Module 14 Quiz:
Machine Learning: Classification

1. Information gain is often used to

A. Measure the quality of the split in a decision tree
B. Decide the number of decision trees to be used in a random forest
C. Find the optimal height of a decision tree
D. Find the optimal branching factor (number of children at each tree node) for the decision tree

Answer: A

2. When building a tree in a random forest, using out-of-bag (OOB) error estimates help reduce the need to perform explicit cross validation.

A. True
B. False

Answer: A

3. In k-fold cross-validation, each data point is used for validation exactly once.

A. True
B. False

Answer: A

4. Reducing a model's ROC curve to a single AUC number (area under the curve) can succinctly summarize the tradeoff between the model's true positive rate and false positive rate (as the model's discrimination threshold varies)

A. True
B. False

Answer: B

5. Supervised learning needs labeled data for building models for data analysis.

A. True
B. False

Answer: A

6. Data has to been separated into training and testing data in the procedure of building machine learning models.

A. True
B. False

Answer: A

7. Logistic regression is a non-linear classification model.

A. True
B. False

Answer: B

8. In the process of building a decision tree, it will implement feature selection.

A. True
B. False

Answer: A

9. Ensemble learning is to build a set of weak learners and combine them to build a strong learner.

A. True
B. False

Answer: A