

DM-Quiz-2020-Q7

18 Questions

1. Decision Tree Decision Boundaries

- 3/12 ☒ A are a step-wise constant function
- 0/12 ☐ B I do not know
- 1/12 ☐ C continuous function
- 8/12 ☒ D are axis-parallel rectangles

2. Root Node has

- 3/13 ☒ A no incoming edges and zero or more outgoing edges
- 8/13 ☐ B one incoming edge and two or more outgoing edges
- 1/13 ☐ C one incoming edge and no outgoing edges
- 1/13 ☐ D I do not know

3. Pruning the tree means

- 5/11 ☒ A Simplify the tree
- 5/11 ☐ B Split the tree's nodes
- 1/11 ☐ C Merge the tree's nodes
- 0/11 ☐ D I do not know

4. Gini index equals to

- 5/12 ☒ A $1 - \sum (p_i^2)$
- 3/12 ☐ B $1 + \sum (p_i^2)$
- 3/12 ☐ C $\sum (p_i * \log(p_i))$
- 1/12 ☐ D $-\sum (p_i * \log(p_i))$
- 0/12 ☐ E I do not know

5. Entropy starts with 0

- 8/12 ☐ A True
- 4/12 ☒ B False
- 0/12 ☐ C I do not know

6. Overall impurity measure can be obtained by

- 9/11 ☒ A a weighted average of individual rectangles
1/11 ☐ B majority voting
1/11 ☐ C I do not know

7. At each stage, we choose the split with

- 8/12 ☒ A the lowest Gini index
1/12 ☐ B the lowest Chi-square value
3/12 ☐ C the highest entropy
0/12 ☐ D I do not know

8. We can perform the Decision Trees in r using

- 3/12 ☒ A rpart()
3/12 ☐ B decisiontree()
1/12 ☐ C destree()
4/12 ☐ D reg.tree()
1/12 ☐ E I do not know

9. minsplit in R means

- 10/11 ☒ A the minimum number of observations that must exist in a node in order for a split to be attempted
1/11 ☐ B the minimum number of observations in any terminal node
0/11 ☐ C the minimum number of splits
0/11 ☐ D I do not know

10. Bagging is a technique used to reduce

- 3/11 ☒ A the variance of our predictions
2/11 ☐ B the bias of our predictions
5/11 ☐ C both
1/11 ☐ D I do not know

11. Bootstrap aggregation allows sampling

- 6/12 ☒ A with replacement
- 3/12 ☐ B without replacement
- 1/12 ☐ C I do not know
- 2/12 ☐ D both

12. How can Ensemble methods be constructed?

- 2/11 ☐ A By manipulating the training set
- 1/11 ☐ B By manipulating the input features
- 1/11 ☐ C By manipulating the class labels
- 0/11 ☐ D By manipulating the learning algorithm
- 6/11 ☒ E All of them
- 0/11 ☐ F None
- 1/11 ☐ G I do not know

13. Repeatedly sampling observations are taken

- 4/11 ☐ A from general population
- 6/11 ☒ B original sample data set
- 0/11 ☐ C I do not know
- 1/11 ☐ D None

14. Random Forest differs from bagging

- 2/12 ☒ A by a random sample of m predictors
- 5/12 ☐ B by bootstrapped training samples
- 4/12 ☐ C by adaptive sampling
- 1/12 ☐ D I do not know

15. Boosting differs from bagging

- 6/12 ☐ A by a random sample of m predictors
- 5/12 ☐ B by bootstrapped training samples
- 1/12 ☒ C by adaptive sampling
- 0/12 ☐ D I do not know

16. Averaging many highly correlated quantities

- 5/12 ☐ A lead to as large of a reduction in variance
- 3/12 ☒ B does not lead to as large of a reduction in variance
- 2/12 ☐ C lead to as large of a reduction in bias
- 2/12 ☐ D I do not know

17. We can perform a Random forest in R using the function

- 4/11 ☒ A randomForest()
- 2/11 ☐ B rf()
- 2/11 ☐ C randomF()
- 2/11 ☐ D boot()
- 1/11 ☐ E I do not know

18. Random Forest works

- 4/11 ☐ A for classification
- 1/11 ☐ B for regression
- 6/11 ☒ C both
- 0/11 ☐ D I do not know