Question #1 of 15

Question ID: 1208611

In machine learning, out-of-sample error equals:

A) forecast error plus expected error plus regression error.

B) bias error plus variance error plus base error.

C) Standard error plus data error plus prediction error.

Explanation

Out-of-sample error equals bias error plus variance error plus base error. Bias error is the extent to which a model fits the training data. Variance error describes the degree to which a model's results change in response to new data from validation and test samples. Base error comes from randomness in the data.

(Study Session 3, Module 7.1, LOS 7.b)

Related Material

SchweserNotes - Book 1

Question #2 of 15

Question ID: 1208605

Which of the following statements about supervised learning is *most accurate*?

A) Supervised learning requires human intervention in machine learning process.

B) Typical data analytics tasks for supervised learning include classification and prediction.

C) Supervised learning does not differentiate between tag and features.

Explanation

Supervised learning utilizes labeled training data to guide the ML program but does not need "human intervention." Typical data analytics tasks for supervised learning include classification and prediction.

(Study Session 3, Module 7.1, LOS 7.a)

Related Material

SchweserNotes - Book 1

Question #3 of 15

Question ID: 1208604

Question #3 of 15

Question ID: 1208604

The technique in which a machine learns to model a set of output data from a given set of inputs is *best* described as:

A) unsupervised learning. ✗

B) deep learning. ✗

C) supervised learning. ✓

Explanation

Supervised learning is a machine learning technique in which a machine is given labelled input and output data and models the output data based on the input data. In unsupervised learning, a machine is given input data in which to identify patterns and relationships, but no output data to model. Deep learning is a technique to identify patterns of increasing complexity and may use supervised or unsupervised learning.

(Study Session 3, Module 7.1, LOS 7.a)

Related Material

SchweserNotes - Book 1

Question #4 of 15

Question ID: 1208609

A rudimentary way to think of machine learning algorithms is that they:

A) “synthesize the pattern, review the pattern.” ✗

B) “find the pattern, apply the pattern.” ✓

C) “develop the pattern, interpret the pattern.” ✗

Explanation

One elementary way to think of ML algorithms is to "find the pattern, apply the pattern."Machine learning attempts to extract knowledge from large amounts of data by learning from known examples in order to determine an underlying structure in the data. The focus is on generating structure or predictions without human intervention.

(Study Session 3, Module 7.1, LOS 7.a)

Related Material

SchweserNotes - Book 1

Question #5 of 15

Question ID: 1208610

The degree to which a machine learning model retains its explanatory power when predicting out-of-sample is *most* commonly described as:

A) predominance.

B) hegemony.

C) generalization.

Explanation

Generalization describes the degree to which, when predicting out-of-sample, a machine learning model retains its explanatory power.

(Study Session 3, Module 7.1, LOS 7.b)

Related Material

SchweserNotes - Book 1

Question #6 of 15

Question ID: 1208617

The unsupervised machine learning algorithm that reduces highly correlated features into fewer uncorrelated composite variables by transforming the feature covariance matrix *best* describes:

A) principal components analysis

B) k-means clustering

C) hierarchical clustering

Explanation

Principal components analysis (PCA) is an unsupervised machine learning algorithm that reduces highly correlated features into fewer uncorrelated composite variables by transforming the feature covariance matrix. K-means partitions observations into a fixed number (k) of non-overlapping clusters. Hierarchical clustering is an unsupervised iterative algorithm used to build a hierarchy of clusters.

(Study Session 3, Module 7.3, LOS 7.d)

Related Material

SchweserNotes - Book 1

Question #7 of 15

Question ID: 1208616

Dimension reduction is *most likely* to be an example of:

A) unsupervised learning.

B) clustering.

C) supervised learning.

Explanation

Dimension reduction and clustering are examples of unsupervised learning algorithms.

(Study Session 3, Module 7.3, LOS 7.d)

Related Material

SchweserNotes - Book 1

Question #8 of 15

Question ID: 1208606

Which of the following about unsupervised learning is *most accurate*?

A) There is no labeled data.

B) Classification is an example of unsupervised learning algorithm.

C) Unsupervised learning has lower forecasting accuracy as compared to supervised learning.

Explanation

In unsupervised learning, the ML program is not given labeled training data. Instead, inputs are provided without any conclusions about those inputs. In the absence of any tagged data, the program seeks out structure or inter-relationships in the data. Clustering is one example of the output of unsupervised ML program while classification is suited for supervised learning.

(Study Session 3, Module 7.1, LOS 7.a)

Related Material

SchweserNotes - Book 1

Question #9 of 15

Question ID: 1208613

Considering the various supervised machine learning algorithms, a penalized regression where the penalty term is the sum of the absolute values of the regression coefficients best describes:

A) support vector machine (SVM).

B) k-nearest neighbor (KNN).

C) least absolute shrinkage and selection operator (LASSO)

Explanation

LASSO (least absolute shrinkage and selection operator) is a popular type of penalized regression in which the penalty term comprises summing the absolute values of the regression coefficients. The more included features, the larger the penalty will be. The result is that a feature needs to make a sufficient contribution to model fit to offset the penalty from including it.

(Study Session 3, Module 7.2, LOS 7.c)

Related Material

SchweserNotes - Book 1

Question #10 of 15

Question ID: 1208615

What is the appropriate remedy in the presence of excessive number of features in a data set?

A) Unsupervised learning.

B) Dimension reduction.

C) Big data analysis.

Explanation

Big Data refers to very large data sets which may include both structured (e.g. spreadsheet) data and unstructured (e.g. emails, text, or pictures) data and includes a large number of features as well as number of observations. Dimension reduction seeks to remove the noise (i.e., those attributes that do not contain much information) when the number of features in a data set (its dimension) is excessive.

(Study Session 3, Module 7.3, LOS 7.d)

Related Material

SchweserNotes - Book 1

Question #11 of 15

Question ID: 1208612

A random forest is *least likely* to:

A) provide a solution to overfitting problem.

B) reduce signal-to-noise ratio.

C) be a classification tree.

Explanation

Random forest is a collection of randomly generated classification trees from the same data set. A randomly selected subset of features is used in creating each tree and hence each tree is slightly different from the others. Since each tree only uses a subset of features, random forests can mitigate the problem of overfitting. Because errors across different trees tend to cancel each other out, using random forests can increase the signal-to-noise ratio.

(Study Session 3, Module 7.2, LOS 7.c)

Related Material

SchweserNotes - Book 1

Question #12 of 15

Question ID: 1208607

Which supervised learning model is *most appropriate* (1) when the Y-variable is continuous and (2) when the Y-variable is categorical

	Continuous Y-variable	Categorical Y-variable
A)	Classification	Neural Networks
B)	Regression	Classification
C)	Decision trees	Regression

Explanation

When the Y-variable is continuous, the appropriate approach is that of regression (used in a broad, ML context). When the Y-variable is categorical (i.e., belonging to a category or classification) or ordinal (i.e., ordered or ranked), a classification model is used.

(Study Session 3, Module 7.1, LOS 7.a)

Related Material

SchweserNotes - Book 1

Question #13 of 15

Question ID: 1208608

Overfitting is *least likely* to result in:

A) inclusion of noise in the model.

B) higher number of features included in the data set.

C) higher forecasting accuracy in out-of-sample data.

Explanation

Overfitting results when a large number of features (i.e., independent variables) are included in the data sample. The resulting model can use the "noise" in the dependent variables to improve the model fit. Overfitting the model in this way will actually decrease the accuracy of model forecasts on other (out-of-sample) data.

(Study Session 3, Module 7.1, LOS 7.b)

Related Material

SchweserNotes - Book 1

Question #14 of 15

Question ID: 1208614

Considering the various supervised machine learning algorithms, a linear classifier that seeks the optimal hyperplane and is typically used for classification, best describes:

A) classification and regression tree (CART)

B) k-nearest neighbor (KNN).

C) support vector machine (SVM).

Explanation

Support vector machine (SVM) is a linear classifier that aims to seek the optimal hyperplane, i.e. the one that separates the two sets of data points by the maximum margin. SVM is typically used for classification.

(Study Session 3, Module 7.2, LOS 7.c)

Related Material

SchweserNotes - Book 1

Question #15 of 15

Question ID: 1208618

An algorithm that involves an agent that performs actions that will maximize its rewards over time, taking into consideration the constraints of its environment, *best* describes:

A) neural networks.

B) deep learning nets.

C) reinforcement learning.

Explanation

Reinforcement learning algorithms involve an agent that will perform actions that will maximize its rewards over time, taking into consideration the constraints of its environment. Neural networks consist of nodes connected by links; learning takes place in the hidden layer nodes, each of which consists of a summation operator and an activation function. Neural networks with many hidden layers (often more than 20) are known as deep learning nets (DLNs) and used in artificial intelligence.

(Study Session 3, Module 7.3, LOS 7.e)

Related Material

SchweserNotes - Book 1