CHAPTER 3

MACHINE LEARNING

1. (B) **unsupervised learning.**

Explanation

Dimension reduction and clustering are examples of unsupervised learning algorithms.

(Module 3.3, LOS 3.d)

Related Material

SchweserNotes - Book 1

2. (C) **Dimension reduction.**

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.

(Module 3.3, LOS 3.d)

Related Material

SchweserNotes - Book 1

3. (A) **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.

(Module 3.2, LOS 3.c)

Related Material

SchweserNotes - Book 1

Hanna Kowalski is a senior fixed-income portfolio analyst at Czarnaskala BP. Kowalski supervises Lena Nowak, who is a junior analyst.

Over the past several years, Kowalski has become aware that investment firms are increasingly to improve their investment decision making. Kowalski has become particularly interested in machine learning techniques and how they might be applied to investment management applications.

Kowalski has read a number of articles about machine learning in various journals for financial analysts. However, she has only a minimal knowledge of how she might source appropriate model inputs, interpret model outputs, and translate those outputs into investment actions.

Kowalski and Nowak meet to discuss plans for incorporating machine learning into their investment model. Kowalski asks Nowak to research machine learning and report back on the types of investment problems that machine learning can address, how the algorithms work, and what the various terminology means.

After spending a few hours researching the topic, Nowak makes a number of statements to Kowalski on the topics of:

- Classification and regression trees (CART)
- Hierarchical clustering
- Neural networks
- Reinforcement learning (RL) algorithms.

Kowalski is left to work out which of Nowak's statements are fully accurate and which are not.

4. (A) discrete target variable, producing a cardinal tree.

Explanation

Classification and regression trees (CART) are generally applied to predict either a continuous target variable, producing a regression tree, or a categorical target variable, producing a classification tree.

(Module 3.2, LOS 3.c)

Related Material

SchweserNotes - Book 1

5. (B) Bottom-up hierarchical clustering begins with each observation being its own cluster.

Explanation

Agglomerative (bottom-up) hierarchical clustering begins with each observation being its own cluster. Then, the algorithm finds the two closest clusters, and combines them into a new, larger cluster. Hierarchical clustering is an unsupervised iterative algorithm. Divisive (top-down) hierarchical clustering progressively partitions clusters into smaller clusters until each cluster contains only one observation.

(Module 3.3, LOS 3.d)

Related Material

SchweserNotes - Book 1

6. (A) **are effective in tasks with non-linearities and complex interactions among variables.**

Explanation

Neural networks have been successfully applied to solve a variety of problems characterized by non-linearities and complex interactions among variables. Neural networks have three types of layers: an input layer, hidden layers, and an output layer. The hidden layer nodes (not the input layer nodes) each consist of a summation operator and an activation function; these nodes are where learning takes place.

(Module 3.3, LOS 3.e)

Related Material

SchweserNotes - Book 1

7. (C) **take into consideration the constraints of its environment.**

Explanation

The reinforcement learning (RL) algorithm involves an agent that will perform actions that will maximize its rewards over time, taking into consideration the constraints of the environment. Unlike supervised learning, reinforcement learning has neither instantaneous feedback nor direct labeled data for each observation.

(Module 3.3, LOS 3.e)

Related Material

SchweserNotes - Book 1

8. (A) **Regression Classification**

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.

(Module 3.1, LOS 3.a)

Related Material

SchweserNotes - Book 1

9. (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.

(Module 3.1, LOS 3.a)

Related Material

SchweserNotes - Book 1

10. (B) principal components analysis.

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.

(Module 3.3, LOS 3.d)

Related Material

SchweserNotes - Book 1

11. (C) There is no labeled data.

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.

(Module 3.1, LOS 3.a)

Related Material

SchweserNotes - Book 1

12. (B) reduce signal-to-noise ratio.

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 over fitting. Because errors across different trees tend to cancel each other out, using random forests can increase the signal-to-noise ratio.

(Module 3.2, LOS 3.c)

Related Material

SchweserNotes - Book 1

13. (C) bias error plus variance error plus base 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.

(Module 3.1, LOS 3.b)

Related Material

SchweserNotes - Book 1

Joyce Tan manages a medium-sized investment fund at Marina Bay Advisors that specializes in international large cap equities. Over the four years that she has been portfolio manager, Tan has been invested in approximately 40 stocks at time.

Tan has used a number of methodologies to select investment opportunities from the universe of investable stocks. In some cases, Tan uses quantitative measures such as accounting ratios to find the most promising investment candidates. In other cases, her team of analysts suggest investments based on qualitative factors and various investment hypotheses.

Tan begins to wonder if her team could leverage financial technology to make better decisions. Specifically, she has read about various machine learning techniques to extract useful information from large financial datasets, in order to uncover new sources of alpha.

14. (A) continuous.

Explanation

Supervised learning can be divided into two categories: regression and classification. If the target variable is categorical or ordinal (e.g., determining a firm's rating), then it is a classification problem. If the target variable to be predicted is continuous, then the task is one of regression.

(Module 3.1, LOS 3.a)

Related Material

SchweserNotes - Book 1

15. (C) k – 1 samples will be used as training samples.

Explanation

In the K-fold cross-validation technique, the data is shuffled randomly and then divided into k equal sub-samples. One sample is saved to be used as a validation sample, and the other k — 1 samples are used as training samples.

(Module 3.1, LOS 3.b)

Related Material

SchweserNotes - Book 1

16. (B) more accurate and more stable.

Explanation

Ensemble learning, which is a technique of combining the predictions from a number of models, generally results in more accurate and more stable predictions than a single model.

(Module 3.2, LOS 3.c)

Related Material

SchweserNotes - Book 1

17. (C) **Neural networks work well in the presence of non-linearities and complex interactions among variables.**

Explanation

Neural networks have been successfully applied to a variety of investment tasks characterized by non-linearities and complex interactions among variables.

Neural networks with at least three hidden layers are known as deep learning nets (DLNs). Reinforcement learning algorithms use an agent that will maximize its rewards over time, within the constraints of its environment.

(Module 3.3, LOS 3.e)

Related Material

SchweserNotes - Book 1

18. (A) **"find the pattern, apply 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.

(Module 3.1, LOS 3.a)

Related Material

SchweserNotes - Book 1

19. (B) **higher forecasting accuracy in out-of-sample data.**

Explanation

Over fitting 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.

(Module 3.1, LOS 3.b)

Related Material

SchweserNotes - Book 1

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

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.

(Module 3.1, LOS 3.a)

Related Material

SchweserNotes - Book 1

21. (B) **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.

(Module 3.2, LOS 3.c)

Related Material

SchweserNotes - Book 1

22. (A) **generalization.**

Explanation

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

(Module 3.1, LOS 3.b)

Related Material

SchweserNotes - Book 1

23. (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.

(Module 3.3, LOS 3.e)

Related Material

SchweserNotes - Book 1