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)
You use the Two-Class Neural Network module in Azure Machine Learning Studio to build a binary classification model. You use the Tune Model Hyperparameters module to tune accuracy for the model.
You need to select the hyperparameters that should be tuned using the Tune Model Hyperparameters module. Which two hyperparameters should you use? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Number of hidden nodes
B. Learning Rate
C. The type of the normalizer
D. Number of learning iterations
E. Hidden layer specification

Answer: DE

Explanation:
D: For Number of learning iterations, specify the maximum number of times the algorithm should process the training cases.
E: For Hidden layer specification, select the type of network architecture to create.
Between the input and output layers you can insert multiple hidden layers. Most predictive tasks can be accomplished easily with only one or a few hidden layers.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-neural-network

NEW QUESTION 2
- (Exam Topic 3)
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

Answer: C

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

NEW QUESTION 3
- (Exam Topic 3)
You are building a binary classification model by using a supplied training set. The training set is imbalanced between two classes.
You need to resolve the data imbalance.
What are three possible ways to achieve this goal? Each correct answer presents a complete solution NOTE: Each correct selection is worth one point.

A. Penalize the classification
B. Resample the data set using under sampling or oversampling
C. Generate synthetic samples in the minority class.
D. Use accuracy as the evaluation metric of the model.
E. Normalize the training feature set.

Answer: BCD

NEW QUESTION 4
- (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.

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 5
- (Exam Topic 3)
You are creating a machine learning model. You have a dataset that contains null rows.
You need to use the Clean Missing Data module in Azure Machine Learning Studio to identify and resolve the
null and missing data in the dataset. Which parameter should you use?

A. Replace with mean
B. Remove entire column
C. Remove entire row
D. Hot Deck

Answer: B

Explanation:
Remove entire row: Completely removes any row in the dataset that has one or more missing values. This is useful if the missing value can be considered randomly missing.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

NEW QUESTION 6
- (Exam Topic 2)
You need to select a feature extraction method. Which method should you use?

A. Mutual information
B. Mood's median test
C. Kendall correlation
D. Permutation Feature Importance

Answer: C

Explanation:
In statistics, the Kendall rank correlation coefficient, commonly referred to as Kendall's tau coefficient (after the Greek letter), is a statistic used to measure the ordinal association between two measured quantities.
It is a supported method of the Azure Machine Learning Feature selection.
Scenario: When you train a Linear Regression module using a property dataset that shows data for property prices for a large city, you need to determine the best features to use in a model. You can choose standard metrics provided to measure performance before and after the feature importance process completes. You must ensure that the distribution of the features across multiple training models is consistent.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/feature-selection-modules

NEW QUESTION 7
- (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 rwur.il 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.

A. Mastered
B. Not Mastered

Answer: A

Explanation:

NEW QUESTION 8
- (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 9
- (Exam Topic 2)
You need to set up the Permutation Feature Importance module according to the model training requirements.
Which properties should you select? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Accuracy
Scenario: You want to configure hyperparameters in the model learning process to speed the learning phase by using hyperparameters. In addition, this configuration should cancel the lowest performing runs at each evaluation interval, thereby directing effort and resources towards models that are more likely to be successful.

Box 2: R-Squared

NEW QUESTION 10
- (Exam Topic 2)
You need to identify the methods for dividing the data according, to the testing requirements.
Which properties should you select? To answer, select the appropriate option-, m the answer area. NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:

NEW QUESTION 10
- (Exam Topic 2)
You need to correct the model fit issue.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Step 1: Augment the data
Scenario: Columns in each dataset contain missing and null values. The datasets also contain many outliers.
Step 2: Add the Bayesian Linear Regression module.
Scenario: You produce a regression model to predict property prices by using the Linear Regression and Bayesian Linear Regression modules.
Step 3: Configure the regularization weight.
Regularization typically is used to avoid overfitting. For example, in L2 regularization weight, type the value to use as the weight for L2 regularization. We recommend that you use a non-zero value to avoid overfitting.
Scenario:
Model fit: The model shows signs of overfitting. You need to produce a more refined regression model that reduces the overfitting.

NEW QUESTION 12

- (Exam Topic 2)
You need to replace the missing data in the AccessibilityToHighway columns.
How should you configure the Clean Missing Data module? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Replace using MICE
Replace using MICE: For each missing value, this option assigns a new value, which is calculated by using a method described in the statistical literature as "Multivariate Imputation using Chained Equations" or "Multiple Imputation by Chained Equations". With a multiple imputation method, each variable with missing data is modeled conditionally using the other variables in the data before filling in the missing values.
Scenario: The AccessibilityToHighway column in both datasets contains missing values. The missing data must be replaced with new data so that it is modeled conditionally using the other variables in the data before filling in the missing values.
Box 2: Propagate
Cols with all missing values indicate if columns of all missing values should be preserved in the output. References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

NEW QUESTION 15
- (Exam Topic 1)
You need to resolve the local machine learning pipeline performance issue. What should you do?

A. Increase Graphic Processing Units (GPUs).
B. Increase the learning rate.
C. Increase the training iterations,
D. Increase Central Processing Units (CPUs).

Answer: A

NEW QUESTION 16
- (Exam Topic 1)
You need to implement a scaling strategy for the local penalty detection data. Which normalization type should you use?

A. Streaming
B. Weight
C. Batch
D. Cosine

Answer: C

Explanation:
Post batch normalization statistics (PBN) is the Microsoft Cognitive Toolkit (CNTK) version of how to evaluate the population mean and variance of Batch Normalization which could be used in inference Original Paper.
In CNTK, custom networks are defined using the BrainScriptNetworkBuilder and described in the CNTK network description language "BrainScript."
Scenario:
Local penalty detection models must be written by using BrainScript. References:
https://docs.microsoft.com/en-us/cognitive-toolkit/post-batch-normalization-statistics

NEW QUESTION 17
- (Exam Topic 1)
You need to implement a model development strategy to determine a user's tendency to respond to an ad. Which technique should you use?

A. Use a Relative Expression Split module to partition the data based on centroid distance.
B. Use a Relative Expression Split module to partition the data based on distance travelled to the event.
C. Use a Split Rows module to partition the data based on distance travelled to the event.
D. Use a Split Rows module to partition the data based on centroid distance.

Answer: A

Explanation:
Split Data partitions the rows of a dataset into two distinct sets.
The Relative Expression Split option in the Split Data module of Azure Machine Learning Studio is helpful when you need to divide a dataset into training and testing datasets using a numerical expression.
Relative Expression Split: Use this option whenever you want to apply a condition to a number column. The number could be a date/time field, a column containing age or dollar amounts, or even a percentage. For example, you might want to divide your data set depending on the cost of the items, group people by age ranges,

or separate data by a calendar date.
Scenario:
Local market segmentation models will be applied before determining a user's propensity to respond to an advertisement.
The distribution of features across training and production data are not consistent References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data

NEW QUESTION 21
- (Exam Topic 1)
You need to define a process for penalty event detection.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

A. Mastered
B. Not Mastered

Answer: A

Explanation:

NEW QUESTION 24
- (Exam Topic 1)
You need to implement a feature engineering strategy for the crowd sentiment local models. What should you do?

A. Apply an analysis of variance (ANOVA).
B. Apply a Pearson correlation coefficient.
C. Apply a Spearman correlation coefficient.
D. Apply a linear discriminant analysis.

Answer: D

Explanation:
The linear discriminant analysis method works only on continuous variables, not categorical or ordinal variables.
Linear discriminant analysis is similar to analysis of variance (ANOVA) in that it works by comparing the means of the variables.
Scenario:
Data scientists must build notebooks in a local environment using automatic feature engineering and model building in machine learning pipelines.
Experiments for local crowd sentiment models must combine local penalty detection data. All shared features for local models are continuous variables.

NEW QUESTION 26
- (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 28
- (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 33
- (Exam Topic 3)
You are developing a hands-on workshop to introduce Docker for Windows to attendees. You need to ensure that workshop attendees can install Docker on their devices.
Which two prerequisite components should attendees install on the devices? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Microsoft Hardware-Assisted Virtualization Detection Tool
B. Kitematic
C. BIOS-enabled virtualization
D. VirtualBox
E. Windows 10 64-bit Professional

Answer: E

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

E: To run Docker, your machine must have a 64-bit operating system running Windows 7 or higher. References:
https://docs.docker.com/toolbox/toolbox_install_windows/ https://blogs.technet.microsoft.com/canitpro/2015/09/08/step-by-step-enabling-hyper-v-for-use-on-windows-10/

NEW QUESTION 37
- (Exam Topic 3)
You plan to use a Data Science Virtual Machine (DSVM) with the open source deep learning frameworks Caffe2 and Theano. You need to select a pre configured DSVM to support the framework.
What should you create?

A. Data Science Virtual Machine for Linux (CentOS)
B. Data Science Virtual Machine for Windows 2012
C. Data Science Virtual Machine for Windows 2016
D. Geo AI Data Science Virtual Machine with ArcGIS
E. Data Science Virtual Machine for Linux (Ubuntu)

Answer: E

NEW QUESTION 39
- (Exam Topic 3)
You are performing sentiment analysis using a CSV file that includes 12,000 customer reviews written in a short sentence format. You add the CSV file to Azure Machine Learning Studio and configure it as the starting point dataset of an experiment. You add the Extract N-Gram Features from Text module to the experiment to extract key phrases from the customer review column in the dataset.
You must create a new n-gram dictionary from the customer review text and set the maximum n-gram size to trigrams.
What should you select? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Vocabulary mode: Create
For Vocabulary mode, select Create to indicate that you are creating a new list of n-gram features. N-Grams size: 3
For N-Grams size, type a number that indicates the maximum size of the n-grams to extract and store. For example, if you type 3, unigrams, bigrams, and trigrams will be created.
Weighting function: Leave blank
The option, Weighting function, is required only if you merge or update vocabularies. It specifies how terms in the two vocabularies and their scores should be weighted against each other.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/extract-n-gram-features-from

NEW QUESTION 40
- (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 43
- (Exam Topic 3)
You create a binary classification model using Azure Machine Learning Studio.
You must use a Receiver Operating Characteristic (RO C) curve and an F1 score to evaluate the model. You need to create the required business metrics.
How should you complete the experiment? To answer, select the appropriate options in the dialog box in the answer area.
NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:

NEW QUESTION 48
- (Exam Topic 3)
You are using C-Support Vector classification to do a multi-class classification with an unbalanced training dataset. The C-Support Vector classification using Python code shown below:

You need to evaluate the C-Support Vector classification code.
Which evaluation statement should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Automatically adjust weights inversely proportional to class frequencies in the input data
The "balanced" mode uses the values of y to automatically adjust weights inversely proportional to class frequencies in the input data as n_samples / (n_classes * np.bincount(y)).
Box 2: Penalty parameter
Parameter: C : float, optional (default=1.0)
Penalty parameter C of the error term. References:
https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html

NEW QUESTION 49
- (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 54
- (Exam Topic 3)
You are performing a filter based feature selection for a dataset 10 build a multi class classifies by using Azure Machine Learning Studio.
The dataset contains categorical features that are highly correlated to the output label column.
You need to select the appropriate feature scoring statistical method to identify the key predictors. Which method should you use?

A. Chi-squared
B. Spearman correlation
C. Kendall correlation
D. Person correlation

Answer: D

Explanation:
Pearson's correlation statistic, or Pearson's correlation coefficient, is also known in statistical models as the r value. For any two variables, it returns a value that indicates the strength of the correlation
Pearson's correlation coefficient is the test statistics that measures the statistical relationship, or association, between two continuous variables. It is known as the best method of measuring the association between variables of interest because it is based on the method of covariance. It gives information about the magnitude of the association, or correlation, as well as the direction of the relationship.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/filter-based-feature-selection https://www.statisticssolutions.com/pearsons-correlation-coefficient/

NEW QUESTION 55
- (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 60
- (Exam Topic 3)
You are building a recurrent neural network to perform a binary classification. You review the training loss, validation loss, training accuracy, and validation accuracy for each training epoch.
You need to analyze model performance.
Which observation indicates that the classification model is over fitted?

A. 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.
B. The training loss increases while the validation loss decreases when training the model.
C. The training loss decreases while the validation loss increases when training the model.
D. The training loss stays constant and the validation loss decreases when training the model.

Answer: B

NEW QUESTION 62
- (Exam Topic 3)
You are a data scientist creating a linear regression model.
You need to determine how closely the data fits the regression line. Which metric should you review?

A. Coefficient of determination
B. Recall
C. Precision
D. Mean absolute error
E. Root Mean Square Error

Answer: A

Explanation:
Coefficient of determination, often referred to as R2, represents the predictive power of the model as a value between 0 and 1. Zero means the model is random (explains nothing); 1 means there is a perfect fit. However, caution should be used in interpreting R2 values, as low values can be entirely normal and high values can be suspect.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model

NEW QUESTION 66
- (Exam Topic 3)
You need to select a pre built development environment for a series of data science experiments. You must use the R language for the experiments.
Which three environments can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

A. MI.NET Library on a local environment
B. Azure Machine Learning Studio
C. Data Science Virtual Machine (OSVM)
D. Azure Data bricks
E. Azure Cognitive Services

Answer: ABD

NEW QUESTION 68
- (Exam Topic 3)
You are evaluating a completed binary classification machine. You need to use the precision as the evaluation metric.
Which visualization should you use?

A. scatter plot
B. coefficient of determination
C. Receiver Operating Characteristic CROC) curve
D. Gradient descent

Answer: C

NEW QUESTION 69
- (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.

A. Mastered
B. Not Mastered

Answer: A

Explanation:

NEW QUESTION 70
- (Exam Topic 3)
You are developing deep learning models to analyze semi-structured, unstructured, and structured data types. You have the following data available for model building:
 Video recordings of sporting events
 Transcripts of radio commentary about events
 Logs from related social media feeds captured during sporting events You need to select an environment for creating the model.
Which environment should you use?

A. Azure Cognitive Services
B. Azure Data Lake Analytics
C. Azure HDInsight with Spark MLib
D. Azure Machine Learning Studio

Answer: A

Explanation:
Azure Cognitive Services expand on Microsoft's evolving portfolio of machine learning APIs and enable developers to easily add cognitive features – such as emotion and video detection; facial, speech, and vision recognition; and speech and language understanding – into their applications. The goal of Azure Cognitive Services is to help developers create applications that can see, hear, speak, understand, and even begin to reason. The catalog of services within Azure Cognitive Services can be categorized into five main pillars - Vision, Speech, Language, Search, and Knowledge.
References:
https://docs.microsoft.com/en-us/azure/cognitive-services/welcome

NEW QUESTION 72
- (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 76
- (Exam Topic 3)
You are analyzing a raw dataset that requires cleaning.
You must perform transformations and manipulations by using Azure Machine Learning Studio. You need to identify the correct modules to perform the transformations.
Which modules should you choose? To answer, drag the appropriate modules to the correct scenarios. Each module 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: Clean Missing Data
Box 2: SMOTE
Use the SMOTE module in Azure Machine Learning Studio 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.
Box 3: Convert to Indicator Values
Use the Convert to Indicator Values module in Azure Machine Learning Studio. The purpose of this module is to convert columns that contain categorical values into a series of binary indicator columns that can more easily be used as features in a machine learning model.
Box 4: Remove Duplicate Rows 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/convert-to-indicator-values

NEW QUESTION 78
- (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.

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 79
- (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 Studio to perform feature engineering on a dataset. You need to normalize values to produce a feature column grouped into bins.
Solution: Apply an Entropy Minimum Description Length (MDL) binning mode.
Does the solution meet the goal?

A. Yes
B. No

Answer: A

Explanation:
Entropy MDL binning mode: This method requires that you select the column you want to predict and the column or columns that you want to group into bins. It then makes a pass over the data and attempts to determine the number of bins that minimizes the entropy. In other words, it chooses a number of bins that allows the data column to best predict the target column. It then returns the bin number associated with each row of your data in a column named <colname>quantized.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

NEW QUESTION 83
- (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 binning mode with a PQuantile 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 88
- (Exam Topic 3)
You arc I mating a deep learning model to identify cats and dogs. You have 25,000 color images.
You must meet the following requirements:
• Reduce the number of training epochs.
• Reduce the size of the neural network.
• Reduce over-fitting of the neural network.
You need to select the image modification values.
Which value should you use? To answer, select the appropriate Options in the answer area. NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:

NEW QUESTION 90
- (Exam Topic 3)
You are evaluating a Python NumPy array that contains six data points defined as follows: data = [10, 20, 30, 40, 50, 60]
You must generate the following output by using the k-fold algorithm implantation in the Python Scikit-learn machine learning library:
train: [10 40 50 60], test: [20 30]
train: [20 30 40 60], test: [10 50]
train: [10 20 30 50], test: [40 60]
You need to implement a cross-validation to generate the output.
How should you complete the code segment? To answer, select the appropriate code segment in the dialog box in the answer area.
NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: k-fold
Box 2: 3
K-F olds cross-validator provides train/test indices to split data in train/test sets. Split dataset into k consecutive folds (without shuffling by default).
The parameter n_splits (int, default=3) is the number of folds. Must be at least 2. Box 3: data
Example: Example:
>>>
>>> from sklearn.model_selection import KFold
>>> X = np.array([[1, 2], [3, 4], [1, 2], [3, 4]])
>>> y = np.array([1, 2, 3, 4])
>>> kf = KFold(n_splits=2)
>>> kf.get_n_splits(X) 2
>>> print(kf)
KFold(n_splits=2, random_state=None, shuffle=False)
>>> for train_index, test_index in kf.split(X): print("TRAIN:", train_index, "TEST:", test_index) X_train, X_test = X[train_index], X[test_index] y_train, y_test = y[train_index], y[test_index] TRAIN: [2 3] TEST: [0 1]
TRAIN: [0 1] TEST: [2 3]
References:
https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html

NEW QUESTION 92
- (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 95
- (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.

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 96
- (Exam Topic 3)
You plan to use a Deep Learning Virtual Machine (DLVM) to train deep learning models using Compute Unified Device Architecture (CUDA) computations.
You need to configure the DLVM to support CUDA. What should you implement?

A. Intel Software Guard Extensions (Intel SGX) technology
B. Solid State Drives (SSD)
C. Graphic Processing Unit (GPU)
D. Computer Processing Unit (CPU) speed increase by using overclocking
E. High Random Access Memory (RAM) configuration

Answer: C

Explanation:
A Deep Learning Virtual Machine is a pre-configured environment for deep learning using GPU instances. References:
https://azuremarketplace.microsoft.com/en-au/marketplace/apps/microsoft-ads.dsvm-deep-learning

NEW QUESTION 99
- (Exam Topic 3)
You are producing a multiple linear regression model in Azure Machine learning Studio. Several independent variables are highly correlated.
You need to select appropriate methods for conducting elective feature engineering on all the data.
Which three actions should you perform in sequence? To answer, move the appropriate Actions from the list of actions to the answer area and arrange them in the correct order.

A. Mastered
B. Not Mastered

Answer: A

Explanation:

NEW QUESTION 101
- (Exam Topic 3)
You are creating a new experiment in Azure Machine Learning Studio. You have a small dataset that has missing values in many columns. The data does not require the application of predictors for each column. You plan to use the Clean Missing Data module to handle the missing data.
You need to select a data cleaning method. Which method should you use?

A. Synthetic Minority Oversampling Technique (SMOTE)
B. Replace using MICE
C. Replace using; Probabilistic PCA
D. Normalization

Answer: A

NEW QUESTION 103
- (Exam Topic 3)
You create an experiment in Azure Machine Learning Studio- You add a training dataset that contains 10.000 rows. The first 9.000 rows represent class 0 (90 percent). The first 1.000 rows represent class 1 (10 percent).
The training set is unbalanced between two Classes. You must increase the number of training examples for class 1 to 4,000 by using data rows. You add the Synthetic Minority Oversampling Technique (SMOTE) module to the experiment.
You need to configure the module.
Which values should you use? To answer, select the appropriate options in the dialog box in the answer area. NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:

NEW QUESTION 107
- (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:
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.

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 112
......

Relate Links

100% Pass Your DP-100 Exam with Exambible 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/