



Microsoft

Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure

About ExamBible

Your Partner of IT Exam

Found in 1998

ExamBible is a company specialized on providing high quality IT exam practice study materials, especially Cisco CCNA, CCDA, CCNP, CCIE, Checkpoint CCSE, CompTIA A+, Network+ certification practice exams and so on. We guarantee that the candidates will not only pass any IT exam at the first attempt but also get profound understanding about the certificates they have got. There are so many alike companies in this industry, however, ExamBible has its unique advantages that other companies could not achieve.

Our Advances

* 99.9% Uptime

All examinations will be up to date.

* 24/7 Quality Support

We will provide service round the clock.

* 100% Pass Rate

Our guarantee that you will pass the exam.

* Unique Gurantee

If you do not pass the exam at the first time, we will not only arrange FULL REFUND for you, but also provide you another exam of your claim, ABSOLUTELY FREE!

NEW QUESTION 1

- (Exam Topic 3)

HOTSPOT

You create a script for training a machine learning model in Azure Machine Learning service. You create an estimator by running the following code:

```
from azureml.core import Workspace, Datastore
from azureml.core.compute import ComputeTarget
from azureml.train.estimator import Estimator
work_space = Workspace.from_config()
data_source = work_space.get_default_datastore()
train_cluster = ComputeTarget(workspace=work_space, name= 'train-cluster')
estimator = Estimator(source_directory =
    'training-experiment',
    script_params = { ' --data-folder' : data_source.as_mount(), ' --regularization':0.8},
    compute_target = train_cluster,
    entry_script = 'train.py',
    conda_packages = ['scikit-learn'])
```

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

NOTE: Each correct selection is worth one point.

	Yes	No
The estimator will look for the files it needs to run an experiment in the training-experiment directory of the local compute environment.	<input type="radio"/>	<input type="radio"/>
The estimator will mount the local data-folder folder and make it available to the script through a parameter.	<input type="radio"/>	<input type="radio"/>
The train.py script file will be created if it does not exist.	<input type="radio"/>	<input type="radio"/>
The estimator can run Scikit-learn experiments.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Yes

Parameter source_directory is a local directory containing experiment configuration and code files needed for a training job.

Box 2: Yes

script_params is a dictionary of command-line arguments to pass to the training script specified in entry_script.

Box 3: No

Box 4: Yes

The conda_packages parameter is a list of strings representing conda packages to be added to the Python environment for the experiment.

NEW QUESTION 2

- (Exam Topic 3)

You train a machine learning model.

You must deploy the model as a real-time inference service for testing. The service requires low CPU utilization and less than 48 MB of RAM. The compute target for the deployed service must initialize automatically while minimizing cost and administrative overhead.

Which compute target should you use?

- A. Azure Kubernetes Service (AKS) inference cluster
- B. Azure Machine Learning compute cluster
- C. Azure Container Instance (ACI)
- D. attached Azure Databricks cluster

Answer: C

Explanation:

Azure Container Instances (ACI) are suitable only for small models less than 1 GB in size. Use it for low-scale CPU-based workloads that require less than 48 GB of RAM.

Note: Microsoft recommends using single-node Azure Kubernetes Service (AKS) clusters for dev-test of larger models.

Reference:

<https://docs.microsoft.com/id-id/azure/machine-learning/how-to-deploy-and-where>

NEW QUESTION 3

- (Exam Topic 3)

You have a dataset that includes confidential data. You use the dataset to train a model.

You must use a differential privacy parameter to keep the data of individuals safe and private. You need to reduce the effect of user data on aggregated results. What should you do?

- A. Decrease the value of the epsilon parameter to reduce the amount of noise added to the data
- B. Increase the value of the epsilon parameter to decrease privacy and increase accuracy
- C. Decrease the value of the epsilon parameter to increase privacy and reduce accuracy
- D. Set the value of the epsilon parameter to 1 to ensure maximum privacy

Answer: C

Explanation:

Differential privacy tries to protect against the possibility that a user can produce an indefinite number of reports to eventually reveal sensitive data. A value known as epsilon measures how noisy, or private, a report is. Epsilon has an inverse relationship to noise or privacy. The lower the epsilon, the more noisy (and private) the data is.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-differential-privacy>

NEW QUESTION 4

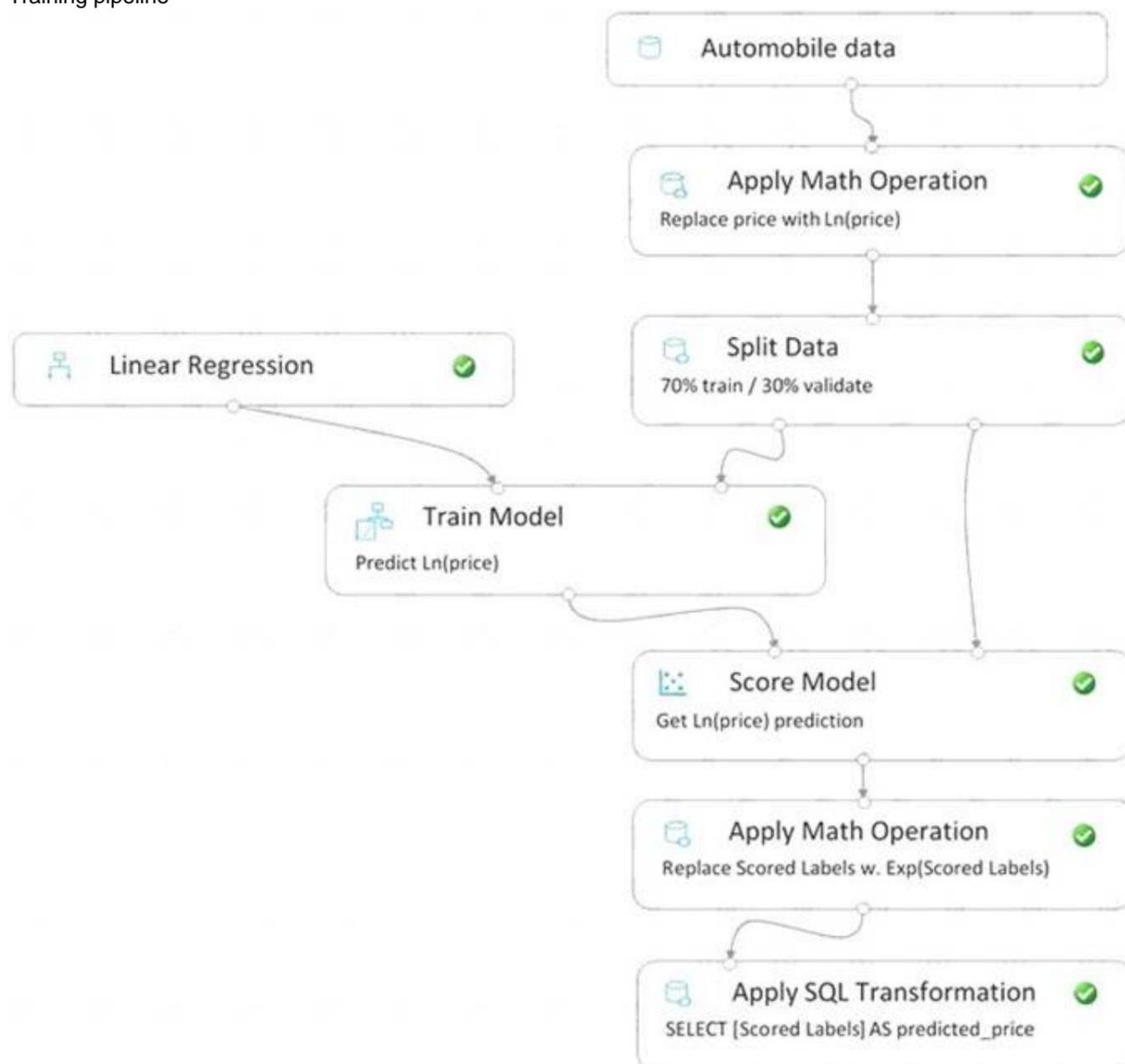
- (Exam Topic 3)

You create a pipeline in designer to train a model that predicts automobile prices.

Because of non-linear relationships in the data, the pipeline calculates the natural log (Ln) of the prices in the training data, trains a model to predict this natural log of price value, and then calculates the exponential of the scored label to get the predicted price.

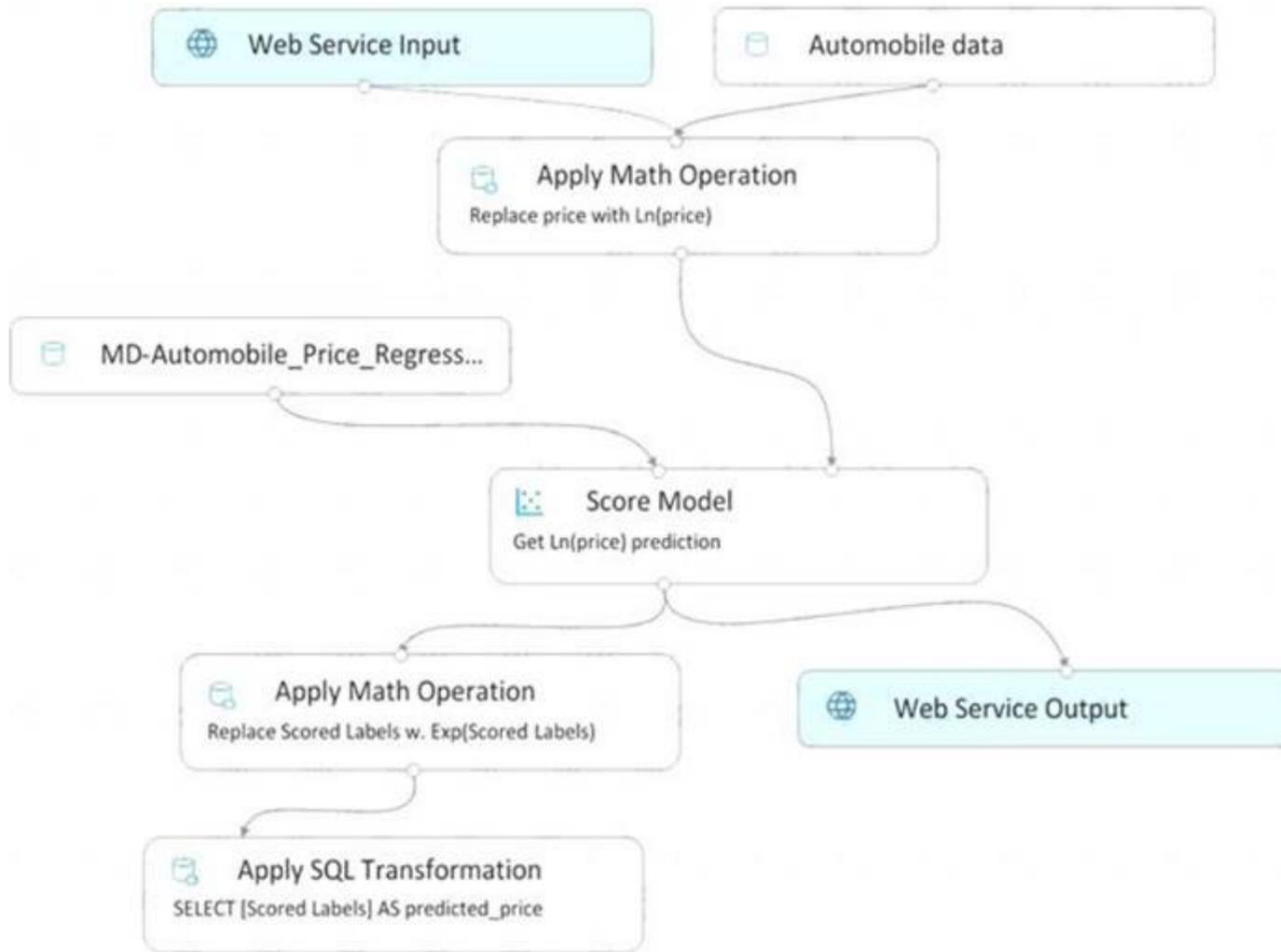
The training pipeline is shown in the exhibit. (Click the Training pipeline tab.)

Training pipeline



You create a real-time inference pipeline from the training pipeline, as shown in the exhibit. (Click the Real-time pipeline tab.)

Real-time pipeline



You need to modify the inference pipeline to ensure that the web service returns the exponential of the scored label as the predicted automobile price and that client applications are not required to include a price value in the input values.

Which three modifications must you make to the inference pipeline? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Connect the output of the Apply SQL Transformation to the Web Service Output module.
- B. Replace the Web Service Input module with a data input that does not include the price column.
- C. Add a Select Columns module before the Score Model module to select all columns other than price.
- D. Replace the training dataset module with a data input that does not include the price column.
- E. Remove the Apply Math Operation module that replaces price with its natural log from the data flow.
- F. Remove the Apply SQL Transformation module from the data flow.

Answer: ACE

NEW QUESTION 5

- (Exam Topic 3)

You create a Python script named train.py and save it in a folder named scripts. The script uses the scikit-learn framework to train a machine learning model.

You must run the script as an Azure Machine Learning experiment on your local workstation. You need to write Python code to initiate an experiment that runs the train.py script.

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.

Answer Area

```

from azureml.core import Experiment, ScriptRunConfig, Environment
from azureml.core.conda_dependencies import CondaDependencies
from azureml.core import Workspace

ws = Workspace.from_config()
py_sk = Environment('sklearn-training')
pkgs = CondaDependencies.create(pip_packages=['scikit-learn', 'azureml-defaults'])
py_sk.python.conda_dependencies = pkgs
script_config = ScriptRunConfig (
    script = 'scripts',
    source_directory = 'train.py',
    arguments = py_sk)

experiment = Experiment(workspace=ws, name='training-experiment')
run = experiment.submit(config=script_config)

```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Graphical user interface, text, application, table, Word Description automatically generated

Box 1: source_directory

source_directory: A local directory containing code files needed for a run. Box 2: script

Script: The file path relative to the source_directory of the script to be run. Box 3: environment

Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.scriptrunconfig>

NEW QUESTION 6

- (Exam Topic 3)

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 analyzing 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.

Solution: Calculate the column median value and use the median value as the replacement for any missing value in the column.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Use the Multiple Imputation by Chained Equations (MICE) method. References: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/>

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

NEW QUESTION 7

- (Exam Topic 3)

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 creating a new experiment in Azure Machine Learning 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. Solution: You use the Stratified split for the sampling mode.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Instead use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode.

Note: SMOTE is used 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.

References:

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

NEW QUESTION 8

- (Exam Topic 3)

You are creating a classification model for a banking company to identify possible instances of credit card fraud. You plan to create the model in Azure Machine Learning by using automated machine learning.

The training dataset that you are using is highly unbalanced. You need to evaluate the classification model.

Which primary metric should you use?

- A. normalized_mean_absolute_error
- B. spearman_correlation
- C. AUC_weighted
- D. accuracy
- E. normalized_root_mean_squared_error

Answer: C

Explanation:

AUC_weighted is a Classification metric.

Note: AUC is the Area under the Receiver Operating Characteristic Curve. Weighted is the arithmetic mean of the score for each class, weighted by the number of true instances in each class.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-understand-automated-ml>

NEW QUESTION 9

- (Exam Topic 3)

You are preparing to build a deep learning convolutional neural network model 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 Machine Learning workspace. The following compute resources are available:

- > a Microsoft Surface device on which Microsoft Office has been installed. Corporate IT policies prevent the installation of additional software
- > a Compute Instance named ds-workstation in the workspace with 2 CPUs and 8 GB of memory
- > an Azure Machine Learning compute target named cpu-cluster with eight CPU-based nodes
- > an Azure Machine Learning compute target named gpu-cluster with four CPU and GPU-based nodes

You need to specify 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.

Which resources should the data scientist use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Resource type

Option

Run code to submit the experiment

	▼
the Microsoft Surface device	
the ds-workstation notebook VM	
the cpu-cluster compute target	
the gpu-cluster compute target	

Run the training script

	▼
the ds-workstation notebook VM	
the cpu-compute target	
the gpu-compute target	
the Microsoft Surface device	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Resource type

Option

Run code to submit the experiment

- the Microsoft Surface device
- the ds-workstation notebook VM
- the cpu-cluster compute target
- the gpu-cluster compute target

Run the training script

- the ds-workstation notebook VM
- the cpu-compute target
- the gpu-compute target
- the Microsoft Surface device

NEW QUESTION 10

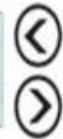
- (Exam Topic 3)

You have a dataset that contains over 150 features. You use the dataset to train a Support Vector Machine (SVM) binary classifier. You need to use the Permutation Feature Importance module in Azure Machine Learning Studio to compute a set of feature importance scores for the dataset. In which order should you perform the actions? To answer, move all actions from the list of actions to the answer area and arrange them in the correct order.

Actions

Answer Area

- Add a Two-Class Support Vector Machine module to initialize the SVM classifier.
- Set the Metric for measuring performance property to **Classification - Accuracy** and then run the experiment.
- Add a Permutation Feature Importance module and connect the trained model and test dataset.
- Add a dataset to the experiment.
- Add a Split Data module to create training and test datasets.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Add a Two-Class Support Vector Machine module to initialize the SVM classifier.

Step 2: Add a dataset to the experiment

Step 3: Add a Split Data module to create training and test dataset.

To generate a set of feature scores requires that you have an already trained model, as well as a test dataset. Step 4: Add a Permutation Feature Importance module and connect to the trained model and test dataset. Step 5: Set the Metric for measuring performance property to Classification - Accuracy and then run the experiment.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-support-vector-mac> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/permutation-feature-importan>

NEW QUESTION 10

- (Exam Topic 3)

You use the designer to create a training pipeline for a classification model. The pipeline uses a dataset that includes the features and labels required for model training.

You create a real-time inference pipeline from the training pipeline. You observe that the schema for the generated web service input is based on the dataset and includes the label column that the model predicts. Client applications that use the service must not be required to submit this value.

You need to modify the inference pipeline to meet the requirement. What should you do?

- A. Add a Select Columns in Dataset module to the inference pipeline after the dataset and use it to select all columns other than the label.
- B. Delete the dataset from the training pipeline and recreate the real-time inference pipeline.

- C. Delete the Web Service Input module from the inference pipeline.
- D. Replace the dataset in the inference pipeline with an Enter Data Manually module that includes data for the feature columns but not the label column.

Answer: A

Explanation:

By default, the Web Service Input will expect the same data schema as the module output data which connects to the same downstream port as it. You can remove the target variable column in the inference pipeline using Select Columns in Dataset module. Make sure that the output of Select Columns in Dataset removing target variable column is connected to the same port as the output of the Web Service Input module.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-deploy>

NEW QUESTION 15

- (Exam Topic 3)

You are developing a machine learning, experiment by using Azure. The following images show the input and output of a machine learning experiment:



Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.
 NOTE: Each correct selection is worth one point.

You need to perform the data transformation applied to the Risk Level column. Which module should you use?

What is the expected input column type for this transformation?

Apply Filter

Build Counting Transform

Convert to Indicator Values

Categorical

Numerical

String

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

You need to perform the data transformation applied to the Risk Level column. Which module should you use?

What is the expected input column type for this transformation?

Apply Filter

Build Counting Transform

Convert to Indicator Values

Categorical

Numerical

String

NEW QUESTION 19

- (Exam Topic 3)

You retrain an existing model.

You need to register the new version of a model while keeping the current version of the model in the registry.

What should you do?

- A. Register a model with a different name from the existing model and a custom property named version with the value 2.
- B. Register the model with the same name as the existing model.
- C. Save the new model in the default datastore with the same name as the existing model.
- D. Do not register the new model.
- E. Delete the existing model and register the new one with the same name.

Answer: B

Explanation:

Model version: A version of a registered model. When a new model is added to the Model Registry, it is added as Version 1. Each model registered to the same model name increments the version number.

Reference:

<https://docs.microsoft.com/en-us/azure/databricks/applications/mlflow/model-registry>

NEW QUESTION 20

- (Exam Topic 3)

You are training machine learning models in Azure Machine Learning. You use Hyperdrive to tune the hyperparameters. In previous model training and tuning runs, many models showed similar performance. You need to select an early termination policy that meets the following requirements:

- accounts for the performance of all previous runs when evaluating the current run
- avoids comparing the current run with only the best performing run to date

Which two early termination policies should you use? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. Bandit
- B. Median stopping
- C. Default
- D. Truncation selection

Answer: BC

Explanation:

The Median Stopping policy computes running averages across all runs and cancels runs whose best performance is worse than the median of the running averages.

If no policy is specified, the hyperparameter tuning service will let all training runs execute to completion. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.medianstoppingpolicy> <https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.truncationselectionpolicy> <https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.banditpolicy>

NEW QUESTION 22

- (Exam Topic 3)

You are running a training experiment on remote compute in Azure Machine Learning.

The experiment is configured to use a conda environment that includes the mlflow and azureml-contrib-run packages.

You must use MLflow as the logging package for tracking metrics generated in the experiment. You need to complete the script for the experiment.

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

NOTE: Each correct selection is worth one point.

```
import numpy as np
# Import library to log metrics
```

```
from azureml.core import Run
import mlflow
import logging
```

```
# Start logging for this run
```

```
run = Run.get_context()
mlflow.start_run()
logger = logging.getLogger('Run')
reg_rate = 0.01
# Log the reg_rate metric
```

```
run.log('reg_rate', np.float(reg_rate))
mlflow.log_metric('reg_rate', np.float(reg_rate))
logger.info(np.float(reg_rate))
```

```
# Stop logging for this run
```

```
run.complete()
mlflow.end_run()
logger.setLevel(logging.INFO)
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: import mlflow

Import the mlflow and Workspace classes to access MLflow's tracking URI and configure your workspace. Box 2: mlflow.start_run()

Set the MLflow experiment name with set_experiment() and start your training run with start_run(). Box 3: mlflow.log_metric('..')

Use log_metric() to activate the MLflow logging API and begin logging your training run metrics. Box 4: mlflow.end_run()

Close the run: run.endRun() Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-mlflow>

NEW QUESTION 24

- (Exam Topic 3)

You run an experiment that uses an AutoMLConfig class to define an automated machine learning task with a maximum of ten model training iterations. The task

will attempt to find the best performing model based on a metric named accuracy.

You submit the experiment with the following code:

You need to create Python code that returns the best model that is generated by the automated machine learning task. Which code segment should you use?

A)

```
best_model = automl_run.get_details()
```

B)

```
best_model = automl_run.get_output()[1]
```

C)

```
best_model = automl_run.get_file_names()[1]
```

D)

```
best_model = automl_run.get_metrics()
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D

Answer: B

Explanation:

The get_output method returns the best run and the fitted model. Reference:

<https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/automated-mach>

NEW QUESTION 27

- (Exam Topic 3)

You are solving a classification task.

You must evaluate your model on a limited data sample by using k-fold cross validation. You start by configuring a k parameter as the number of splits.

You need to configure the k parameter for the cross-validation. Which value should you use?

- A. k=0.5
- B. k=0
- C. k=5
- D. k=1

Answer: C

Explanation:

Leave One Out (LOO) cross-validation

Setting $K = n$ (the number of observations) yields n-fold and is called leave-one out cross-validation (LOO), a special case of the K-fold approach.

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 tradeoff.

NEW QUESTION 30

- (Exam Topic 3)

You create an Azure Machine Learning workspace.

You must create a custom role named DataScientist that meets the following requirements:

- Role members must not be able to delete the workspace.
- Role members must not be able to create, update, or delete compute resource in the workspace.
- Role members must not be able to add new users to the workspace.

You need to create a JSON file for the DataScientist role in the Azure Machine Learning workspace. The custom role must enforce the restrictions specified by the IT Operations team.

Which JSON code segment should you use?

A)

```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": ["*"],
  "NotActions": [
    "Microsoft.MachineLearningServices/workspaces/*/delete",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft/MachineLearningServices/workspaces/ml-ws"
  ]
}
```

B)

```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": ["*"],
  "NotActions": [],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
  ]
}
```

C)

```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": [
    "Microsoft.MachineLearningServices/workspaces/*/delete",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "NotActions": [],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
  ]
}
```

D)

```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": [],
  "NotActions": ["*"],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
  ]
}
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D

Answer: A

Explanation:

The following custom role can do everything in the workspace except for the following actions:

- It can't create or update a compute resource.
- It can't delete a compute resource.
- It can't add, delete, or alter role assignments.
- It can't delete the workspace.

To create a custom role, first construct a role definition JSON file that specifies the permission and scope for the role. The following example defines a custom role named "Data Scientist Custom" scoped at a specific workspace level:

data_scientist_custom_role.json :

```
{
  "Name": "Data Scientist Custom", "IsCustom": true,
  "Description": "Can run experiment but can't create or delete compute.", "Actions": ["*"],
  "NotActions": [
    "Microsoft.MachineLearningServices/workspaces/*/delete", "Microsoft.MachineLearningServices/workspaces/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write", "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "AssignableScopes": [ "/subscriptions/<subscription_id>/resourceGroups/<resource_group_name>/providers/Microsoft.MachineLearnin"
]
}
```

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-assign-roles>

NEW QUESTION 35

- (Exam Topic 3)

You are creating a binary classification by using a two-class logistic regression model. You need to evaluate the model results for imbalance. Which evaluation metric should you use?

- A. Relative Absolute Error
- B. AUC Curve
- C. Mean Absolute Error

D. Relative Squared Error

Answer: B

Explanation:

One can inspect the true positive rate vs. the false positive rate in the Receiver Operating Characteristic (ROC) curve and the corresponding Area Under the Curve (AUC) value. The closer this curve is to the upper left corner, the better the classifier's performance is (that is maximizing the true positive rate while minimizing the false positive rate). Curves that are close to the diagonal of the plot, result from classifiers that tend to make predictions that are close to random guessing.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio/evaluate-model-performance#evaluating-a-bina>

NEW QUESTION 39

- (Exam Topic 3)

You are moving a large dataset from Azure Machine Learning Studio to a Weka environment. You need to format the data for the Weka environment. Which module should you use?

- A. Convert to CSV
- B. Convert to Dataset
- C. Convert to ARFF
- D. Convert to SVMLight

Answer: C

Explanation:

Use the Convert to ARFF module in Azure Machine Learning Studio, to convert datasets and results in Azure Machine Learning to the attribute-relation file format used by the Weka toolset. This format is known as ARFF.

The ARFF data specification for Weka supports multiple machine learning tasks, including data preprocessing, classification, and feature selection. In this format, data is organized by entities and their attributes, and is contained in a single text file.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/convert-to-arff>

NEW QUESTION 44

- (Exam Topic 3)

You deploy a model in Azure Container Instance.

You must use the Azure Machine Learning SDK to call the model API.

You need to invoke the deployed model using native SDK classes and methods.

How should you complete the command? To answer, select the appropriate options in the answer areas.

NOTE: Each correct selection is worth one point.

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

```
from azureml.core.webservice import requests
from azureml.core.webservice import Webservice
from azureml.core.webservice import LocalWebservice
```

```
import json
ws = Workspace.from_config()
service_name = "mlmodel1-service"
service = Webservice(name=service_name, workspace=ws)
x_new = [[2,101.5,1,24,21], [1,89.7,4,41,21]]
input_json = json.dumps({"data": x_new})
```

```
predictions = service.run(input_json)
predictions = requests.post(service.scoring_uri, input_json)
predictions = service.deserialize(ws, input_json)
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: from azureml.core.webservice import Webservice

The following code shows how to use the SDK to update the model, environment, and entry script for a web service to Azure Container Instances:

```
from azureml.core import Environment
```

```
from azureml.core.webservice import Webservice
```

```
from azureml.core.model import Model, InferenceConfig
```

Example: The following code demonstrates sending data to the service:

```
import json
test_sample = json.dumps({'data': [ [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
[10, 9, 8, 7, 6, 5, 4, 3, 2, 1]
]})
```

```
test_sample = bytes(test_sample, encoding='utf8') prediction = service.run(input_data=test_sample)
```

```
print(prediction)
```

Reference:

<https://docs.microsoft.com/bs-latn-ba/azure/machine-learning/how-to-deploy-azure-container-instance> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment>

NEW QUESTION 47

- (Exam Topic 3)

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.

```
from azureml.train.hyperdrive import RandomParameterSampling
param_sampling = RandomParameterSampling( {
"learning_rate" :
"batch_size":
"keep_probability" :
}
```

uniform(10,3)
normal(10,3)
choice(10,3)
Loguniform(10,3)

choice(16,32,64)
choice(range(16,64))
normal(16,32,64)
normal(range(16,64))

choice(range(0.05, 0.1))
uniform(0.05, 0.1)
normal(0.05, 0.1)
lognormal(0.05, 0.1)

- A. Mastered
- B. Not Mastered

Answer: A

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>

NEW QUESTION 50

- (Exam Topic 3)

You are creating a machine learning model. You need to identify outliers in the data.

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

NOTE: Each correct selection is worth one point. NOTE: Each correct selection is worth one point.

- A. box plot
- B. scatter
- C. random forest diagram
- D. Venn diagram
- E. ROC curve

Answer: AB

Explanation:

The box-plot algorithm can be used to display outliers.

One other way to quickly identify Outliers visually is to create scatter plots. References:

<https://blogs.msdn.microsoft.com/azuredev/2017/05/27/data-cleansing-tools-in-azure-machine-learning/>

NEW QUESTION 52

- (Exam Topic 3)

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>

NEW QUESTION 56

- (Exam Topic 3)

You are conducting feature engineering to prepuce data for further analysis. The data includes seasonal patterns on inventory requirements.

You need to select the appropriate method to conduct feature engineering on the data. Which method should you use?

- A. Exponential Smoothing (ETS) function.
- B. One Class Support Vector Machine module
- C. Time Series Anomaly Detection module
- D. Finite Impulse Response (FIR) Filter module.

Answer: D

NEW QUESTION 58

- (Exam Topic 3)

You are using the Hyperdrive feature in Azure Machine Learning to train a model. You configure the Hyperdrive experiment by running the following code:

```
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, 128)
    "number_of_hidden_layers": choice(range(3,5))
})
```

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

NOTE: Each correct selection is worth one point.

	Yes	No
By defining sampling in this manner, every possible combination of the parameters will be tested.	<input type="radio"/>	<input type="radio"/>
Random values of the <code>learning_rate</code> parameter will be selected from a normal distribution with a mean of 10 and a standard deviation of 3.	<input type="radio"/>	<input type="radio"/>
The <code>keep_probability</code> parameter value will always be either 0.05 or 0.1 .	<input type="radio"/>	<input type="radio"/>
Random values for the <code>number_of_hidden_layers</code> parameter will be selected from a normal distribution with a mean of 3 and a standard deviation of 5.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Yes

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.

Box 2: Yes

learning_rate has a normal distribution with mean value 10 and a standard deviation of 3.

Box 3: No

keep_probability has a uniform distribution with a minimum value of 0.05 and a maximum value of 0.1.

Box 4: No

number_of_hidden_layers takes on one of the values [3, 4, 5].

Reference:

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

NEW QUESTION 62

- (Exam Topic 3)

You create a binary classification model by using Azure Machine Learning Studio.

You must tune hyperparameters by performing a parameter sweep of the model. The parameter sweep must meet the following requirements:

- > iterate all possible combinations of hyperparameters
- > minimize computing resources required to perform the sweep
- > You need to perform a parameter sweep of the model.

Which parameter sweep mode should you use?

- A. Random sweep
- B. Sweep clustering
- C. Entire grid
- D. Random grid
- E. Random seed

Answer: D

Explanation:

Maximum number of runs on random grid: This option also controls the number of iterations over a random sampling of parameter values, but the values are not generated randomly from the specified range; instead, a matrix is created of all possible combinations of parameter values and a random sampling is taken over the matrix. This method is more efficient and less prone to regional oversampling or undersampling.

If you are training a model that supports an integrated parameter sweep, you can also set a range of seed values to use and iterate over the random seeds as well. This is optional, but can be useful for avoiding bias introduced by seed selection.

NEW QUESTION 66

- (Exam Topic 3)

You have a multi-class image classification deep learning model that uses a set of labeled photographs. You create the following code to select hyperparameter values when training the model.

```
from azureml.train.hyperdrive import BayesianParameterSampling
param_sampling = BayesianParametersSampling ({
    "learning_rate": uniform(0.01, 0.1),
    "batch_size": choice(16, 32, 64, 128)}
)
```

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

NOTE: Each correct selection is worth one point.

Yes

No

Hyperparameter combinations for the runs are selected based on how previous samples performed in the previous experiment run.

The learning rate value 0.09 might be used during model training.

You can define an early termination policy for this hyperparameter tuning run.

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Yes

Hyperparameters are adjustable parameters you choose to train a model that govern the training process itself. Azure Machine Learning allows you to automate hyperparameter exploration in an efficient manner, saving you significant time and resources. You specify the range of hyperparameter values and a maximum number of training runs. The system then automatically launches multiple simultaneous runs with different parameter configurations and finds the configuration that results in the best performance, measured by the metric you choose. Poorly performing training runs are automatically early terminated, reducing wastage of compute resources. These resources are instead used to explore other hyperparameter configurations.

Box 2: Yes

uniform(low, high) - Returns a value uniformly distributed between low and high

Box 3: No

Bayesian sampling does not currently support any early termination policy. Reference:

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

NEW QUESTION 70

- (Exam Topic 3)

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

```
from azureml.core.compute import ComputeTarget, AmlCompute
from azureml.core.compute_target import ComputeTargetException
ws = Workspace.from_config()
cluster_name = 'aml-cluster'
try:
    training_compute = ComputeTarget(workspace=ws, name=cluster_name)
except ComputeTargetException:
    compute_config = AmlCompute.provisioning_configuration(vm_size='STANDARD_D2_V2', vm_priority='lowpriority',
max_nodes=4)
    training_compute = ComputeTarget.create(ws, cluster_name, compute_config)
    training_compute.wait_for_completion(show_output=True)
```

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

NOTE: Each correct selection is worth one point.

	Yes	No
If a training cluster named aml-cluster already exists in the workspace, it will be deleted and replaced.	<input type="radio"/>	<input type="radio"/>
The wait_for_completion() method will not return until the aml-cluster compute has four active nodes.	<input type="radio"/>	<input type="radio"/>
If the code creates a new aml-cluster compute target, it may be preempted due to capacity constraints.	<input type="radio"/>	<input type="radio"/>
The aml-cluster compute target is deleted from the workspace after the training experiment completes.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: No

If a training cluster already exists it will be used. Box 2: Yes

The wait_for_completion method waits for the current provisioning operation to finish on the cluster. Box 3: Yes

Low Priority VMs use Azure's excess capacity and are thus cheaper but risk your run being pre-empted.

Box 4: No

Need to use training_compute.delete() to deprovision and delete the AmlCompute target. Reference:

<https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/training/train-on> <https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.computetarget>

NEW QUESTION 74

- (Exam Topic 3)

You create a multi-class image classification deep learning model that uses a set of labeled images. You create a script file named train.py that uses the PyTorch 1.3 framework to train the model.

You must run the script by using an estimator. The code must not require any additional Python libraries to be installed in the environment for the estimator. The time required for model training must be minimized.

You need to define the estimator that will be used to run the script. Which estimator type should you use?

- A. TensorFlow
- B. PyTorch
- C. SKLearn
- D. Estimator

Answer: B

Explanation:

For PyTorch, TensorFlow and Chainer tasks, Azure Machine Learning provides respective PyTorch, TensorFlow, and Chainer estimators to simplify using these frameworks.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-ml-models>

NEW QUESTION 79

- (Exam Topic 3)

You create a batch inference pipeline by using the Azure ML SDK. You run the pipeline by using the following code:

```
from azureml.pipeline.core import Pipeline
from azureml.core.experiment import Experiment
pipeline = Pipeline(workspace=ws, steps=[parallelrun_step]) pipeline_run = Experiment(ws, 'batch_pipeline').submit(pipeline)
You need to monitor the progress of the pipeline execution.
```

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. Run the following code in a notebook:

```
from azureml.contrib.interpret.explanation.explanation_client import ExplanationClient
client = ExplanationClient.from_run(pipeline_run)
explanation = client.download_model_explanation()
explanation = client.download_model_explanation(top_k=4)
global_importance_values = explanation.get_ranked_global_values()
global_importance_names = explanation.get_ranked_global_names()
print('global importance values: {}'.format(global_importance_values))
print('global importance names: {}'.format(global_importance_names))
```

B. Use the Inference Clusters tab in Machine Learning Studio.

C. Use the Activity log in the Azure portal for the Machine Learning workspace.

D. Run the following code in a notebook:

```
from azureml.widgets import RunDetails
RunDetails(pipeline_run).show()
```

E. Run the following code and monitor the console output from the PipelineRun object:

```
pipeline_run.wait_for_completion(show_output=True)
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D
- E. Option E

Answer: DE

Explanation:

A batch inference job can take a long time to finish. This example monitors progress by using a Jupyter widget. You can also manage the job's progress by using:

➤ Azure Machine Learning Studio.

➤ Console output from the PipelineRun object. `from azureml.widgets import RunDetails RunDetails(pipeline_run).show()`

`pipeline_run.wait_for_completion(show_output=True)` Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-parallel-run-step#monitor-the-parallel-run>

NEW QUESTION 80

- (Exam Topic 3)

You plan to build a team data science environment. Data for training models in machine learning pipelines will be over 20 GB in size.

You have the following requirements:

➤ Models must be built using Caffe2 or Chainer frameworks.

➤ Data scientists must be able to use a data science environment to build the machine learning pipelines and train models on their personal devices in both connected and disconnected network environments.

➤ Personal devices must support updating machine learning pipelines when connected to a network. You need to select a data science environment.

Which environment should you use?

- A. Azure Machine Learning Service
- B. Azure Machine Learning Studio
- C. Azure Databricks
- D. Azure Kubernetes Service (AKS)

Answer: A

Explanation:

The Data Science Virtual Machine (DSVM) is a customized VM image on Microsoft's Azure cloud built specifically for doing data science. Caffe2 and Chainer are supported by DSVM.

DSVM integrates with Azure Machine Learning.

NEW QUESTION 81

- (Exam Topic 3)

You create a Python script that runs a training experiment in Azure Machine Learning. The script uses the Azure Machine Learning SDK for Python.

You must add a statement that retrieves the names of the logs and outputs generated by the script. You need to reference a Python class object from the SDK for the statement.

Which class object should you use?

- A. Run
- B. ScriptRunConfig
- C. Workspace
- D. Experiment

Answer: A

Explanation:

A run represents a single trial of an experiment. Runs are used to monitor the asynchronous execution of a trial, log metrics and store output of the trial, and to analyze results and access artifacts generated by the trial.

The run Class get_all_logs method downloads all logs for the run to a directory. Reference:

[https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run\(class\)](https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run(class))

NEW QUESTION 83

- (Exam Topic 3)

You are building a machine learning model for translating English language textual content into French language textual content.

You need to build and train the machine learning model to learn the sequence of the textual content. Which type of neural network should you use?

- A. Multilayer Perceptions (MLPs)
- B. Convolutional Neural Networks (CNNs)
- C. Recurrent Neural Networks (RNNs)
- D. Generative Adversarial Networks (GANs)

Answer: C

Explanation:

To translate a corpus of English text to French, we need to build a recurrent neural network (RNN).

Note: RNNs are designed to take sequences of text as inputs or return sequences of text as outputs, or both. They're called recurrent because the network's hidden layers have a loop in which the output and cell state from each time step become inputs at the next time step. This recurrence serves as a form of memory. It allows contextual information to flow through the network so that relevant outputs from previous time steps can be applied to network operations at the current time step.

References:

<https://towardsdatascience.com/language-translation-with-rnns-d84d43b40571>

NEW QUESTION 87

- (Exam Topic 3)

You use the following code to run a script as an experiment in Azure Machine Learning:

```
from azureml.core import Workspace, Experiment, Run
from azureml.core import RunConfig, ScriptRunConfig
ws = Workspace.from_config()
run_config = RunConfiguration()
run_config.target='local'
script_config = ScriptRunConfig(source_directory='./script', script='experiment.py', run_config=run_config)
experiment = Experiment(workspace=ws, name='script experiment')
run = experiment.submit(config=script_config)
run.wait_for_completion()
```

You must identify the output files that are generated by the experiment run. You need to add code to retrieve the output file names.

Which code segment should you add to the script?

- A. files = run.get_properties()
- B. files= run.get_file_names()
- C. files = run.get_details_with_logs()
- D. files = run.get_metrics()
- E. files = run.get_details()

Answer: B

Explanation:

You can list all of the files that are associated with this run record by called run.get_file_names() Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-track-experiments>

NEW QUESTION 90

- (Exam Topic 3)

You are the owner of an Azure Machine Learning workspace.

You must prevent the creation or deletion of compute resources by using a custom role. You must allow all other operations inside the workspace.

You need to configure the custom role.

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

NOTE: Each correct selection is worth one point.

Answer Area

```
{
  "Name": "Data Scientist Custom",
  "IsCustom": true
  "Description": "Description"
  "Actions": [
    Microsoft.MachineLearningServices/workspaces/*/read
    Microsoft.MachineLearningServices/workspaces/computes/*/write
    Microsoft.MachineLearningServices/workspaces/delete
    Microsoft.MachineLearningServices/workspaces/*/write
    Microsoft.MachineLearningServices/workspaces/computes/*/write
    Microsoft.MachineLearningServices/workspaces/delete
  ],
  "NotActions": [
    Microsoft.MachineLearningServices/workspaces/*/read
    Microsoft.MachineLearningServices/workspaces/*/write
    Microsoft.MachineLearningServices/workspaces/computes/*/delete
    Microsoft.MachineLearningServices/workspaces/*/read
    Microsoft.MachineLearningServices/workspaces/*/write
    Microsoft.MachineLearningServices/workspaces/computes/*/write
  ],
  "AssignableScopes": [
    "/subscriptions/<subscription_id>"
  ]
}
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Graphical user interface, application Description automatically generated
 Graphical user interface, application Description automatically generated

Box 1: Microsoft.MachineLearningServices/workspaces/*/read

Reader role: Read-only actions in the workspace. Readers can list and view assets, including datastore credentials, in a workspace. Readers can't create or update these assets.

Box 2: Microsoft.MachineLearningServices/workspaces/*/write

If the roles include Actions that have a wildcard (*), the effective permissions are computed by subtracting the NotActions from the allowed Actions.

Box 3: Microsoft.MachineLearningServices/workspaces/computes/*/delete

Box 4: Microsoft.MachineLearningServices/workspaces/computes/*/write Reference:

<https://docs.microsoft.com/en-us/azure/role-based-access-control/overview#how-azure-rbac-determines-if-a-use>

NEW QUESTION 94

- (Exam Topic 3)

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 creating a new experiment in Azure Machine Learning Studio.

One class has a much smaller number of observations than tin- other classes in the training set. You need to select an appropriate data sampling strategy to compensate for the class imbalance. Solution: You use the Principal Components Analysis (PCA) sampling mode.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Instead use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode.

Note: SMOTE is used 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.

NEW QUESTION 96

- (Exam Topic 3)

You create and register a model in an Azure Machine Learning workspace.

You must use the Azure Machine Learning SDK to implement a batch inference pipeline that uses a ParallelRunStep to score input data using the model. You

must specify a value for the ParallelRunConfig compute_target setting of the pipeline step.
You need to create the compute target. Which class should you use?

- A. BatchCompute
- B. AdlaCompute
- C. AmlCompute
- D. Aks Compute

Answer: C

Explanation:

Compute target to use for ParallelRunStep. This parameter may be specified as a compute target object or the string name of a compute target in the workspace. The compute_target target is of AmlCompute or string.

Note: An Azure Machine Learning Compute (AmlCompute) is a managed-compute infrastructure that allows you to easily create a single or multi-node compute. The compute is created within your workspace region as a resource that can be shared with other users

Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-contrib-pipeline-steps/azureml.contrib.pipeline.steps.parall> [https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute\(class\)](https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute(class))

NEW QUESTION 98

- (Exam Topic 3)

You are performing clustering by using the K-means algorithm. You need to define the possible termination conditions. Which three conditions can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. A fixed number of iterations is executed.
- B. The residual sum of squares (RSS) rises above a threshold.
- C. The sum of distances between centroids reaches a maximum.
- D. The residual sum of squares (RSS) falls below a threshold.
- E. Centroids do not change between iterations.

Answer: ADE

Explanation:

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/k-means-clustering> <https://nlp.stanford.edu/IR-book/html/htmledition/k-means-1.html>

NEW QUESTION 99

- (Exam Topic 3)

You plan to create a speech recognition deep learning model. The model must support the latest version of Python. You need to recommend a deep learning framework for speech recognition to include in the Data Science Virtual Machine (DSVM). What should you recommend?

- A. Apache Drill
- B. Tensorflow
- C. Rattle
- D. Weka

Answer: B

Explanation:

TensorFlow is an open source library for numerical computation and large-scale machine learning. It uses Python to provide a convenient front-end API for building applications with the framework

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 PDE (partial differential equation) based simulations.

References:

<https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html>

NEW QUESTION 100

- (Exam Topic 3)

You plan to run a Python script as an Azure Machine Learning experiment. The script must read files from a hierarchy of folders. The files will be passed to the script as a dataset argument. You must specify an appropriate mode for the dataset argument. Which two modes can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. to_pandas_dataframe ()
- B. as_download()
- C. as_upload()
- D. as_mount ()

Answer: B

Explanation:

Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.data.filedataset?view=azure-ml-py>

NEW QUESTION 104

- (Exam Topic 3)

You have a feature set containing the following numerical features: X, Y, and Z.
 The Poisson correlation coefficient (r-value) of X, Y, and Z features is shown in the following image:

	X	Y	Z
X	1	0.149676	-0.106276
Y	0.149676	1	0.859122
Z	-0.106276	0.859122	1

Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.
 NOTE: Each correct selection is worth one point.

What is the r-value for the correlation of Y to Z?

▼

-0.106276

0.149676

0.859122

1

Which type of relationship exists between Z and Y in the feature set?

▼

a positive linear relationship

a negative linear relationship

no linear relationship

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: 0.859122
 Box 2: a positively linear relationship
 +1 indicates a strong positive linear relationship
 -1 indicates a strong negative linear correlation
 0 denotes no linear relationship between the two variables. References:
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/compute-linear-correlation>

NEW QUESTION 107

- (Exam Topic 3)

You previously deployed a model that was trained using a tabular dataset named training-dataset, which is based on a folder of CSV files. Over time, you have collected the features and predicted labels generated by the model in a folder containing a CSV file for each month. You have created two tabular datasets based on the folder containing the inference data: one named predictions-dataset with a schema that matches the training data exactly, including the predicted label; and another named features-dataset with a schema containing all of the feature columns and a timestamp column based on the filename, which includes the day, month, and year. You need to create a data drift monitor to identify any changing trends in the feature data since the model was trained. To accomplish this, you must define the required datasets for the data drift monitor. Which datasets should you use to configure the data drift monitor? To answer, drag the appropriate datasets to the correct data drift monitor options. Each source 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.

Target datasets

training-dataset

predictions-dataset

features-dataset

Answer Area

Baseline dataset

Target dataset

Target dataset

Target dataset

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Text Description automatically generated with medium confidence
 Box 1: training-dataset
 Baseline dataset - usually the training dataset for a model. Box 2: predictions-dataset
 Target dataset - usually model input data - is compared over time to your baseline dataset. This comparison means that your target dataset must have a timestamp column specified.
 The monitor will compare the baseline and target datasets. Reference:
<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-monitor-datasets>

NEW QUESTION 110

- (Exam Topic 3)

You are using a decision tree algorithm. You have trained a model that generalizes well at a tree depth equal to 10. You need to select the bias and variance properties of the model with varying tree depth values.

Which properties should you select for each tree depth? To answer, select the appropriate options in the answer area.

Answer Area

Tree Depth	Bias	Variance
5	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; text-align: right;">▼</div> <div style="padding: 2px;">High</div> <div style="padding: 2px;">Low</div> <div style="padding: 2px;">Identical</div> </div>	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; text-align: right;">▼</div> <div style="padding: 2px;">High</div> <div style="padding: 2px;">Low</div> <div style="padding: 2px;">Identical</div> </div>
15	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; text-align: right;">▼</div> <div style="padding: 2px;">High</div> <div style="padding: 2px;">Low</div> <div style="padding: 2px;">Identical</div> </div>	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; text-align: right;">▼</div> <div style="padding: 2px;">High</div> <div style="padding: 2px;">Low</div> <div style="padding: 2px;">Identical</div> </div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

In decision trees, the depth of the tree determines the variance. A complicated decision tree (e.g. deep) has low bias and high variance.

Note: In statistics and machine learning, the bias–variance tradeoff is the property of a set of predictive models whereby models with a lower bias in parameter estimation have a higher variance of the parameter estimates across samples, and vice versa. Increasing the bias will decrease the variance. Increasing the variance will decrease the bias.

References:

<https://machinelearningmastery.com/gentle-introduction-to-the-bias-variance-trade-off-in-machine-learning/>

NEW QUESTION 114

- (Exam Topic 3)

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 have a Python script named train.py in a local folder named scripts. The script trains a regression model by using scikit-learn. The script includes code to load a training data file which is also located in the scripts folder.

You must run the script as an Azure ML experiment on a compute cluster named aml-compute.

You need to configure the run to ensure that the environment includes the required packages for model training. You have instantiated a variable named aml-compute that references the target compute cluster.

Solution: Run the following code:

```
from azureml.train.estimator import Estimator
sk_est = Estimator(source_directory='./scripts',
compute_target=aml-compute,
entry_script='train.py',
conda_packages=['scikit-learn'])
```

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

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

Example:

```
from azureml.train.sklearn import SKLearn
}
estimator = SKLearn(source_directory=project_folder, compute_target=compute_target, entry_script='train_iris.py')
)
```

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-scikit-learn>

NEW QUESTION 115

- (Exam Topic 3)

You are developing a data science workspace that uses an Azure Machine Learning service. You need to select a compute target to deploy the workspace. What should you use?

- A. Azure Data Lake Analytics

- B. Azure Databrick .
- C. Apache Spark for HDInsight.
- D. Azure Container Service

Answer: D

Explanation:

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

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-deploy-and-where>

NEW QUESTION 116

- (Exam Topic 3)

You create a multi-class image classification deep learning model.

The model must be retrained monthly with the new image data fetched from a public web portal. You create an Azure Machine Learning pipeline to fetch new data, standardize the size of images, and retrain the model.

You need to use the Azure Machine Learning SDK to configure the schedule for the pipeline.

Which four 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.

Actions	Answer Area
Publish the pipeline.	
Retrieve the pipeline ID.	
Create a ScheduleRecurrence(frequency= 'Month', interval=1, start_time='2019-01-01T00:00:00') object.	
Define a pipeline parameter named RunDate .	
Define a new Azure Machine Learning pipeline StepRun object with the step ID of the first step in the pipeline.	
Define an Azure Machine Learning pipeline schedule using the schedule.create method with the defined recurrence specification.	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Publish the pipeline.

To schedule a pipeline, you'll need a reference to your workspace, the identifier of your published pipeline, and the name of the experiment in which you wish to create the schedule.

Step 2: Retrieve the pipeline ID. Needed for the schedule.

Step 3: Create a ScheduleRecurrence..

To run a pipeline on a recurring basis, you'll create a schedule. A Schedule associates a pipeline, an experiment, and a trigger.

First create a schedule. Example: Create a Schedule that begins a run every 15 minutes: recurrence = ScheduleRecurrence(frequency="Minute", interval=15)

Step 4: Define an Azure Machine Learning pipeline schedule.. Example, continued:

```
recurring_schedule = Schedule.create(ws, name="MyRecurringSchedule", description="Based on time", pipeline_id=pipeline_id, experiment_name=experiment_name, recurrence=recurrence)
```

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-schedule-pipelines>

NEW QUESTION 119

- (Exam Topic 3)

You create a multi-class image classification deep learning model that uses the PyTorch deep learning framework.

You must configure Azure Machine Learning Hyperdrive to optimize the hyperparameters for the classification model.

You need to define a primary metric to determine the hyperparameter values that result in the model with the best accuracy score.

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

NOTE: Each correct selection is worth one point.

- A. Set the primary_metric_goal of the estimator used to run the bird_classifier_train.py script to maximize.
- B. Add code to the bird_classifier_train.py script to calculate the validation loss of the model and log it as a float value with the key loss.
- C. Set the primary_metric_goal of the estimator used to run the bird_classifier_train.py script to minimize.
- D. Set the primary_metric_name of the estimator used to run the bird_classifier_train.py script to accuracy.
- E. Set the primary_metric_name of the estimator used to run the bird_classifier_train.py script to loss.
- F. Add code to the bird_classifier_train.py script to calculate the validation accuracy of the model and log it as a float value with the key accuracy.

Answer: ADF

Explanation:

AD:

primary_metric_name="accuracy", primary_metric_goal=PrimaryMetricGoal.MAXIMIZE

Optimize the runs to maximize "accuracy". Make sure to log this value in your training script. Note:

primary_metric_name: The name of the primary metric to optimize. The name of the primary metric needs to exactly match the name of the metric logged by the training script.

primary_metric_goal: It can be either PrimaryMetricGoal.MAXIMIZE or PrimaryMetricGoal.MINIMIZE and determines whether the primary metric will be maximized or minimized when evaluating the runs.

F: The training script calculates the val_accuracy and logs it as "accuracy", which is used as the primary metric.

NEW QUESTION 124

- (Exam Topic 3)

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 have a Python script named train.py in a local folder named scripts. The script trains a regression model by using scikit-learn. The script includes code to load a training data file which is also located in the scripts folder.

You must run the script as an Azure ML experiment on a compute cluster named aml-compute.

You need to configure the run to ensure that the environment includes the required packages for model training. You have instantiated a variable named aml-compute that references the target compute cluster.

Solution: Run the following code:

```
from azureml.train.sklearn import SKLearn
sk_est = SKLearn(source_directory='./scripts',
                 compute_target=aml-compute,
                 entry_script='train.py')
```

Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

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

Example:

```
from azureml.train.sklearn import SKLearn
}
estimator = SKLearn(source_directory=project_folder, compute_target=compute_target, entry_script='train_iris.py'
)
```

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-scikit-learn>

NEW QUESTION 126

- (Exam Topic 3)

You train and register a machine learning model. You create a batch inference pipeline that uses the model to generate predictions from multiple data files.

You must publish the batch inference pipeline as a service that can be scheduled to run every night. You need to select an appropriate compute target for the inference service.

Which compute target should you use?

- A. Azure Machine Learning compute instance
- B. Azure Machine Learning compute cluster
- C. Azure Kubernetes Service (AKS)-based inference cluster
- D. Azure Container Instance (ACI) compute target

Answer: B

Explanation:

Azure Machine Learning compute clusters is used for Batch inference. Run batch scoring on serverless compute. Supports normal and low-priority VMs. No support for real-time inference.

Reference:

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

NEW QUESTION 127

- (Exam Topic 3)

The finance team asks you 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 Machine Learning workspace and ensure that an error will be raised if the container does not exist.

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

NOTE: Each correct selection is worth one point.

```

datastore = Datastore.  (workspace = ws,
register_azure_blob_container
register_azure_file_share
register_azure_data_lake
register_azure_sql_database

datastore_name = 'finance_datastore',
container_name = 'finance-data',
account_name = 'fintrainingdatastorage',
account_key = 'FWUYORRv3XoyNe...',


create_if_not_exists = True
create_if_not_exists = False
overwrite = True
overwrite = False

```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: register_azure_blob_container
 Register an Azure Blob Container to the datastore.
 Box 2: create_if_not_exists = False
 Create the file share if it does not exists, defaults to False. Reference:
<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.datastore.datastore>

NEW QUESTION 131

- (Exam Topic 3)

You use the Azure Machine Learning Python SDK to define a pipeline to train a model. The data used to train the model is read from a folder in a datastore. You need to ensure the pipeline runs automatically whenever the data in the folder changes. What should you do?

- A. Set the regenerate_outputs property of the pipeline to True
- B. Create a ScheduleRecurrance object with a Frequency of aut
- C. Use the object to create a Schedule for the pipeline
- D. Create a PipelineParameter with a default value that references the location where the training data is stored
- E. Create a Schedule for the pipelin
- F. Specify the datastore in the datastore property, and the folder containing the training data in the path_on_datascor property

Answer: D

Explanation:

Reference:
<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-trigger-published-pipeline>

NEW QUESTION 132

- (Exam Topic 3)

You use the Azure Machine Learning SDK in a notebook to run an experiment using a script file in an experiment folder. The experiment fails. You need to troubleshoot the failed experiment. What are two possible ways to achieve this goal? Each correct answer presents a complete solution.

- A. Use the get_metrics() method of the run object to retrieve the experiment run logs.
- B. Use the get_details_with_logs() method of the run object to display the experiment run logs.
- C. View the log files for the experiment run in the experiment folder.
- D. View the logs for the experiment run in Azure Machine Learning studio.
- E. Use the get_output() method of the run object to retrieve the experiment run logs.

Answer: BD

Explanation:

Use get_details_with_logs() to fetch the run details and logs created by the run. You can monitor Azure Machine Learning runs and view their logs with the Azure Machine Learning studio. Reference:
<https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.steprun> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-monitor-view-training-logs>

NEW QUESTION 136

- (Exam Topic 3)

You are determining if two sets of data are significantly different from one another by using Azure Machine Learning Studio.

Estimated values in one set of data may be more than or less than reference values in the other set of data. You must produce a distribution that has a constant Type I error as a function of the correlation.

You need to produce the distribution.

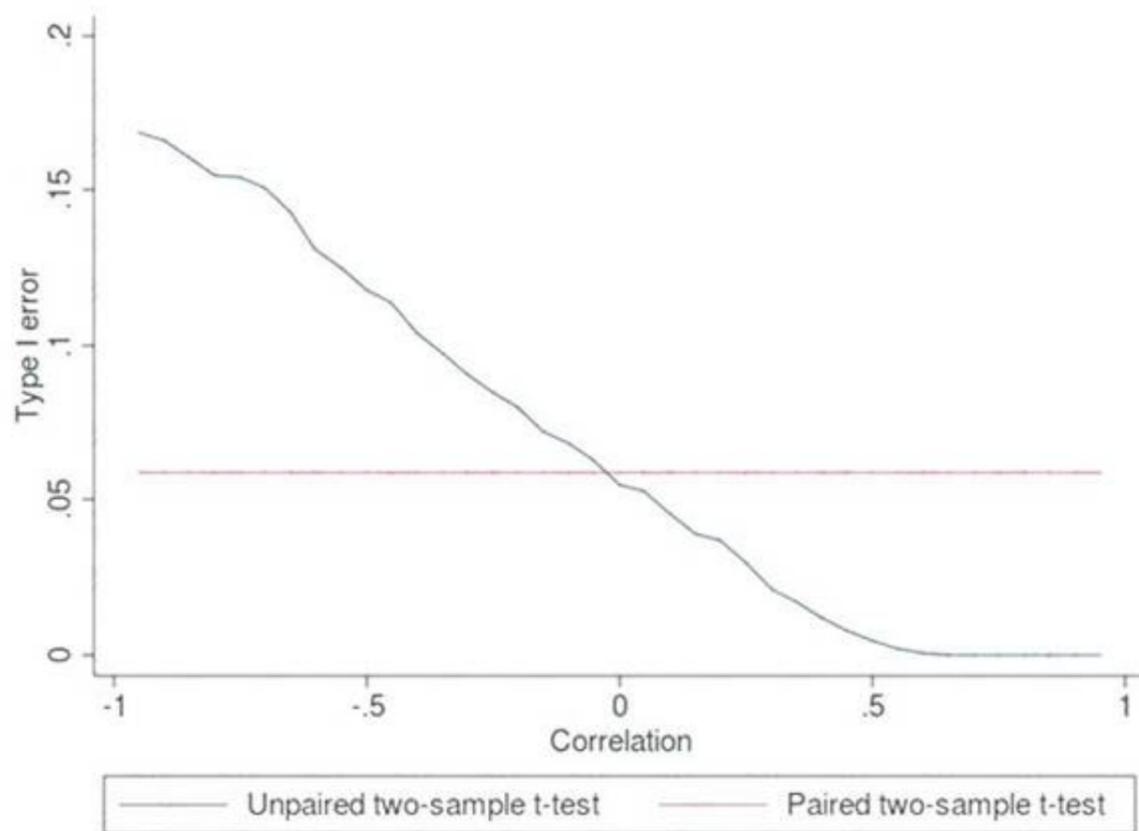
Which type of distribution should you produce?

- A. Paired t-test with a two-tail option
- B. Unpaired t-test with a two tail option
- C. Paired t-test with a one-tail option
- D. Unpaired t-test with a one-tail option

Answer: A

Explanation:

Choose a one-tail or two-tail test. The default is a two-tailed test. This is the most common type of test, in which the expected distribution is symmetric around zero. Example: Type I error of unpaired and paired two-sample t-tests as a function of the correlation. The simulated random numbers originate from a bivariate normal distribution with a variance of 1.



Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/test-hypothesis-using-t-test> https://en.wikipedia.org/wiki/Student%27s_t-test

NEW QUESTION 139

- (Exam Topic 3)

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 create a model to forecast weather conditions based on historical data.

You need to create a pipeline that runs a processing script to load data from a datastore and pass the processed data to a machine learning model training script.

Solution: Run the following code:

```
datastore = ws.get_default_datastore()
data_input = PipelineData("raw_data", datastore=rawdatastore)
data_output = PipelineData("processed_data", datastore=datastore)
process_step = PythonScriptStep(script_name="process.py",
    arguments=["--data_for_train", data_input],
    outputs=[data_output], compute_target=aml_compute,
    source_directory=process_directory)
train_step = PythonScriptStep(script_name="train.py",
    arguments=["--data_for_train", data_input], inputs=[data_output],
    compute_target=aml_compute, source_directory=train_directory)
pipeline = Pipeline(workspace=ws, steps=[process_step, train_step])
```

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Note: Data used in pipeline can be produced by one step and consumed in another step by providing a PipelineData object as an output of one step and an input

of one or more subsequent steps.

Compare with this example, the pipeline train step depends on the process_step_output output of the pipeline process step:

```
from azureml.pipeline.core import Pipeline, PipelineData
from azureml.pipeline.steps import PythonScriptStep
datastore = ws.get_default_datastore()
process_step_output = PipelineData("processed_data", datastore=datastore)
process_step = PythonScriptStep(script_name="process.py",
arguments=["--data_for_train", process_step_output],
outputs=[process_step_output],
compute_target=aml_compute,
source_directory=process_directory)
train_step = PythonScriptStep(script_name="train.py",
arguments=["--data_for_train", process_step_output],
inputs=[process_step_output],
compute_target=aml_compute,
source_directory=train_directory)
pipeline = Pipeline(workspace=ws, steps=[process_step, train_step])
```

Reference:
<https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinedata?view=azu>

NEW QUESTION 141

- (Exam Topic 3)

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 create an Azure Machine Learning service datastore in a workspace. The datastore contains the following files:

- /data/2018/Q1 .csv
- /data/2018/Q2.csv
- /data/2018/Q3.csv
- /data/2018/Q4.csv
- /data/2019/Q1.csv

All files store data in the following format: id,M,f2,l

1,1,2,0
 2,1,1,1
 32,10

You run the following code:

```
data_store = Datastore.register_azure_blob_container(workspace=ws,
datastore_name='data_store',
container_name='quarterly_data',
account_name='companydata',
account_key='NRPxk8duxbM3...'
create_if_not_exists=False)
```

You need to create a dataset named training_data and load the data from all files into a single data frame by using the following code:

```
data_frame = training_data.to_pandas_dataframe()
```

Solution: Run the following code:

```
from azureml.core import Dataset
paths = (data_store, 'data/*/*.csv')
training_data = Dataset.Tabular.from_delimited_files(paths)
```

Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

NEW QUESTION 146

- (Exam Topic 3)

You train and register a model by using the Azure Machine Learning SDK on a local workstation. Python 3.6 and Visual Studio Code are installed on the workstation.

When you try to deploy the model into production as an Azure Kubernetes Service (AKS)-based web service, you experience an error in the scoring script that causes deployment to fail.

You need to debug the service on the local workstation before deploying the service to production.

Which four 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.

Actions	Answer Area
Create an AksWebservice deployment configuration for the service and deploy the model to it	<div style="display: flex; flex-direction: column; align-items: center;"> <div style="margin-bottom: 10px;">➤</div> </div>
Install Docker on the workstation	
Create a LocalWebservice deployment configuration for the service and deploy the model to it	
Debug and modify the scoring script as necessary. Use the reload() method of the service after each modification	
Create an AciWebservice deployment configuration for the service and deploy the model to it	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Graphical user interface, text, application, email Description automatically generated

Step 1: Install Docker on the workstation

Prerequisites include having a working Docker installation on your local system. Build or download the dockerfile to the compute node.

Step 2: Create an AksWebservice deployment configuration and deploy the model to it

To deploy a model to Azure Kubernetes Service, create a deployment configuration that describes the compute resources needed.

If deploying to a cluster configured for dev/test, ensure that it was created with enough

cores and memory to handle this deployment configuration. Note that memory is also used by

things such as dependencies and AML components.

```
deployment_config = AksWebservice.deploy_configuration(cpu_cores = 1, memory_gb = 1)
```

```
service = Model.deploy(ws, "myservice", [model], inference_config, deployment_config, aks_target) service.wait_for_deployment(show_output = True)
```

```
print(service.state) print(service.get_logs())
```

Step 3: Create a LocalWebservice deployment configuration for the service and deploy the model to it

To deploy locally, modify your code to use LocalWebservice.deploy_configuration() to create a deployment configuration. Then use Model.deploy() to deploy the service.

Step 4: Debug and modify the scoring script as necessary. Use the reload() method of the service after each modification.

During local testing, you may need to update the score.py file to add logging or attempt to resolve any problems that you've discovered. To reload changes to the score.py file, use reload(). For example, the following code reloads the script for the service, and then sends data to it.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-azure-kubernetes-service> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment-local>

NEW QUESTION 147

- (Exam Topic 3)

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.

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()
```

- A. Mastered
- B. Not Mastered

Answer: A

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.

NEW QUESTION 152

- (Exam Topic 3)

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 json, os
from sklearn.metrics import roc_auc_score
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
os.makedirs("outputs", exist_ok = True)
with open("outputs/AUC.txt", "w") as file_cur:
    file_cur.write(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>

NEW QUESTION 157

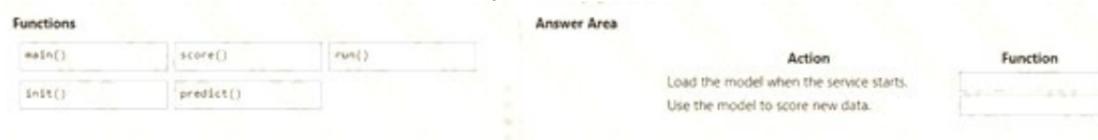
- (Exam Topic 3)

You use Azure Machine Learning to deploy a model as a real-time web service.

You need to create an entry script for the service that ensures that the model is loaded when the service starts and is used to score new data as it is received.

Which functions should you include in the script? To answer, drag the appropriate functions to the correct actions. Each function 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.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: `init()`

The entry script has only two required functions, `init()` and `run(data)`. These functions are used to initialize the service at startup and run the model using request data passed in by a client. The rest of the script handles loading and running the model(s).

Box 2: `run()` Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-existing-model>

NEW QUESTION 158

- (Exam Topic 3)

You are analyzing a dataset containing historical data from a local taxi company. You are developing a regression a regression model.

You must predict the fare of a taxi trip.

You need to select performance metrics to correctly evaluate the regression model. Which two metrics can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. an F1 score that is high
- B. an R Squared value close to 1
- C. an R-Squared value close to 0
- D. a Root Mean Square Error value that is high
- E. a Root Mean Square Error value that is low
- F. an F 1 score that is low.

Answer: BE

Explanation:

References:
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model>

NEW QUESTION 162

- (Exam Topic 3)

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	<ul style="list-style-type: none"> • 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: <ul style="list-style-type: none"> • 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: Install the Azure ML SDK on the Surface Book. Run Python code to connect to the workspace and then run the training script as an experiment on local compute.

- A. Yes
- B. No

Answer: B

Explanation:

Need to attach the mlvm virtual machine as a compute target in the Azure Machine Learning workspace. Reference:
<https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target>

NEW QUESTION 164

- (Exam Topic 3)

A biomedical research company plans to enroll people in an experimental medical treatment trial.

You create and train a binary classification model to support selection and admission of patients to the trial. The model includes the following features: Age, Gender, and Ethnicity.

The model returns different performance metrics for people from different ethnic groups.

You need to use Fairlearn to mitigate and minimize disparities for each category in the Ethnicity feature. Which technique and constraint should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Option	Value
Technique	<div style="border: 1px solid #ccc; padding: 5px;"> <div style="text-align: right; margin-bottom: 5px;">▼</div> <div style="padding: 2px 5px;">Grid search</div> <div style="padding: 2px 5px;">Outlier detection</div> <div style="padding: 2px 5px;">Dimensionality reduction</div> </div>
Constraint	<div style="border: 1px solid #ccc; padding: 5px;"> <div style="text-align: right; margin-bottom: 5px;">▼</div> <div style="padding: 2px 5px;">Demographic parity</div> <div style="padding: 2px 5px;">False-positive rate parity</div> </div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Graphical user interface, text, application, chat or text message Description automatically generated

Box 1: Grid Search

Fairlearn open-source package provides postprocessing and reduction unfairness mitigation algorithms: ExponentiatedGradient, GridSearch, and ThresholdOptimizer.

Note: The Fairlearn open-source package provides postprocessing and reduction unfairness mitigation algorithms types:

- > Reduction: These algorithms take a standard black-box machine learning estimator (e.g., a LightGBM model) and generate a set of retrained models using a sequence of re-weighted training datasets.
- > Post-processing: These algorithms take an existing classifier and the sensitive feature as input.

Box 2: Demographic parity

The Fairlearn open-source package supports the following types of parity constraints: Demographic parity, Equalized odds, Equal opportunity, and Bounded group loss.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-fairness-ml>

NEW QUESTION 169

- (Exam Topic 3)

You must store data in Azure Blob Storage to support Azure Machine Learning. You need to transfer the data into Azure Blob Storage.

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

NOTE: Each correct selection is worth one point.

- A. Bulk Insert SQL Query
- B. AzCopy
- C. Python script
- D. Azure Storage Explorer
- E. Bulk Copy Program (BCP)

Answer: BCD

Explanation:

You can move data to and from Azure Blob storage using different technologies: Azure Storage-Explorer

AzCopy Python SSIS

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/move-azure-blob>

NEW QUESTION 174

- (Exam Topic 3)

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 creating a new experiment in Azure Machine Learning 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. Solution: You use the Scale and Reduce sampling mode.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Instead use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode.

Note: SMOTE is used 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.

References:

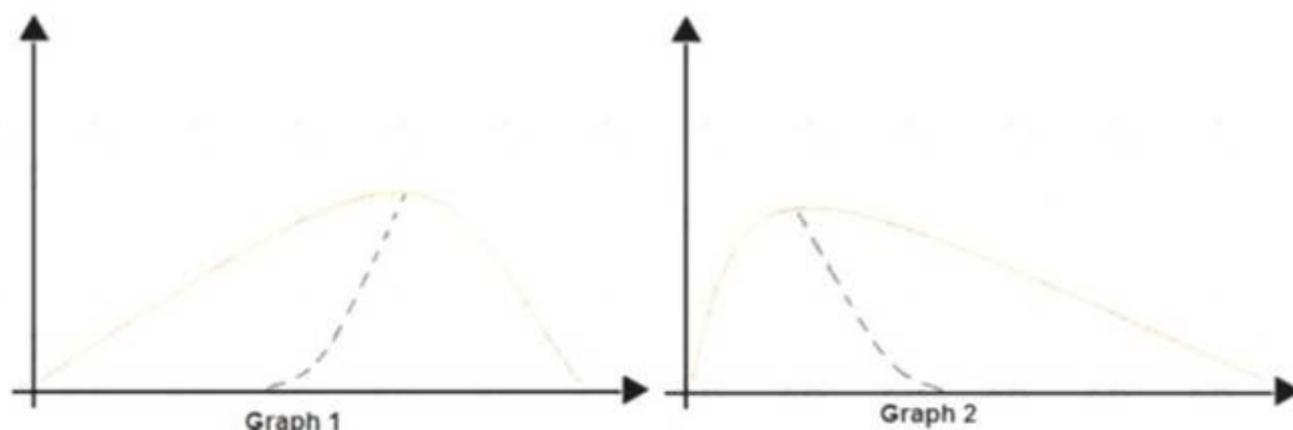
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote>

NEW QUESTION 178

- (Exam Topic 3)

You are analyzing the asymmetry in a statistical distribution.

The following image contains two density curves that show the probability distribution of two datasets.



Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.

NOTE: Each correct selection is worth one point.

Question

Answer choice

Which type of distribution is shown for the dataset density curve of Graph 1?

▼

- Negative skew
- Positive skew
- Normal distribution
- Bimodal distribution

Which type of distribution is shown for the dataset density curve of Graph 2?

▼

- Negative skew
- Positive skew
- Normal distribution
- Bimodal distribution

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Positive skew

Positive skew values means the distribution is skewed to the right. Box 2: Negative skew

Negative skewness values mean the distribution is skewed to the left. References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/compute-elementary-statistic>

NEW QUESTION 182

- (Exam Topic 3)

You are building recurrent neural network to perform a binary classification.

The training loss, validation loss, training accuracy, and validation accuracy of each training epoch has been provided. You need to identify whether the classification model is over fitted.

Which of the following is correct?

- A. The training loss increases while the validation loss decreases when training the model.
- B. The training loss decreases while the validation loss increases when training the model.
- C. The training loss stays constant and the validation loss decreases when training the model.
- D. The training loss .stays constant and the validation loss stays on a constant value and close to the training loss value when training the model.

Answer: B

Explanation:

An overfit model is one where performance on the train set is good and continues to improve, whereas performance on the validation set improves to a point and then begins to degrade.

References:

<https://machinelearningmastery.com/diagnose-overfitting-underfitting-lstm-models/>

NEW QUESTION 184

- (Exam Topic 3)

You train and register a model in your Azure Machine Learning workspace.

You must publish a pipeline that enables client applications to use the model for batch inferencing. You must use a pipeline with a single ParallelRunStep step that runs a Python inferencing script to get predictions from the input data.

You need to create the inferencing script for the ParallelRunStep pipeline step.

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

NOTE: Each correct selection is worth one point.

- A. run(mini_batch) D
- B. main()
- C. batch()
- D. init()
- E. score(mini_batch)

Answer: AD

Explanation:

Reference:

<https://github.com/Azure/MachineLearningNotebooks/tree/master/how-to-use-azureml/machine-learningpipelin>

NEW QUESTION 187

- (Exam Topic 3)

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 create an Azure Machine Learning service datastore in a workspace. The datastore contains the following files:

- /data/2018/Q1.csv
- /data/2018/Q2.csv

- /data/2018/Q3.csv
- /data/2018/Q4.csv
- /data/2019/Q1.csv

All files store data in the following format: id,f1,f2i

```
1,1.2,0
2,1,1,
1 3,2.1,0
```

You run the following code:

```
data_store = Datastore.register_azure_blob_container(workspace=ws,
datastore_name='data_store',
container_name='quarterly_data',
account_name='companydata',
account_key='NRPxk8duxbM3...'
create_if_not_exists=False)
```

You need to create a dataset named training_data and load the data from all files into a single data frame by using the following code:

```
data_frame = training_data.to_pandas_dataframe()
```

Solution: Run the following code:

```
from azureml.core import Dataset
paths = [(data_store, 'data/2018/*.csv'), (data_store, 'data/2019/*.csv')]
training_data = Dataset.File.from_files(paths)
```

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Use two file paths.

Use Dataset.Tabular_from_delimited, instead of Dataset.File.from_files as the data isn't cleansed. Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-register-datasets>

NEW QUESTION 188

- (Exam Topic 3)

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 plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:

```
from azureml.core import Run import pandas as pd
run = Run.get_context()
data = pd.read_csv('data.csv') label_vals = data['label'].unique()
# Add code to record metrics here run.complete()
```

The experiment must record the unique labels in the data as metrics for the run that can be reviewed later.

You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.

Solution: Replace the comment with the following code: for label_val in label_vals:

```
run.log('Label Values', label_val) Does the solution meet the goal?
```

- A. Yes
- B. No

Answer: A

Explanation:

The run_log function is used to log the contents in label_vals: for label_val in label_vals:

```
run.log('Label Values', label_val) Reference:
```

<https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai>

NEW QUESTION 191

- (Exam Topic 3)

You plan to use the Hyperdrive feature of Azure Machine Learning to determine the optimal hyperparameter values when training a model.

You must use Hyperdrive to try combinations of the following hyperparameter values. You must not apply an early termination policy.

learning_rate: any value between 0.001 and 0.1

- batch_size: 16, 32, or 64

You need to configure the sampling method for the Hyperdrive experiment

Which two sampling methods can you use? Each correct answer is a complete solution. NOTE: Each correct selection is worth one point.

- A. Grid sampling
- B. No sampling
- C. Bayesian sampling
- D. Random sampling

Answer: CD

Explanation:

C: Bayesian sampling is based on the Bayesian optimization algorithm and makes intelligent choices on the hyperparameter values to sample next. It picks the

sample based on how the previous samples performed, such that the new sample improves the reported primary metric.

Bayesian sampling does not support any early termination policy Example:

```
from azureml.train.hyperdrive import BayesianParameterSampling from azureml.train.hyperdrive import uniform, choice param_sampling = BayesianParameterSampling( { "learning_rate": uniform(0.05, 0.1), "batch_size": choice(16, 32, 64, 128) } )
```

D: 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.

Reference:

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

NEW QUESTION 194

- (Exam Topic 3)

You are building an experiment using the Azure Machine Learning designer.

You split a dataset into training and testing sets. You select the Two-Class Boosted Decision Tree as the algorithm.

You need to determine the Area Under the Curve (AUC) of the model.

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

Modules

- Export Data
- Tune Model Hyperparameters
- Cross Validate Model
- Evaluate Model
- Score Model
- Train Model

Answer Area

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Train Model

Two-Class Boosted Decision Tree

First, set up the boosted decision tree model.

* 1. Find the Two-Class Boosted Decision Tree module in the module palette and drag it onto the canvas.

* 2. Find the Train Model module, drag it onto the canvas, and then connect the output of the Two-Class Boosted Decision Tree module to the left input port of the Train Model module.

The Two-Class Boosted Decision Tree module initializes the generic model, and Train Model uses training data to train the model.

* 3. Connect the left output of the left Execute R Script module to the right input port of the Train Model

module (in this tutorial you used the data coming from the left side of the Split Data module for training). This portion of the experiment now looks something like this:



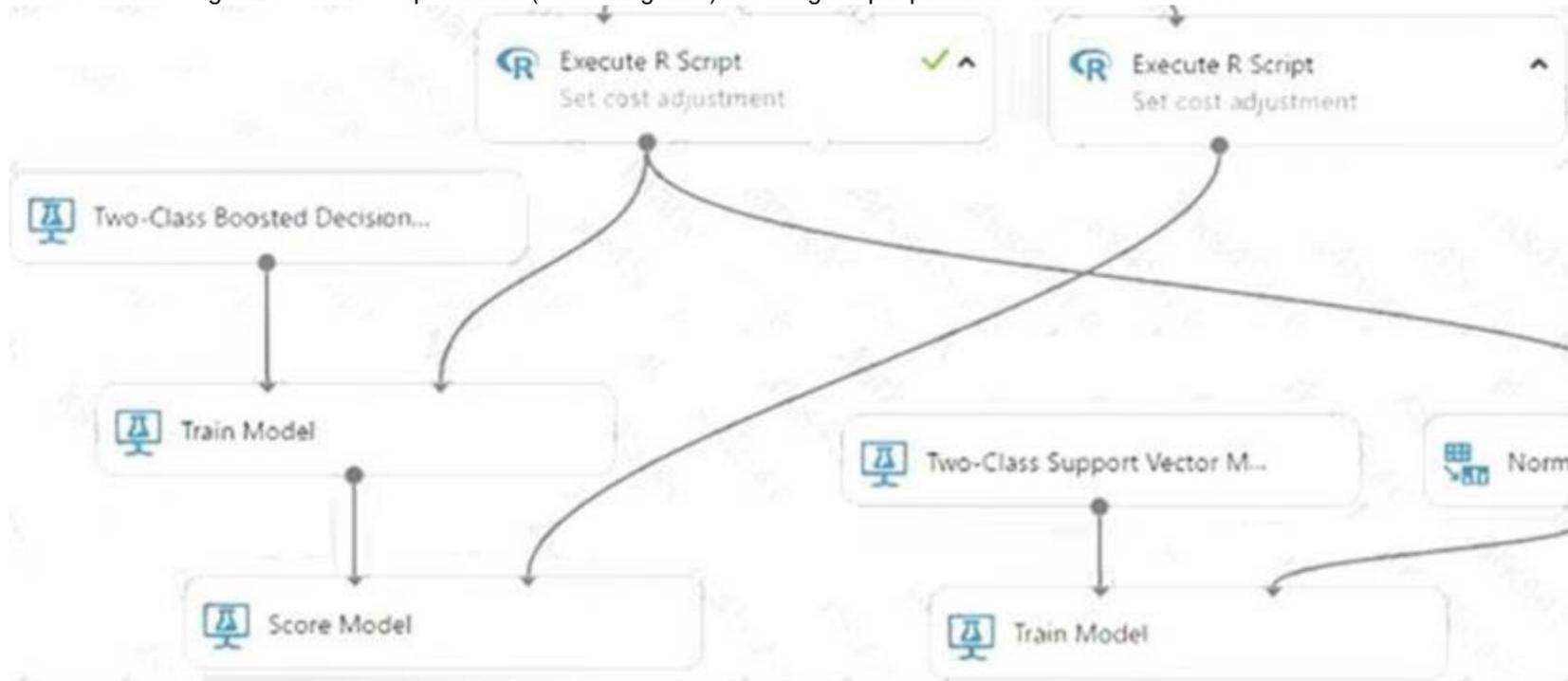
Step 2: Score Model

Score and evaluate the models

You use the testing data that was separated out by the Split Data module to score our trained models. You can then compare the results of the two models to see which generated better results.

Add the Score Model modules

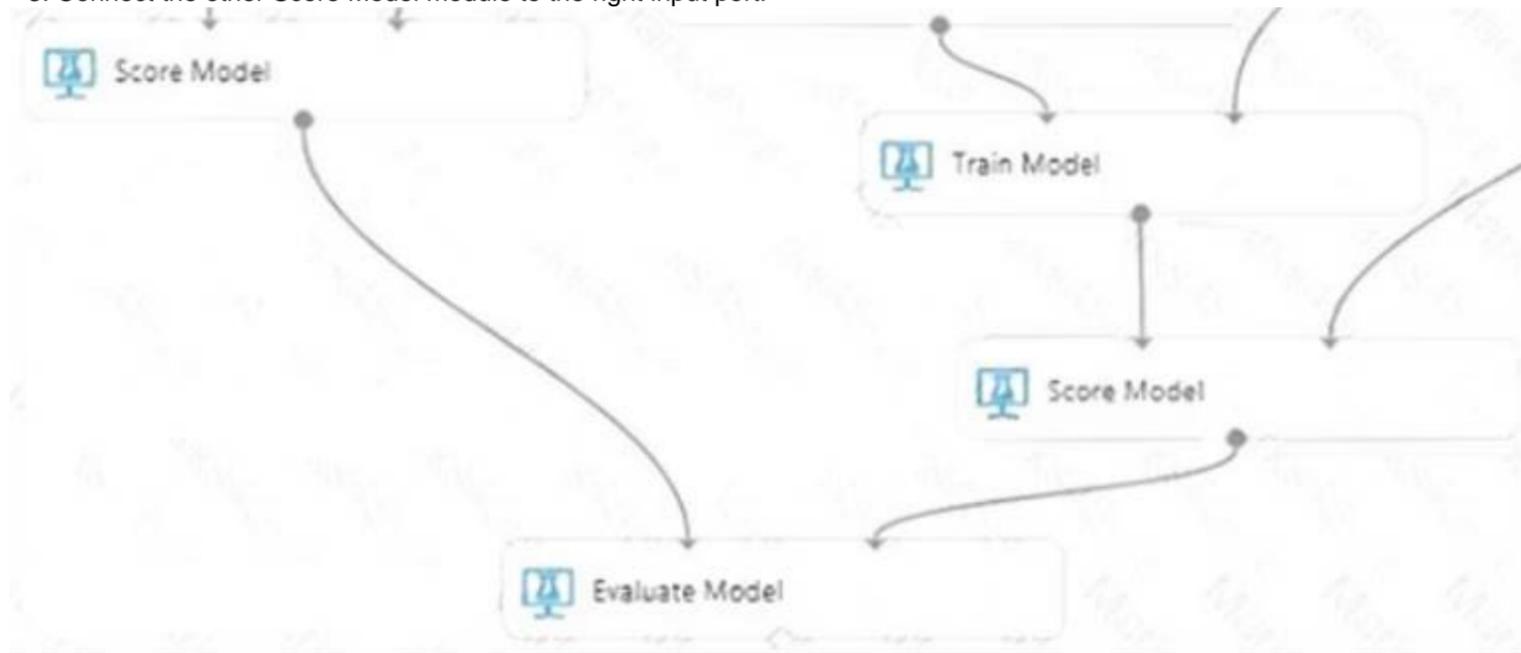
- * 1. Find the Score Model module and drag it onto the canvas.
- * 2. Connect the Train Model module that's connected to the Two-Class Boosted Decision Tree module to the left input port of the Score Model module.
- * 3. Connect the right Execute R Script module (our testing data) to the right input port of the Score Model module.



Step 3: Evaluate Model

To evaluate the two scoring results and compare them, you use an Evaluate Model module.

- * 1. Find the Evaluate Model module and drag it onto the canvas.
- * 2. Connect the output port of the Score Model module associated with the boosted decision tree model to the left input port of the Evaluate Model module.
- * 3. Connect the other Score Model module to the right input port.



NEW QUESTION 199

- (Exam Topic 3)

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 creating a model to predict the price of a student's artwork depending on the following variables: the student's length of education, degree type, and art form.

You start by creating a linear regression model. You need to evaluate the linear regression model.

Solution: Use the following metrics: Accuracy, Precision, Recall, F1 score and AUC. Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Those are metrics for evaluating classification models, instead use: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error, Relative Squared Error, and the Coefficient of Determination.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model>

NEW QUESTION 204

- (Exam Topic 3)

You publish a batch inferencing pipeline that will be used by a business application.

The application developers need to know which information should be submitted to and returned by the REST interface for the published pipeline.

You need to identify the information required in the REST request and returned as a response from the published pipeline.

Which values should you use in the REST request and to expect in the response? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

REST Request	Value
Request Header	JSON containing the run ID JSON containing the pipeline ID JSON containing the experiment name JSON containing an OAuth bearer token
Response	JSON containing the run ID JSON containing the pipeline ID JSON containing the experiment name JSON containing an OAuth bearer token
Response	JSON containing the run ID JSON containing a list of predictions JSON containing the experiment name JSON containing a path to the parallel_run_step.txt output file

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: JSON containing an OAuth bearer token Specify your authentication header in the request.

To run the pipeline from the REST endpoint, you need an OAuth2 Bearer-type authentication header. Box 2: JSON containing the experiment name

Add a JSON payload object that has the experiment name. Example:

```
rest_endpoint = published_pipeline.endpoint response = requests.post(rest_endpoint, headers=auth_header, json={"ExperimentName": "batch_scoring", "ParameterAssignments": {"process_count_per_node": 6}}) run_id = response.json()["Id"]
```

Box 3: JSON containing the run ID

Make the request to trigger the run. Include code to access the Id key from the response dictionary to get the value of the run ID.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-pipeline-batch-scoring-classification>

NEW QUESTION 209

- (Exam Topic 3)

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 train and register a machine learning model.

You plan to deploy the model as a real-time web service. Applications must use key-based authentication to use the model.

You need to deploy the web service.

Solution:

Create an AksWebservice instance.

Set the value of the auth_enabled property to False.

Set the value of the token_auth_enabled property to True.

Deploy the model to the service. Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Instead use only auth_enabled = TRUE Note: Key-based authentication.

Web services deployed on AKS have key-based auth enabled by default. ACI-deployed services have

key-based auth disabled by default, but you can enable it by setting auth_enabled = TRUE when creating the ACI web service. The following is an example of creating an ACI deployment configuration with key-based auth enabled.

```
deployment_config <- aci_webservice_deployment_config(cpu_cores = 1, memory_gb = 1,
```

```
auth_enabled = TRUE) Reference:
```

<https://azure.github.io/azureml-sdk-for-r/articles/deploying-models.html>

NEW QUESTION 214

- (Exam Topic 3)

You configure a Deep Learning Virtual Machine for Windows.

You need to recommend tools and frameworks to perform the following:

- > Build deep neural network (DNN) models
- > Perform interactive data exploration and visualization

Which tools and frameworks should you recommend? To answer, drag the appropriate tools to the correct tasks. Each tool 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.

Tools	Answer Area						
<div style="border: 1px solid black; padding: 2px; margin-bottom: 2px;">Vowpal Wabbit</div> <div style="border: 1px solid black; padding: 2px; margin-bottom: 2px;">PowerBI Desktop</div> <div style="border: 1px solid black; padding: 2px; margin-bottom: 2px;">Azure Data Factory</div> <div style="border: 1px solid black; padding: 2px;">Microsoft Cognitive Toolkit</div>	<table border="1"> <thead> <tr> <th>Task</th> <th>Tool</th> </tr> </thead> <tbody> <tr> <td>Build DNN models</td> <td>Tool</td> </tr> <tr> <td>Enable interactive data exploration and visualization</td> <td>Tool</td> </tr> </tbody> </table>	Task	Tool	Build DNN models	Tool	Enable interactive data exploration and visualization	Tool
Task	Tool						
Build DNN models	Tool						
Enable interactive data exploration and visualization	Tool						

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Vowpal Wabbit

Use the Train Vowpal Wabbit Version 8 module in Azure Machine Learning Studio (classic), to create a machine learning model by using Vowpal Wabbit.

Box 2: PowerBI Desktop

Power BI Desktop is a powerful visual data exploration and interactive reporting tool

BI is a name given to a modern approach to business decision making in which users are empowered to find, explore, and share insights from data across the enterprise.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/train-vowpal-wabbit-version-> <https://docs.microsoft.com/en-us/azure/architecture/data-guide/scenarios/interactive-data-exploration>

NEW QUESTION 216

- (Exam Topic 3)

You are creating a machine learning 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.

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.

```

from sklearn import linear_model
import
    pandas as df
    numpy as df
    scipy as df

dataset = df.read_csv("data\\ProductSales.csv")
ProductCategoryMapping = {"Bikes":1, "Cars":2, "Boats": 3,
"Vans": 4}
dataset['ProductCategoryMapping'] =
dataset['ProductCategory'].
    map[ProductCategoryMapping]
    reduce[ProductCategoryMapping]
    transpose[ProductCategoryMapping]

regr = linear_model.LinearRegression()
X_train = dataset[['ProductCategoryMapping', 'ProductSize',
'ProductCost']]
y_train = dataset[['Sales']]
regr.fit(X_train, y_train)
    
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: pandas as df

Pandas takes data (like a CSV or TSV file, or a SQL database) and creates a Python object with rows and columns called data frame that looks very similar to table in a statistical software (think Excel or SPSS for example).

Box 2: transpose[ProductCategoryMapping] Reshape the data from the pandas Series to columns. Reference: <https://datascienceplus.com/linear-regression-in-python/>

NEW QUESTION 218

- (Exam Topic 3)

```
train_cluster = ComputeTarget(workspace=work_space, name='train-cluster')
estimator = Estimator(source_directory =
    'training-experiment',
    script_params = {'--data-folder': data_source.as_mount(), '--regularization': 0.8},
    compute_target = train_cluster,
    entry_script = 'train.py',
    conda_packages = ['scikit-learn'])
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No. NOTE: Each correct selection is worth one point.

Answer Area

The estimator will look for the files it needs to run an experiment in the training-experiment directory of the local compute environment.

Yes No

The estimator will mount the local data-folder folder and make it available to the script through a parameter.

This is the script The estimator will look for the files it needs to run an experiment in the training-experiment directory of the local compute environment.

The train.py script file will be created if it does not exist.

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

The estimator will look for the files it needs to run an experiment in the training-experiment directory of the local compute environment.

Yes No

The estimator will mount the local data-folder folder and make it available to the script through a parameter.

This is the script The estimator will look for the files it needs to run an experiment in the training-experiment directory of the local compute environment.

The train.py script file will be created if it does not exist.

NEW QUESTION 219

- (Exam Topic 3)

```
source_directory=scripts_folder,
entry_script="batch_pipeline.py",
mini_batch_size="5",
error_threshold=10,
output_action="append_row",
environment=batch_env,
compute_target=compute_target,
logging_level="DEBUG",
node_count=4)
```

You need to obtain the output from the pipeline execution. Where will you find the output?

- A. the Activity Log in the Azure portal for the Machine Learning workspace
- B. a file named parallel_run_step.txt located in the output folder
- C. the digitidentification.py script
- D. the Inference Clusters tab in Machine Learning studio
- E. the debug log

Answer: B

Explanation:

output_action (str): How the output is to be organized. Currently supported values are 'append_row' and 'summary_only'.

> 'append_row' – All values output by run() method invocations will be aggregated into one unique file named parallel_run_step.txt that is created in the output location.

> 'summary_only' Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-contrib-pipeline-steps/azureml.contrib.pipeline.steps.parall>

NEW QUESTION 223

- (Exam Topic 3)

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 train a classification model by using a logistic regression algorithm.

You must be able to explain the model's predictions by calculating the importance of each feature, both as an overall global relative importance value and as a measure of local importance for a specific set of predictions.

You need to create an explainer that you can use to retrieve the required global and local feature importance values.

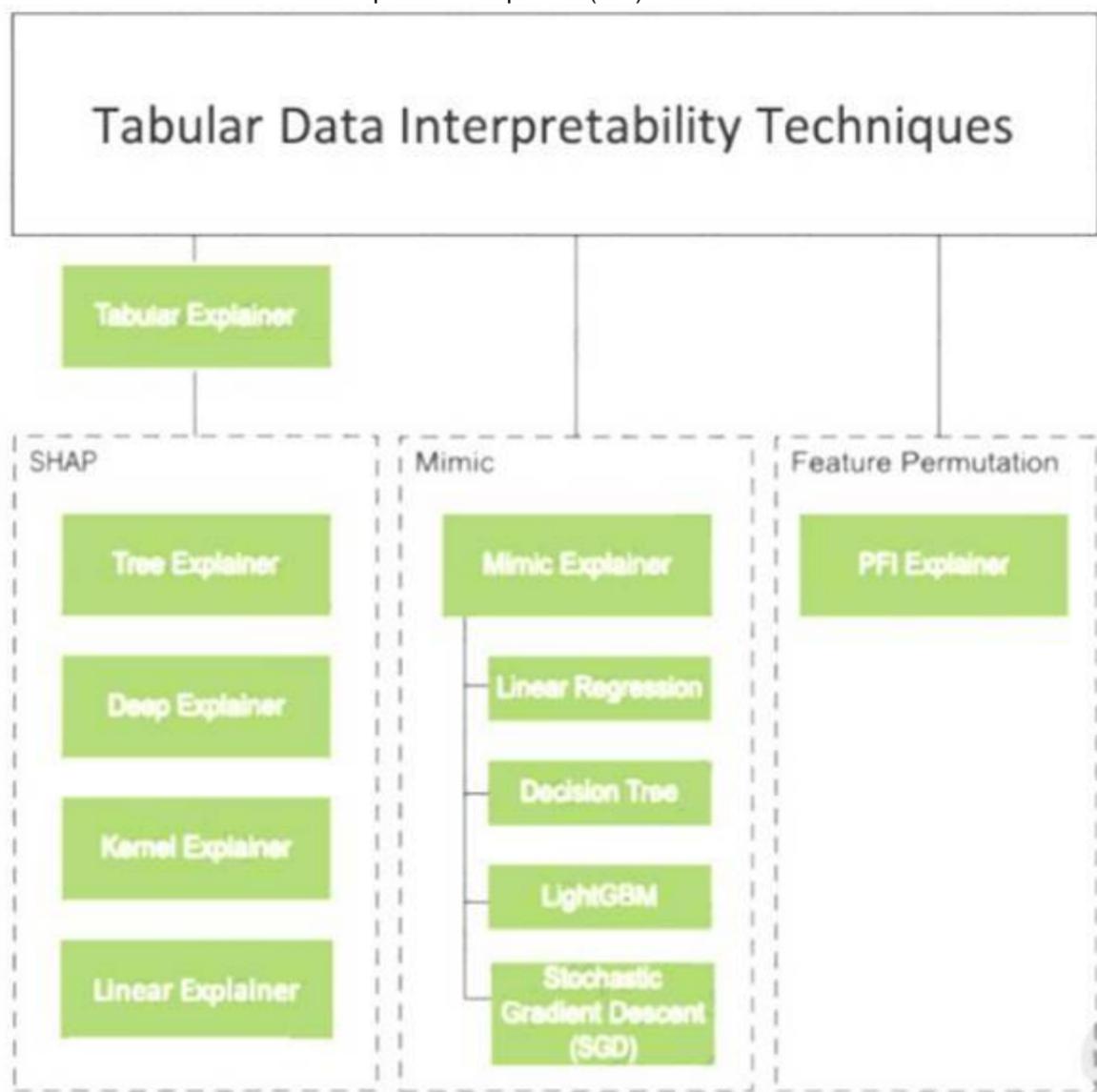
Solution: Create a TabularExplainer. Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Instead use Permutation Feature Importance Explainer (PFI). Note 1:



Note 2: Permutation Feature Importance Explainer (PFI): Permutation Feature Importance is a technique used to explain classification and regression models. At a high level, the way it works is by randomly shuffling data one feature at a time for the entire dataset and calculating how much the performance metric of interest changes. The larger the change, the more important that feature is. PFI can explain the overall behavior of any underlying model but does not explain individual predictions.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability>

NEW QUESTION 228

- (Exam Topic 3)

You create a classification model with a dataset that contains 100 samples with Class A and 10,000 samples with Class B. The variation of Class B is very high. You need to resolve imbalances. Which method should you use?

- A. Partition and Sample
- B. Cluster Centroids
- C. Tomek links
- D. Synthetic Minority Oversampling Technique (SMOTE)

Answer: D

NEW QUESTION 229

- (Exam Topic 3)

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 a data scientist using Azure Machine Learning Studio.

You need to normalize values to produce an output column into bins to predict a target column. Solution: Apply an Equal Width with Custom Start and Stop binning mode.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

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

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins>

NEW QUESTION 234

- (Exam Topic 3)

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 a data scientist using Azure Machine Learning Studio.

You need to normalize values to produce an output column into bins to predict a target column. Solution: Apply a Quantiles normalization with a QuantileIndex normalization.

Does the solution meet the GOAL?

- A. Yes
- B. No

Answer: B

Explanation:

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

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins>

NEW QUESTION 237

- (Exam Topic 3)

You train a classification model by using a decision tree algorithm.

You create an estimator by running the following Python code. The variable `feature_names` is a list of all feature names, and `class_names` is a list of all class names.

```
from interpret.ext.blackbox import TabularExplainer
```

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

You need to explain the predictions made by the model for all classes by determining the importance of all features.

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

NOTE: Each correct selection is worth one point.

	Yes	No
The SHAP TreeExplainer will be used to interpret the model.	<input type="radio"/>	<input type="radio"/>
If you omit the features and classes parameters in the TabularExplainer instantiation, the explainer still works as expected.	<input type="radio"/>	<input type="radio"/>
You could interpret the model by using a MimicExplainer instead of a TabularExplainer.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

s and visualizations more informative, you can choose to pass in feature names and output class names if doing classification.

Box 3: No

TabularExplainer automatically selects the most appropriate one for your use case, but you can call each of its three underlying explainers underneath (TreeExplainer, DeepExplainer, or KernelExplainer) directly.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability-aml>

NEW QUESTION 241

- (Exam Topic 3)

You use Azure Machine Learning Studio to build a machine learning experiment. You need to divide data into two distinct datasets. Which module should you use?

- A. Partition and Sample
- B. Assign Data to Clusters
- C. Group Data into Bins
- D. Test Hypothesis Using t-Test

Answer: A

Explanation:

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

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample>

NEW QUESTION 243

- (Exam Topic 3)

You use the Azure Machine Learning designer to create and run a training pipeline. You then create a real-time inference pipeline. You must deploy the real-time inference pipeline as a web service. What must you do before you deploy the real-time inference pipeline?

- A. Run the real-time inference pipeline.
- B. Create a batch inference pipeline.
- C. Clone the training pipeline.
- D. Create an Azure Machine Learning compute cluster.

Answer: D

Explanation:

You need to create an inferencing cluster. Deploy the real-time endpoint

After your AKS service has finished provisioning, return to the real-time inferencing pipeline to complete deployment.

- > Select Deploy above the canvas.
- > Select Deploy new real-time endpoint.
- > Select the AKS cluster you created.
- > Select Deploy. Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-deploy>

NEW QUESTION 248

- (Exam Topic 3)

You create machine learning models by using Azure Machine Learning.

You plan to train and score models by using a variety of compute contexts. You also plan to create a new compute resource in Azure Machine Learning studio.

You need to select the appropriate compute types.

Which compute types should you select? To answer, drag the appropriate compute types to the correct requirements. Each compute type 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.

Compute types	Answer Area	
	Requirement	Compute type
<div style="border: 1px solid gray; padding: 2px; margin-bottom: 2px; width: fit-content;">Attached compute</div> <div style="border: 1px solid gray; padding: 2px; margin-bottom: 2px; width: fit-content;">Inference cluster</div> <div style="border: 1px solid gray; padding: 2px; margin-bottom: 2px; width: fit-content;">Training cluster</div>	<p>Train models by using the Azure Machine Learning designer.</p> <p>Score new data through a trained model published as a real-time web service.</p> <p>Train models by using an Azure Databricks cluster.</p> <p>Deploy models by using the Azure Machine Learning designer.</p>	<div style="border: 1px solid gray; padding: 2px; margin-bottom: 2px; width: fit-content;">Compute type</div> <div style="border: 1px solid gray; padding: 2px; margin-bottom: 2px; width: fit-content;">Compute type</div> <div style="border: 1px solid gray; padding: 2px; margin-bottom: 2px; width: fit-content;">Compute type</div> <div style="border: 1px solid gray; padding: 2px; margin-bottom: 2px; width: fit-content;">Compute type</div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Attached compute

Training targets	Automated ML	ML pipelines	Azure Machine Learning designer
Local computer	yes		
Azure Machine Learning compute cluster	yes & hyperparameter tuning	yes	yes
Azure Machine Learning compute instance	yes & hyperparameter tuning	yes	yes

Box 2: Inference cluster Box 3: Training cluster Box 4: Attached compute

NEW QUESTION 249

- (Exam Topic 3)

You have the following code. The code prepares an experiment to run a script:

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

ws = Workspace.from_config()
script_config = ScriptRunConfig(source_directory='experiment_files',
                                script='experiment.py')

script_experiment = Experiment(workspace=ws, name='script-experiment')
```

The experiment must be run on local computer using the default environment. You need to add code to start the experiment and run the script. Which code segment should you use?

- A. run = script_experiment.start_logging()
- B. run = Run(experiment=script_experiment)
- C. ws.get_run(run_id=experiment.id)
- D. run = script_experiment.submit(config=script_config)

Answer: D

Explanation:

The experiment class submit method submits an experiment and return the active created run.
 Syntax: submit(config, tags=None, **kwargs) Reference:
<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.experiment.experiment>

NEW QUESTION 253

- (Exam Topic 3)

You plan to preprocess text from CSV files. You load the Azure Machine Learning Studio default stop words list. You need to configure the Preprocess Text module to meet the following requirements:

- > Ensure that multiple related words from a single canonical form.
- > Remove pipe characters from text.
- > Remove words to optimize information retrieval.

Which three options should you select? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Preprocess Text

Language

English

Remove by part of speech

False

Text column to clean

Selected columns:
 Column names: **String, Feature**

Launch column selector

Remove stop words

Lemmatization

Detect sentences

Normalize case to lowercase

Remove numbers

Remove special characters

Remove duplicate characters

Remove email addresses

Remove URLs

Expand verb contractions

Normalize backslashes to slashes

Split tokens on special characters

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Remove stop words

Remove words to optimize information retrieval.

Remove stop words: Select this option if you want to apply a predefined stopword list to the text column. Stop word removal is performed before any other processes.

Box 2: Lemmatization

Ensure that multiple related words from a single canonical form. Lemmatization converts multiple related words to a single canonical form

Box 3: Remove special characters

Remove special characters: Use this option to replace any non-alphanumeric special characters with the pipe | character.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/preprocess-text>

NEW QUESTION 258

- (Exam Topic 3)

You create an Azure Machine Learning 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.

Which values should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

Parameter	Value
Estimator	Estimator
Training compute	SKLearn PyTorch Tensorflow Chainer
Training compute	12 vCPU, 48 GB memory, 96 GB SSD 12 vCPU, 112 GB memory, 680 GB SSD, 2 GPU, 24 GB GPU memory 16 vCPU, 128 GB memory, 160 GB HDD, 80 GB NVME disk (4000 MBps) 44 vCPU, 352 GB memory, 3.4 GHz CPU frequency all cores

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Tensorflow

TensorFlow represents an estimator for training in TensorFlow experiments. Box 2: 12 vCPU, 112 GB memory...,2 GPU,.. Use GPUs for the deep neural network. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.dnn>

NEW QUESTION 259

- (Exam Topic 3)

You create an Azure Machine Learning workspace named workspace1. You assign a custom role to a user of workspace1. The custom role has the following JSON definition:

```
{
  "Name": "MyRole",
  "IsCustom": true,
  "Description": "New custom role description.",
  "Actions": ["*"],
  "NotActions": [
    "Microsoft.MachineLearningServices/workspaces/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "AssignableScopes": [
    "/subscriptions/<subscription_id>/resourceGroups/resourcegroup1/providers/Microsoft.MachineLearningServices/workspaces/workspace1"
  ]
}
```

Instructions: For each of the following statements, select Yes if the statement is true. Otherwise, select No.
 NOTE: Each correct selection is worth one point.

Statements	Yes	No
The user can perform all actions in the workspace	<input type="radio"/>	<input type="radio"/>
The user can delete a compute resource in the workspace	<input type="radio"/>	<input type="radio"/>
The user can write metrics to the workspace	<input type="radio"/>	<input type="radio"/>

A.

Answer:

Explanation:

Graphical user interface, text, application, email Description automatically generated

Box 1: No

The actions listed in NotActions are prohibited.

If the roles include Actions that have a wildcard (*), the effective permissions are computed by subtracting the NotActions from the allowed Actions.

Box 2: No

Deleting compute resources in the workspace is in the NotActions list. Box 3: Yes
Writing metrics is not listed in NotActions. Reference:
<https://docs.microsoft.com/en-us/azure/role-based-access-control/overview#how-azure-rbac-determines-if-a-use>

NEW QUESTION 262

- (Exam Topic 3)

You have a Jupyter Notebook that contains Python code that is used to train a model.
You must create a Python script for the production deployment. The solution must minimize code maintenance.
Which two actions should you perform? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

- A. Refactor the Jupyter Notebook code into functions
- B. Save each function to a separate Python file
- C. Define a main() function in the Python script
- D. Remove all comments and functions from the Python script

Answer: AC

Explanation:

Reference:

<https://www.guru99.com/learn-python-main-function-with-examples-understand-main.html> <https://towardsdatascience.com/from-jupyter-notebook-to-deployment-a-straightforward-example-1838c203a43>

NEW QUESTION 263

- (Exam Topic 3)

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. HDInsight
- B. Azure Databricks
- C. Azure Kubernetes Services
- D. the existing Machine Learning Compute resource
- E. a new Machine Learning Compute resource

Answer: C

Explanation:

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

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

NEW QUESTION 265

- (Exam Topic 3)

You have a comma-separated values (CSV) file containing data from which you want to train a classification model.
You are using the Automated Machine Learning interface in Azure Machine Learning studio to train the classification model. You set the task type to Classification.
You need to ensure that the Automated Machine Learning process evaluates only linear models. What should you do?

- A. Add all algorithms other than linear ones to the blocked algorithms list.
- B. Set the Exit criterion option to a metric score threshold.
- C. Clear the option to perform automatic featurization.
- D. Clear the option to enable deep learning.
- E. Set the task type to Regression.

Answer: C

Explanation:

Automatic featurization can fit non-linear models. Reference: <https://econml.azurewebsites.net/spec/estimation/dml.html>

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-automated-ml-for-ml-models>

NEW QUESTION 268

- (Exam Topic 3)

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: BC

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>

NEW QUESTION 273

- (Exam Topic 3)

You are developing a linear regression model in Azure Machine Learning Studio. You run an experiment to compare different algorithms. The following image displays the results dataset output:

Algorithm	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error
Bayesian Linear	3.276025	4.655442	0.511436	0.282138
Neural Network	2.676538	3.621476	0.417847	0.17073
Boosted Decision Tree	2.168847	2.878077	0.338589	0.107831
Linear	6.350005	8.720718	0.99133	0.99002
Decision Forest	2.390206	3.315164	0.373146	0.14307

Use the drop-down menus to select the answer choice that answers each question based on the information presented in the image.

NOTE: Each correct selection is worth one point.

Question

Answer choice

Which algorithm minimizes differences between actual and predicted values?

▼

- Bayesian Linear Regression
- Neural Network Regression
- Boosted Decision Tree Regression
- Linear Regression
- Decision Forest Regression

Which approach should you use to find the best parameters for a Linear Regression model for the Online Gradient Descent method?

▼

- Set the Decrease learning rate option to True.
- Set the Decrease learning rate option to True.
- Set the Create trainer mode option to Parameter Range.
- Increase the number of epochs.
- Decrease the number of epochs.

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Boosted Decision Tree Regression

Mean absolute error (MAE) measures how close the predictions are to the actual outcomes; thus, a lower score is better.

Box 2:

Online Gradient Descent: If you want the algorithm to find the best parameters for you, set Create trainer mode option to Parameter Range. You can then specify multiple values for the algorithm to try.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/linear-regression>

NEW QUESTION 274

.....

Relate Links

100% Pass Your DP-100 Exam with Exam Bible Prep Materials

<https://www.exambible.com/DP-100-exam/>

Contact us

We are proud of our high-quality customer service, which serves you around the clock 24/7.

Viste - <https://www.exambible.com/>