DP-100 Q&As

Designing and Implementing a Data Science Solution on Azure

Pass Microsoft DP-100 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.geekcert.com/dp-100.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Microsoft Official Exam Center

- **Instant Download** After Purchase
- **100% Money Back** Guarantee
- **365 Days** Free Update
- **800,000+** Satisfied Customers

QUESTION 1

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

Which Azure tool should you use?

A. Anaconda Data Science Platform

B. Azure BatchAI

C. Azure Notebooks

D. Azure Machine Learning Service

Correct Answer: C

References: https://notebooks.azure.com/

QUESTION 2

You need to implement a model development strategy to determine a user\\'s tendency to respond to an ad. Which technique should you use?

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

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

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

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

Correct Answer: A

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

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

Relative Expression Split: Use this option whenever you want to apply a condition to a number column. The number could be a date/time field, a column containing age or dollar amounts, or even a percentage. For example, you might want to

divide your data set depending on the cost of the items, group people by age ranges, or separate data by a calendar date.

Scenario:

Local market segmentation models will be applied before determining a user\\'s propensity to respond to an advertisement.

The distribution of features across training and production data are not consistent

References:

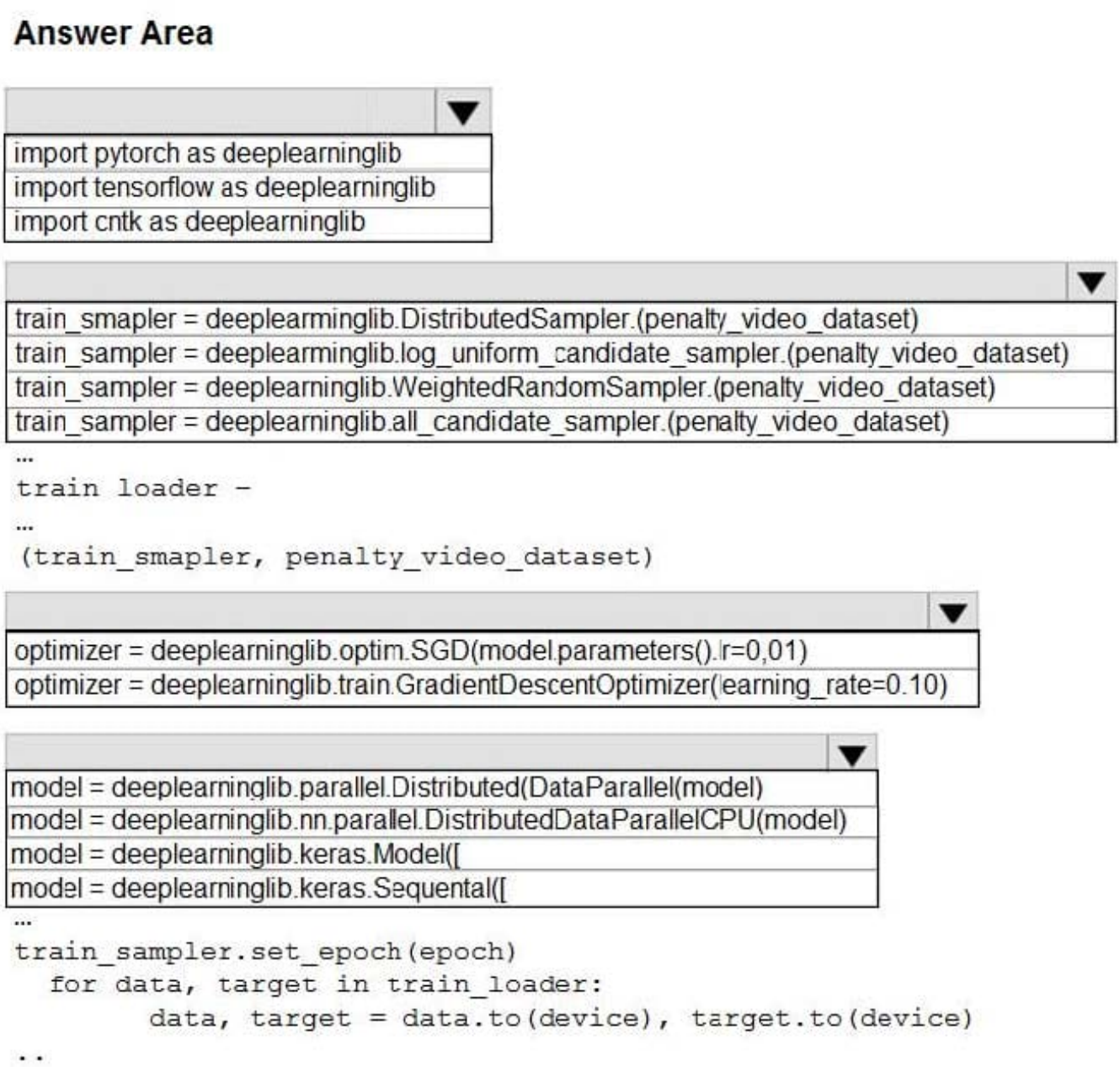
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data

QUESTION 3

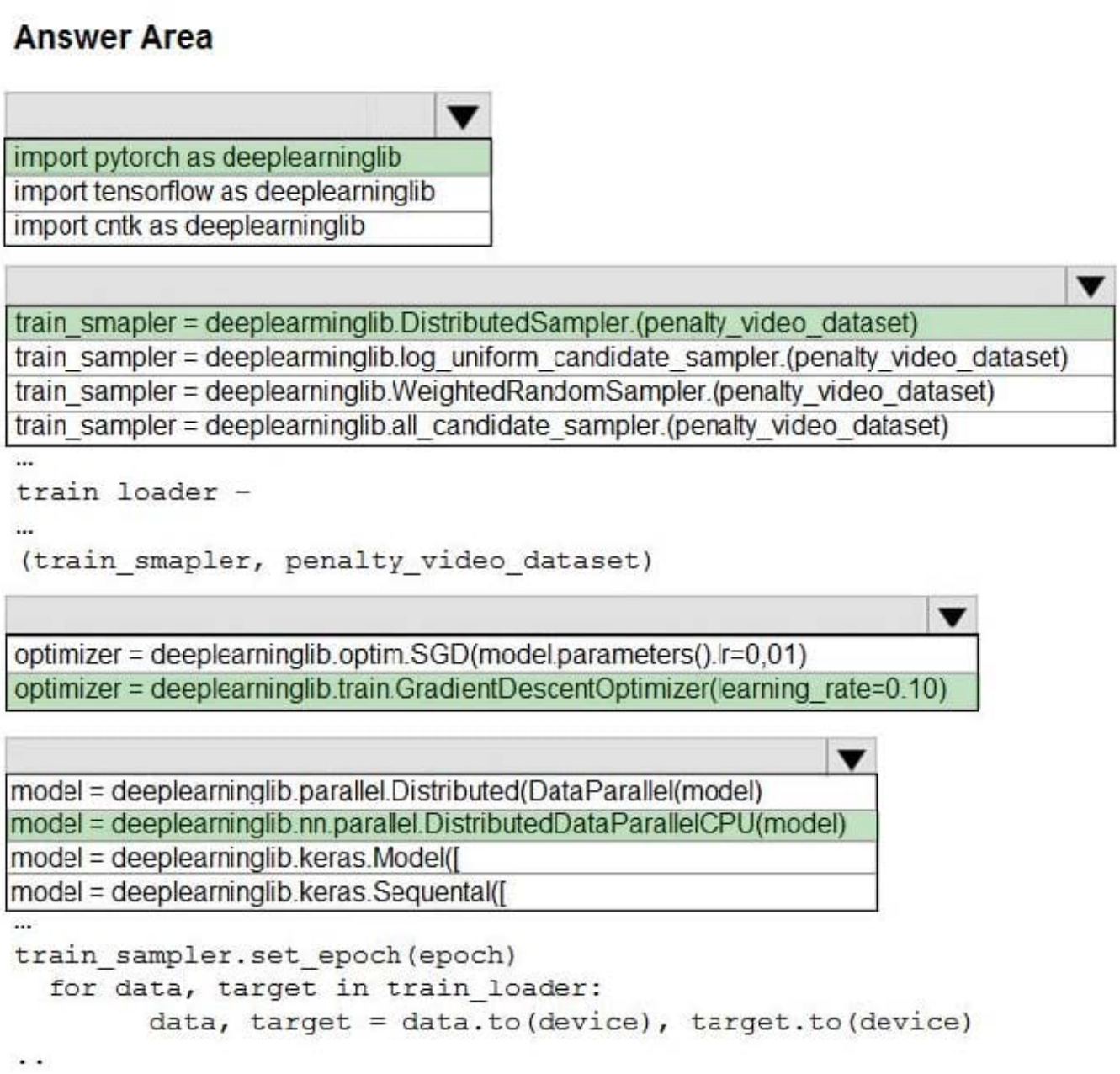
You need to use the Python language to build a sampling strategy for the global penalty detection models. How should you complete the code segment? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area



Correct Answer:



Box 1: import pytorch as deeplearninglib

Box 2: ..DistributedSampler(Sampler)..

DistributedSampler(Sampler):

Sampler that restricts data loading to a subset of the dataset.

It is especially useful in conjunction with class:`torch.nn.parallel.DistributedDataParallel`. In such case, each process can pass a DistributedSampler instance as a DataLoader sampler, and load a subset of the original dataset that is exclusive

to it.

Scenario: Sampling must guarantee mutual and collective exclusively between local and global segmentation models that share the same features.

Box 3: optimizer = deeplearninglib.train. GradientDescentOptimizer(learning_rate=0.10)

Incorrect Answers: ..SGD..

Scenario: All penalty detection models show inference phases using a Stochastic Gradient Descent (SGD) are running too slow.

Box 4: .. nn.parallel.DistributedDataParallel.. DistributedSampler(Sampler): The sampler that restricts data loading to a subset of the dataset. It is especially useful in conjunction with :class:`torch.nn.parallel.DistributedDataParallel`.

References: https://github.com/pytorch/pytorch/blob/master/torch/utils/data/distributed.py

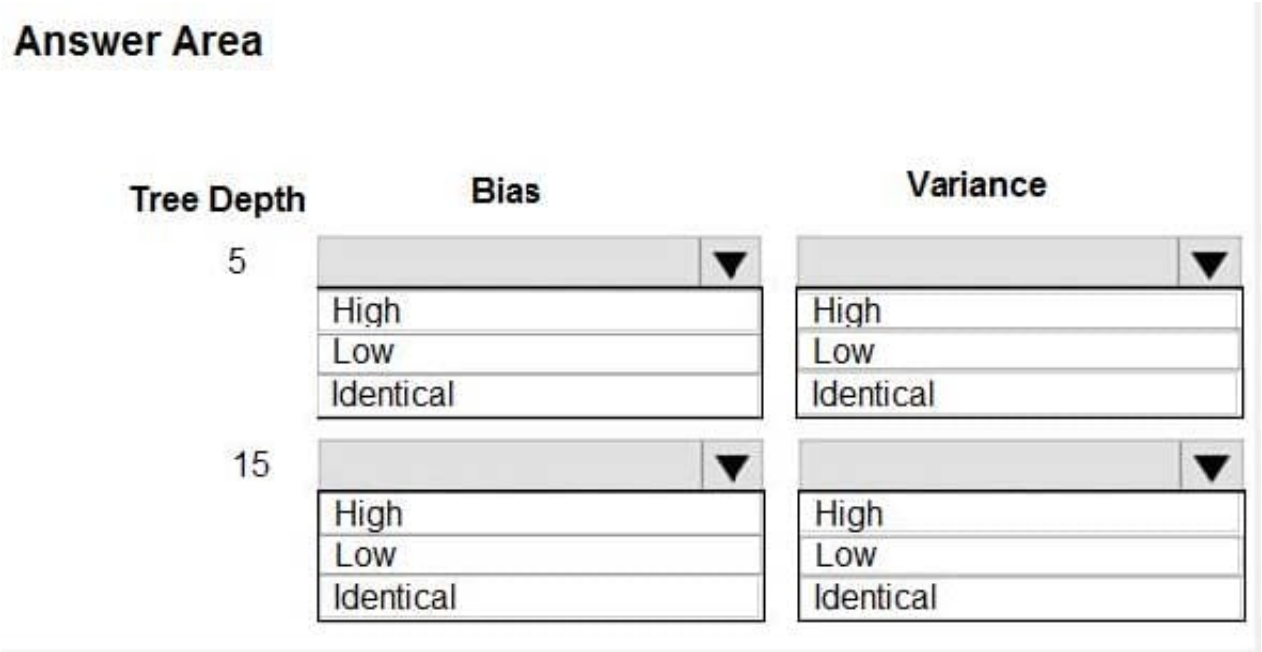
QUESTION 4

You are using a decision tree algorithm. You have trained a model that generalizes well at a tree depth equal to 10.

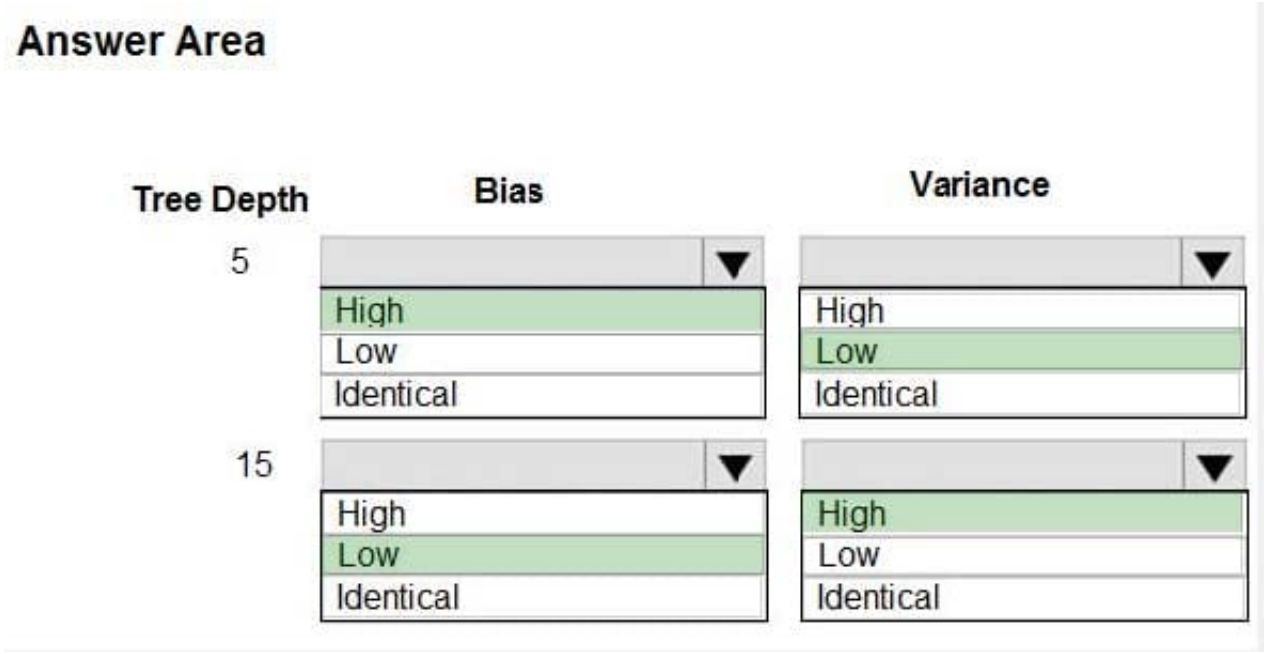
You need to select the bias and variance properties of the model with varying tree depth values.

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

Hot Area:



Correct Answer:



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/

QUESTION 5

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 Principal Components Analysis (PCA) sampling mode.

Does the solution meet the goal?

A. Yes

B. No

Correct Answer: B

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

Note: SMOTE is used to increase the number of underepresented 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.

Incorrect Answers:

The Principal Component Analysis module in Azure Machine Learning Studio (classic) is used to reduce the dimensionality of your training data. The module analyzes your data and creates a reduced feature set that captures all the

information contained in the dataset, but in a smaller number of features.

References:

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

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