3 MACHINE LEARNING

1. Dimension reduction is most likely to be an example of:
   (A) supervised learning.
   (B) unsupervised learning.
   (C) clustering

2. What is the appropriate remedy in the presence of excessive number of features in a data set?
   (A) Big data analysis.
   (B) Unsupervised learning.
   (C) Dimension reduction.

3. 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) least absolute shrinkage and selection operator (LASSO).
   (B) k-nearest neighbor (KNN).
   (C) support vector machine (SVM).

4. Nowak first tries to explain classification and regression tree (CART) to Kowalski. CART is least likely to be applied to predict a:
   (A) discrete target variable, producing a cardinal tree.
   (B) categorical target variable, producing a classification tree.
   (C) continuous target variable, producing a regression tree.

5. Which of the following statements Nowak makes about hierarchical clustering is most accurate?
   (A) In divisive hierarchical clustering, the algorithm seeks out the two closest clusters.
   (B) Bottom-up hierarchical clustering begins with each observation being its own cluster.
   (C) Hierarchical clustering is a supervised iterative algorithm that is used to build a hierarchy of clusters.

6. Which of the following statements Nowak makes about neural networks is most accurate? Neural networks:
   (A) are effective in tasks with non-linearities and complex interactions among variables.
   (B) have four types of layers: an input layer, agglomerative layers, regularization layers, and an output layer.
   (C) have an input layer node that consists of a summation operator and an activation unction.

7. Nowak tries to explain the reinforcement learning (RL) algorithm to Kowalski and makes a number of statements about it. The reinforcement learning (RL) algorithm involves an agent that is most likely to:
   (A) perform actions that will minimize costs over time.
   (B) make use of direct labeled data and instantaneous feedback.
   (C) take into consideration the constraints of its environment.

8. which supervised learning model is most appropriate (1) when the Y-variable is continuous and (2) when Y-variable is categorical.

	Continuous Y – variable	Continuous Y – variable
(A)	Regression	Classification
(B)	Decision trees	Regression
(C)	Classification	Neural Networks

9. 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.

10. The unsupervised machine learning algorithm that reduces highly correlated features into fewer uncorrelated composite variables by transforming the feature covariance matrix best describes:
   (A) k-means clustering
   (B) principal components analysis
   (C) hierarchical clustering

11. Which of the following about unsupervised learning is most accurate?
   (A) Unsupervised learning has lower forecasting accuracy as compared to supervised learning.
   (B) Classification is an example of unsupervised learning algorithm.
   (C) There is no labeled data.

12. A random forest is least likely to:
    (A) be a classification tree.
    (B) reduce signal-to-noise ratio.
    (C) provide a solution to overfitting problem.

13. In machine learning, out-of-sample error equals:
    (A) Standard error plus data error plus prediction error.
    (B) forecast error plus expected error plus regression error.
    (C) bias error plus variance error plus base error.

14. Tan is interested in using a supervised learning algorithm to analyze stocks. This task is least likely to be a classification problem if the target variable is:
    (A) continuous.
    (B) ordinal.
    (C) categorical.

15. After Tan implements a particular new supervised machine learning algorithm, she begins to suspect that the holdout samples she is using are reducing the training set size too much. As a result begins to make use of K-fold cross-validation. In the K-fold cross-validation technique, after Tan shuffles the data randomly it is most likely that:
    (A) k – 1 samples will be used as validation samples.
    (B) the data will be divided into k – 1 equal sub-samples.
    (C) k – 1 samples will be used as training samples.

16. At first Tan bases her stock picks on the results of a single machine-learning model, but then begins to wonder if she should instead be using the predictions of a group of models. Compared to a single machine-learning model, an ensemble machine learning algorithm is most likely to produce predictions that are:
    (A) less reliable but more steady.
    (B) more accurate and more stable.
    (C) more precise but less dependable.

17. Tan is interested in applying neural networks, deep learning nets, and reinforcement learning to her investment process. Regarding these techniques, which of the following statements is most accurate?
    (A) Neural networks with one or more hidden layers would be considered deep learning nets (DLNs).
    (B) Reinforcement learning algorithms achieve maximum performance when they stay as far away from their constraints as possible.
    (C) Neural networks work well in the presence of non-linearities and complex interactions among variables.

18. A rudimentary way to think of machine learning algorithms is that they:
    (A) "find the pattern, apply the pattern."
    (B) "develop the pattern, interpret the pattern."
    (C) "synthesize the pattern, review the pattern."

19. Overfitting is least likely to result in:
    (A) higher number of features included in the data set.
    (B) higher forecasting accuracy in out-of-sample data.
    (C) inclusion of noise in the model.

20. 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.

21. Considering the various supervised machine learning algorithms, a linear classifier that seeks the optimal hyperplane and is typically used for classification, best describes:
    (A) k-nearest neighbor (KNN).
    (B) support vector machine (SVM).
    (C) classification and regression tree (CART).

22. The degree to which a machine learning model retains its explanatory power when predicting out-of sample is most commonly described as:
    (A) generalization.
    (B) hegemony.
    (C) predominance.

23. 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.