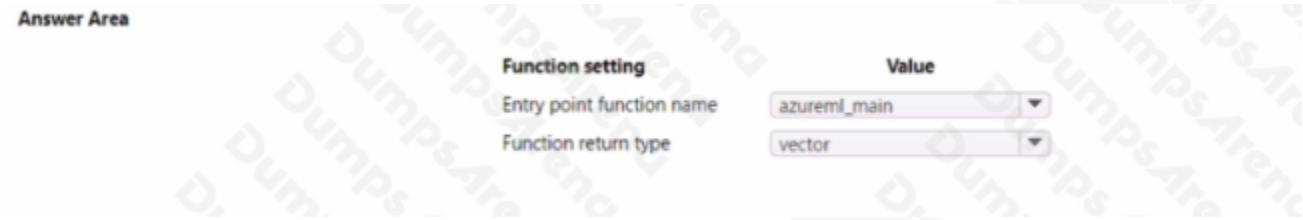
Which configurations should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

ANSWER: seetheanswerbelow.

Explanation:

See below image



QUESTION NO: 5

You need to visually identify whether outliers exist in the Age column and quantify the outliers before the outliers are removed.

Which three Azure Machine Learning Studio modules should you use? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

A. Create Scatterplot

B. Summarize Data

C. Clip Values

D. Replace Discrete Values

E. Build Counting Transform

ANSWER: A B C

Explanation:

B: To have a global view, the summarize data module can be used. Add the module and connect it to the data set that needs to be visualized. A: One way to quickly identify Outliers visually is to create scatter plots.

C: The easiest way to treat the outliers in Azure ML is to use the Clip Values module. It can identify and optionally replace data values that are above or below a specified threshold.

You can use the Clip Values module in Azure Machine Learning Studio, to identify and optionally replace data values that are above or below a specified threshold. This is useful when you want to remove outliers or replace them with a mean, a constant, or other substitute value.

Reference:

https://blogs.msdn.microsoft.com/azuredev/2017/05/27/data-cleansing-tools-in-azure-machine-learning/
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clip-values

QUESTION NO: 6

You use the Azure Machine Learning SDK to run a training experiment that trains a classification model and calculates its accuracy metric.

The model will be retrained each month as new data is available.

You must register the model for use in a batch inference pipeline.

You need to register the model and ensure that the models created by subsequent retraining experiments are registered only if their accuracy is higher than the currently registered model.

What are two possible ways to achieve this goal? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

A. Specify a different name for the model each time you register it.

B. Register the model with the same name each time regardless of accuracy, and always use the latest version of the model in the batch inferencing pipeline.

C. Specify the model framework version when registering the model, and only register subsequent models if this value is higher.

D. Specify a property named accuracy with the accuracy metric as a value when registering the model, and only register subsequent models if their accuracy is higher than the accuracy property value of the currently registered model.

E. Specify a tag named accuracy with the accuracy metric as a value when registering the model, and only register subsequent models if their accuracy is higher than the accuracy tag value of the currently registered model.

ANSWER: C E

Explanation:

E: Using tags, you can track useful information such as the name and version of the machine learning library used to train the model. Note that tags must be alphanumeric.

Reference: https://notebooks.azure.com/xavierheriat/projects/azureml-getting-started/html/how-to-use-azureml/deployment/register-model-create-image-deploy-service/register-model-create-image-deployservice.ipynb

QUESTION NO: 7

You are evaluating a completed binary classification machine learning model.

You need to use the precision as the evaluation metric.

Which visualization should you use?

A. violin plot

B. Gradient descent

C. Scatter plot

D. Receiver Operating Characteristic (ROC) curve

ANSWER: D

Explanation:

Receiver operating characteristic (or ROC) is a plot of the correctly classified labels vs. the incorrectly classified labels for a particular model.

Incorrect Answers:

A: A violin plot is a visual that traditionally combines a box plot and a kernel density plot.

B: Gradient descent is a first-order iterative optimization algorithm for finding the minimum of a function. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient (or approximate gradient) of the function at the current point.

C: A scatter plot graphs the actual values in your data against the values predicted by the model. The scatter plot displays the actual values along the X-axis, and displays the predicted values along the Yaxis. It also displays a line that illustrates the perfect prediction, where the predicted value exactly matches the actual value.

Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/how-to-understand-automated-ml#confusion-matrix

QUESTION NO: 8

You create an Azure Machine Learning compute resource to train models. The compute resource is configured as follows:

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

You need to reconfigure the compute resource.

What are three possible ways to achieve this goal? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

A. Use the Azure Machine Learning studio.

B. Run the update method of the AmlCompute class in the Python SDK.

C. Use the Azure portal.

D. Use the Azure Machine Learning designer.

E. Run the refresh_state() method of the BatchCompute class in the Python SDK

ANSWER: A B C

Explanation:

Reference:

https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute(class)

QUESTION NO: 9

You need to implement a model development strategy to determine a user's tendency to respond to an ad.

Which technique should you use?

A. Use a Relative Expression Split module to partition the data based on centroid distance.

B. Use a Relative Expression Split module to partition the data based on distance travelled to the event.

C. Use a Split Rows module to partition the data based on distance travelled to the event.

D. Use a Split Rows module to partition the data based on centroid distance.

ANSWER: A

Explanation:

Split Data partitions the rows of a dataset into two distinct sets.

The Relative Expression Split option in the Split Data module of Azure Machine Learning Studio is helpful when you need to divide a dataset into training and testing datasets using a numerical expression.

Relative Expression Split: Use this option whenever you want to apply a condition to a number column. The number could be a date/time field, a column containing age or dollar amounts, or even a percentage. For example, you might want to divide your data set depending on the cost of the items, group people by age ranges, or separate data by a calendar date.

Scenario:

Local market segmentation models will be applied before determining a user's propensity to respond to an advertisement. The distribution of features across training and production data are not consistent

Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data

QUESTION NO: 10

Note: This question is part of 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.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are using Azure Machine Learning to run an experiment that trains a classification model.

You want to use Hyperdrive to find parameters that optimize the AUC metric for the model. You configure a HyperDriveConfig for the experiment by running the following code:

```
hyperdrive = HyperDriveConfig(estimator=your_estimator,
 hyperparameter_sampling=your_params,
 policy=policy,
 primary_metric_name='AUC',
 primary_metric_goal=PrimaryMetricGoal.MAXIMIZE,
 max_total_runs=6,
 max_concurrent_runs=4)
```

You plan to use this configuration to run a script that trains a random forest model and then tests it with validation data. The label values for the validation data are stored in a variable named y_test variable, and the predicted probabilities from the model are stored in a variable named y_predicted.

You need to add logging to the script to allow Hyperdrive to optimize hyperparameters for the AUC metric.

Solution: Run the following code:

```
import numpy as np
from sklearn.metrics import roc_auc_score
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
print(np.float(auc))
```

Does the solution meet the goal?

A. Yes

B. No

ANSWER: B

Explanation:

Use a solution with logging.info(message) instead.

Note: Python printing/logging example: logging.info(message)

Destination: Driver logs, Azure Machine Learning designer

Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines

QUESTION NO: 11 - (DRAG DROP)

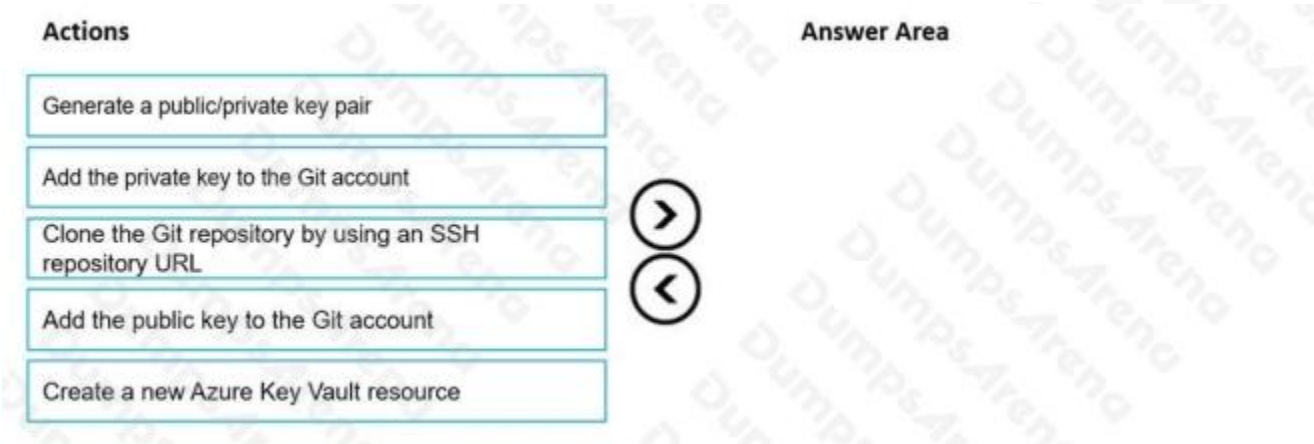
DRAG DROP

You are using a Git repository to track work in an Azure Machine Learning workspace.

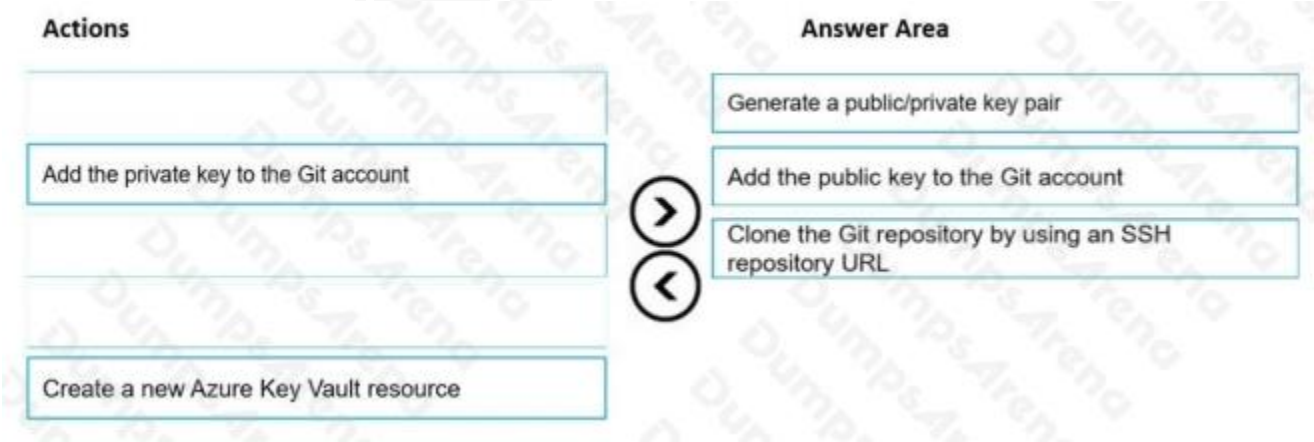
You need to authenticate a Git account by using SSH.

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

Select and Place:



ANSWER:



Explanation:

Authenticate your Git Account with SSH:

Step 1: Generating a public/private key pair

Generate a new SSH key

1. Open the terminal window in the Azure Machine Learning Notebook Tab.

2. Paste the text below, substituting in your email address.

ssh-keygen -t rsa -b 4096 -C "your_email@example.com" This creates a new ssh key, using the provided email as a label.

> Generating public/private rsa key pair.

Step 2: Add the public key to the Git Account

In your terminal window, copy the contents of your public key file.

Step 3: Clone the Git repository by using an SSH repository URL 1. Copy the SSH Git clone URL from the Git repo.

2. Paste the url into the git clone command below, to use your SSH Git repo URL. This will look something like:

git clone git@example.com:GitUser/azureml-example.git Cloning into 'azureml-example'.

Reference: https://docs.microsoft.com/en-us/azure/machine-learning/concept-train-model-git-integration

QUESTION NO: 12 - (SIMULATION)

You are implementing hyperparameter tuning for a model training from a notebook. The notebook is in an Azure Machine Learning workspace. You add code that imports all relevant Python libraries.

You must configure Bayesian sampling over the search space for the num_hidden_layers and batch_size hyperparameters.

You need to complete the following Python code to configure Bayesian sampling.

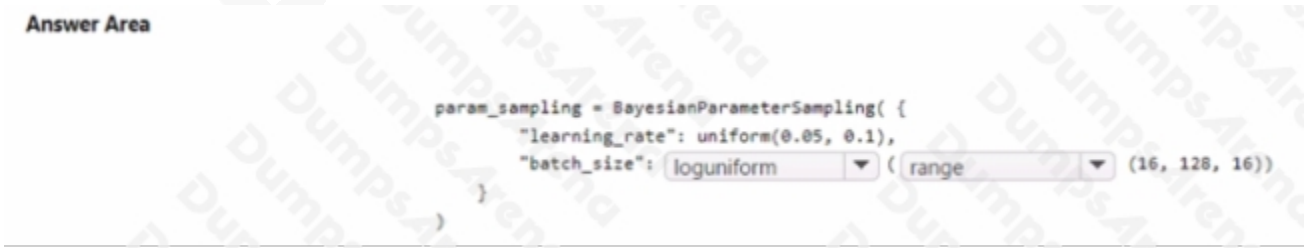
Which code segments should you use? To answer, select the appropriate options in the answer area

NOTE: Each correct selection is worth one point.

ANSWER: seetheanswerbelow.

Explanation:

See below image



QUESTION NO: 13 - (HOTSPOT)

HOTSPOT

You are using the Azure Machine Learning Service to automate hyperparameter exploration of your neural network classification model.

You must define the hyperparameter space to automatically tune hyperparameters using random sampling according to following requirements:

- The learning rate must be selected from a normal distribution with a mean value of 10 and a standard deviation of 3.
- Batch size must be 16, 32 and 64.
- Keep probability must be a value selected from a uniform distribution between the range of 0.05 and 0.1.

You need to use the param_sampling method of the Python API for the Azure Machine Learning Service.

How should you complete the code segment? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

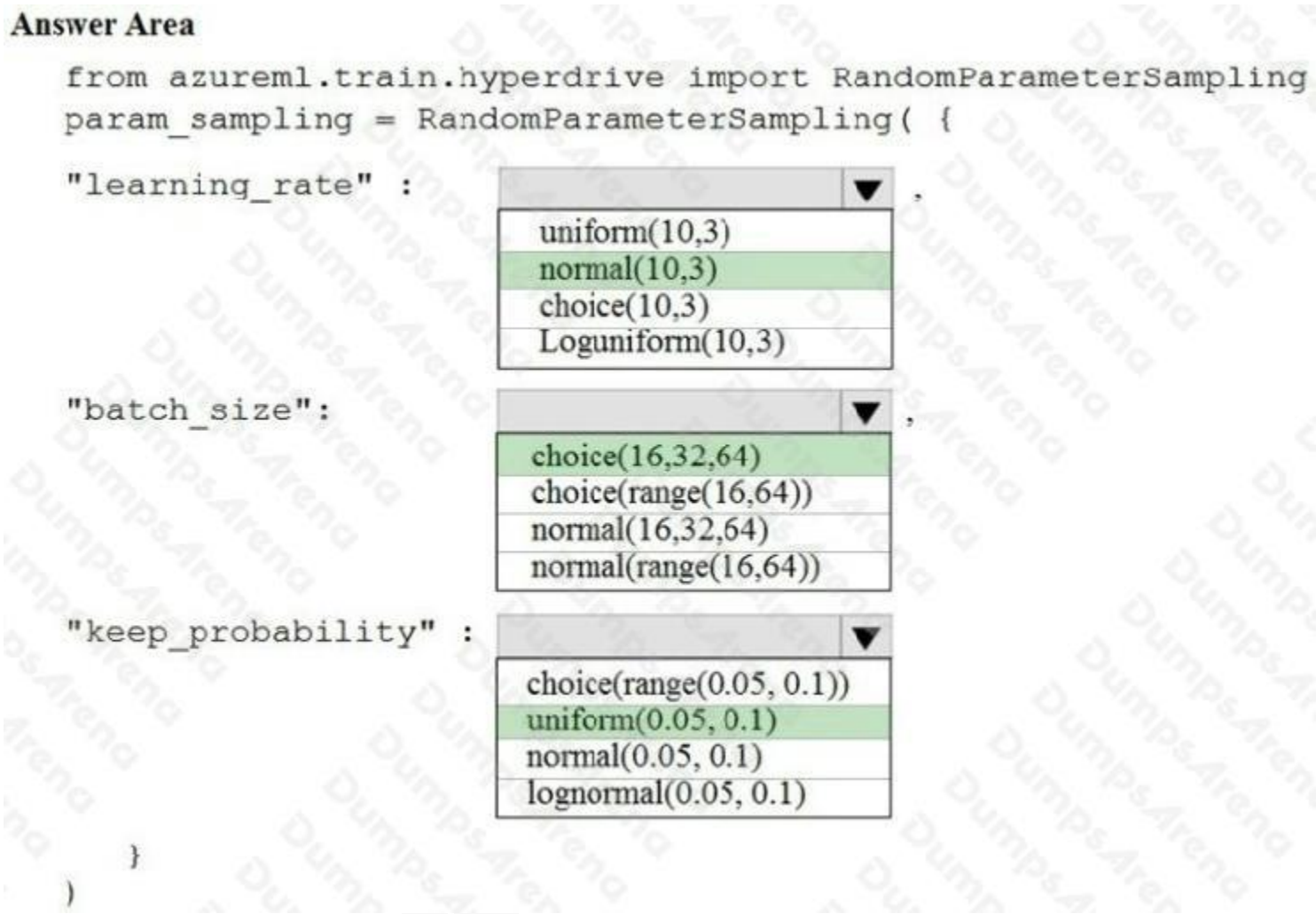
Hot Area:

Answer Area

```
from azureml.train.hyperdrive import RandomParameterSampling
param_sampling = RandomParameterSampling( {
"learning_rate" :  [                    ] ,
                   [ uniform(10,3)      ]
                   [ normal(10,3)       ]
                   [ choice(10,3)       ]
                   [ Loguniform(10,3)   ]
"batch_size":      [                    ] ,
                   [ choice(16,32,64)   ]
                   [ choice(range(16,64)) ]
                   [ normal(16,32,64)   ]
                   [ normal(range(16,64)) ]
"keep_probability" : [                  ]
                   [ choice(range(0.05, 0.1)) ]
                   [ uniform(0.05, 0.1) ]
                   [ normal(0.05, 0.1)  ]
                   [ lognormal(0.05, 0.1) ]
    }
)
```

ANSWER:

Answer Area



Explanation:

In random sampling, hyperparameter values are randomly selected from the defined search space. Random sampling allows the search space to include both discrete and continuous hyperparameters.

Example:

from azureml.train.hyperdrive import RandomParameterSampling param_sampling = RandomParameterSampling({

"learning_rate": normal(10, 3),

"keep_probability": uniform(0.05, 0.1),

"batch_size": choice(16, 32, 64)

}

Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-tune-hyperparameters

QUESTION NO: 14

You use Azure Machine Learning designer to create a real-time service endpoint. You have a single Azure Machine Learning service compute resource.

You train the model and prepare the real-time pipeline for deployment.

You need to publish the inference pipeline as a web service.

Which compute type should you use?

A. a new Machine Learning Compute resource

B. Azure Kubernetes Services

C. HDInsight

D. the existing Machine Learning Compute resource

E. Azure Databricks

ANSWER: B

Explanation:

Azure Kubernetes Service (AKS) can be used real-time inference.

Reference:

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

QUESTION NO: 15 - (DRAG DROP)

DRAG DROP

You plan to explore demographic data for home ownership in various cities. The data is in a CSV file with the following format:

age,city,income,home_owner

21,Chicago,50000,0

35,Seattle,120000,1

23,Seattle,65000,0

45,Seattle,130000,1 18,Chicago,48000,0

You need to run an experiment in your Azure Machine Learning workspace to explore the data and log the results. The experiment must log the following information:

- the number of observations in the dataset
- a box plot of income by home_owner
- a dictionary containing the city names and the average income for each city

You need to use the appropriate logging methods of the experiment's run object to log the required information.

How should you complete the code? To answer, drag the appropriate code segments to the correct locations. Each code segment may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Select and Place:

Code segments

- log
- log_list
- log_row
- log_table
- log_image

Answer Area

```
from azureml.core import Experiment, Run
import pandas as pd
import matplotlib.pyplot as plt
# Create an Azure ML experiment in workspace
experiment = Experiment(workspace = ws, name = "demo-experiment")
# Start logging data from the experiment
run = experiment.start_logging()
# load the dataset
data = pd.read_csv('research/demographics.csv')
# Log the number of observations
row_count = (len(data))
run. Segment ("observations", row_count)
# Log box plot for income by home_owner
fig = plt.figure(figsize=(9, 6))
ax = fig.gca()
data.boxplot(column = 'income', by = "home_owner", ax = ax)
ax.set_title('income by home_owner')
ax.set_ylabel('income')
run. Segment (name = 'income_by_home_owner', plot = fig)
# Create a dataframe of mean income per city
mean_inc_df = data.groupby('city')['income'].agg(np.mean).to_frame().reset_index()
# Convert to a dictionary
mean_inc_dict = mean_inc_df.to_dict('dict')
# Log city names and average income dictionary
run. Segment (name="mean_income_by_city", value= mean_inc_dict)
# Complete tracking and get link to details
run.complete()
```

ANSWER:

Code segments

log_list

log_row

Answer Area

```
from azureml.core import Experiment, Run
import pandas as pd
import matplotlib.pyplot as plt
# Create an Azure ML experiment in workspace
experiment = Experiment(workspace = ws, name = "demo-experiment")
# Start logging data from the experiment
run = experiment.start_logging()
# load the dataset
data = pd.read_csv('research/demographics.csv')
# Log the number of observations
row_count = (len(data))
run. log          ("observations", row_count)
# Log box plot for income by home_owner
fig = plt.figure(figsize=(9, 6))
ax = fig.gca()
data.boxplot(column = 'income', by = "home_owner", ax = ax)
ax.set_title('income by home_owner')
ax.set_ylabel('income')
run. log_image    (name = 'income_by_home_owner', plot = fig)
# Create a dataframe of mean income per city
mean_inc_df = data.groupby('city')['income'].agg(np.mean).to_frame().reset_index()
# Convert to a dictionary
mean_inc_dict = mean_inc_df.to_dict('dict')
# Log city names and average income dictionary
run. log_table    (name="mean_income_by_city", value= mean_inc_dict)
# Complete tracking and get link to details
run.complete()
```

Explanation:

Box 1: log

The number of observations in the dataset.

run.log(name, value, description='')

Scalar values: Log a numerical or string value to the run with the given name. Logging a metric to a run causes that metric to be stored in the run record in the experiment. You can log the same metric multiple times within a run, the result being considered a vector of that metric. Example: run.log("accuracy", 0.95)

Box 2: log_image

A box plot of income by home_owner. log_image Log an image to the run record. Use log_image to log a .PNG image file or a matplotlib plot to the run. These images will be visible and comparable in the run record. Example: run.log_image("ROC", plot=plt)

Box 3: log_table

A dictionary containing the city names and the average income for each city. log_table: Log a dictionary object to the run with the given name.

QUESTION NO: 16

You have been tasked with ascertaining if two sets of data differ considerably. You will make use of Azure Machine Learning Studio to complete your task.

You plan to perform a paired t-test.

Which of the following are conditions that must apply to use a paired t-test? (Choose all that apply.)

A. All scores are independent from each other.

B. You have a matched pairs of scores.

C. The sampling distribution of d is normal.

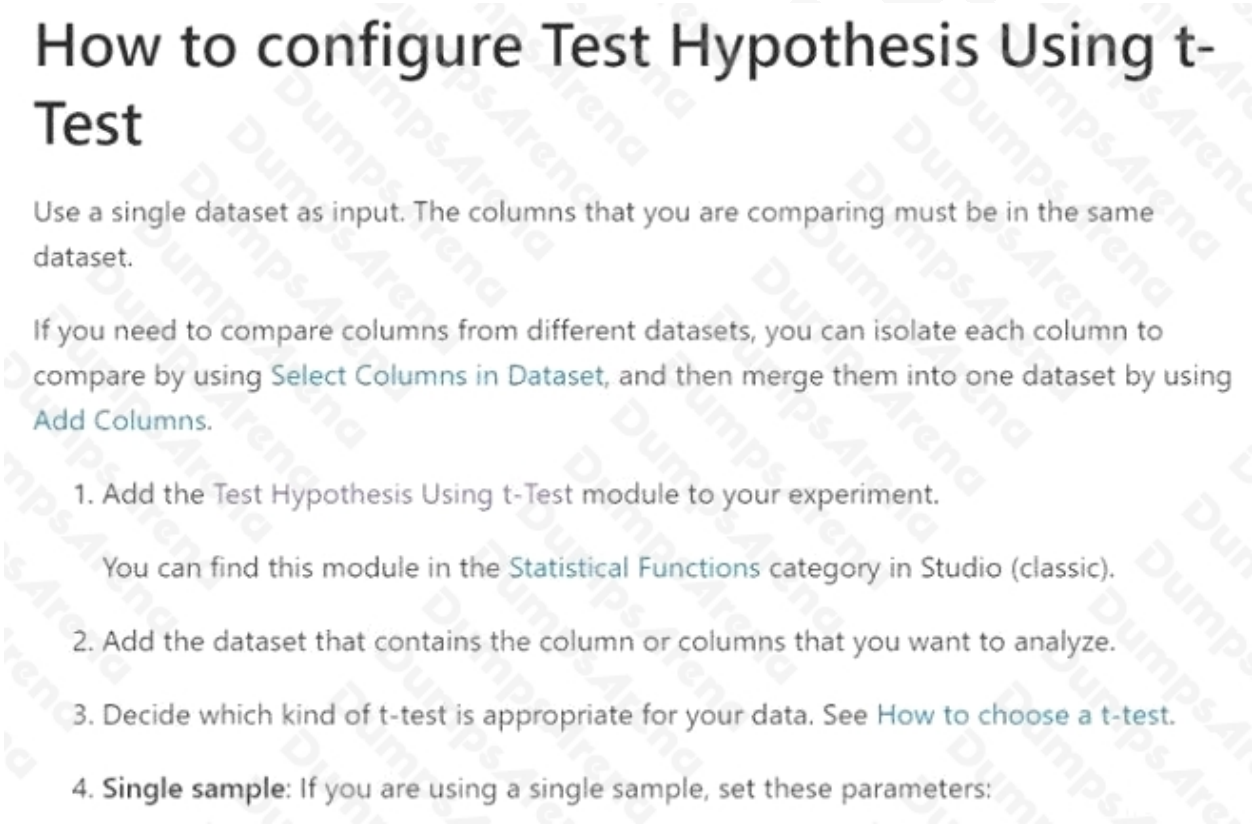
D. The sampling distribution of x1- x2 is normal.

ANSWER: B C

Explanation:

Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/test-hypothesis-using-t-test



How to configure Test Hypothesis Using t-Test

Use a single dataset as input. The columns that you are comparing must be in the same dataset.

If you need to compare columns from different datasets, you can isolate each column to compare by using Select Columns in Dataset, and then merge them into one dataset by using Add Columns.

1. Add the Test Hypothesis Using t-Test module to your experiment.

   You can find this module in the Statistical Functions category in Studio (classic).

2. Add the dataset that contains the column or columns that you want to analyze.

3. Decide which kind of t-test is appropriate for your data. See How to choose a t-test.

4. **Single sample**: If you are using a single sample, set these parameters:

QUESTION NO: 17

Note: This question is part of 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.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

An IT department creates the following Azure resource groups and resources:

Resource group	Resources
ml_resources	• an Azure Machine Learning workspace named amlworkspace • an Azure Storage account named amlworkspace12345 • an Application Insights instance named amlworkspace54321 • an Azure Key Vault named amlworkspace67890 • an Azure Container Registry named amlworkspace09876
general_compute	A virtual machine named mlvm with the following configuration: • Operating system: Ubuntu Linux • Software installed: Python 3.6 and Jupyter Notebooks • Size: NC6 (6 vCPUs, 1 vGPU, 56 Gb RAM)

The IT department creates an Azure Kubernetes Service (AKS)-based inference compute target named aks-cluster in the Azure Machine Learning workspace.

You have a Microsoft Surface Book computer with a GPU. Python 3.6 and Visual Studio Code are installed.

You need to run a script that trains a deep neural network (DNN) model and logs the loss and accuracy metrics.

Solution: Attach the mlvm virtual machine as a compute target in the Azure Machine Learning workspace. Install the Azure ML SDK on the Surface Book and run Python code to connect to the workspace. Run the training script as an experiment on the mlvm remote compute resource.

A. Yes

B. No

ANSWER: A

Explanation:

Use the VM as a compute target.

Note: A compute target is a designated compute resource/environment where you run your training script or host your service deployment. This location may be your local machine or a cloud-based compute resource.

Reference:

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

QUESTION NO: 18

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.

Which two prerequisite components should attendees install on the devices? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

A. Microsoft Hardware-Assisted Virtualization Detection Tool

B. Kitematic

C. BIOS-enabled virtualization

D. VirtualBox

E. Windows 10 64-bit Professional

ANSWER: C E

Explanation:

C: Make sure your Windows system supports Hardware Virtualization Technology and that virtualization is enabled. Ensure that hardware virtualization support is turned on in the BIOS settings. For example:

E: To run Docker, your machine must have a 64-bit operating system running Windows 7 or higher.

Reference: https://docs.docker.com/toolbox/toolbox_install_windows/
https://blogs.technet.microsoft.com/canitpro/2015/09/08/step-by-step-enabling-hyper-v-for-use-on-windows-10/

QUESTION NO: 19

You create a binary classification model.

You need to evaluate the model performance.

Which two metrics can you use? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

A. relative absolute error

B. precision

C. accuracy

D. mean absolute error

E. coefficient of determination

ANSWER: B C

Explanation:

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

Note: A very natural question is: 'Out of the individuals whom the model, how many were classified correctly (TP)?'

This question can be answered by looking at the Precision of the model, which is the proportion of positives that are classified correctly.

References:

https://docs.microsoft.com/en-us/azure/machine-learning/studio/evaluate-model-performance

QUESTION NO: 20

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.

You are in the process of creating a machine learning model. Your dataset includes rows with null and missing values.

You plan to make use of the Clean Missing Data module in Azure Machine Learning Studio to detect and fix the null and missing values in the dataset.

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

Will the requirements be satisfied?

A. Yes

B. No

ANSWER: A

Explanation:

Remove entire row: Completely removes any row in the dataset that has one or more missing values. This is useful if the missing value can be considered randomly missing.

Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data