

Large Scale Machine Learning and Deep Learning

Review Questions 3

1. Which of the following is/are true about Bagging Trees and Boosting Trees?
 - (a) In Bagging Trees, individual trees (weak learners) are independent of each other.
 - (b) Bagging is a method for improving the performance by aggregating the results of weak learners.
 - (c) In Boosting Trees, individual trees (weak learners) are independent of each other.
 - (d) Boosting is a method for improving the performance by aggregating the results of weak learners.

2. Which of the following is/are true about individual tree in Random Forest?
 - (a) Individual tree is built on a subset of the features.
 - (b) Individual tree is built on all the features.
 - (c) Individual tree is built on a subset of instances.
 - (d) Individual tree is built on full set of instances.

3. Ensemble model estimators (such as Random Forest) in Spark have a parameter called `featureSubsetStrategy`. What does it do?

4. Explain why the entropy becomes zero when all class partitions are pure?

5. Explain why the Gini impurity becomes zero when all class partitions are pure?