

# DM-Spring-2020-Q7-Grade

100% (18/18)

- ✓ 1. (Classification) Decision Tree Decision Boundaries
- ☐ A are a step-wise constant function
  - ☐ B I do not know
  - ☐ C continuous function
  - ☒ D are axis-parallel rectangles
- ✓ 2. Root Node has
- ☒ A no incoming edges and zero or more outgoing edges
  - ☐ B one incoming edge and two or more outgoing edges
  - ☐ C one incoming edge and no outgoing edges
  - ☐ D I do not know
- ✓ 3. Pruning the tree means
- ☒ A Simplify the tree
  - ☐ B Split the tree's nodes
  - ☐ C Merge the tree's nodes
  - ☐ D I do not know
- ✓ 4. Gini index equals to
- ☒ A  $1 - \sum (p_i^2)$
  - ☐ B  $1 + \sum (p_i^2)$
  - ☐ C  $\sum (p_i * \log(p_i))$
  - ☐ D  $-\sum (p_i * \log(p_i))$
  - ☐ E I do not know

- ✓ 5. Entropy starts with 0 (rough mathematically)
- ☐ A True
  - ☒ B False
  - ☐ C I do not know
- ✓ 6. Overall impurity measure can be obtained by
- ☒ A a weighted average of individual rectangles
  - ☐ B majority voting
  - ☐ C I do not know
- ✓ 7. At each stage, we choose the split with
- ☒ A the lowest Gini index
  - ☐ B the lowest Chi-square value
  - ☐ C the highest entropy
  - ☐ D I do not know
- ✓ 8. We can perform the Decision Trees in R using
- ☒ A `rpart()`
  - ☐ B `decisiontree()`
  - ☐ C `destree()`
  - ☐ D `reg.tree()`
  - ☐ E I do not know
- ✓ 9. `minsplit` in R means
- ☒ A the minimum number of observations that must exist in a node in order for a split to be attempted
  - ☐ B the minimum number of observations in any terminal node
  - ☐ C the minimum number of splits
  - ☐ D I do not know

- ✓ 10. Bagging is a technique used to reduce
- ☒ A the variance of our predictions
  - ☐ B the bias of our predictions
  - ☐ C both
  - ☐ D I do not know
- ✓ 11. Bootstrap aggregation allows sampling
- ☒ A with replacement
  - ☐ B without replacement
  - ☐ C I do not know
  - ☐ D both
- ✓ 12. How can Ensemble methods be constructed?
- ☐ A By manipulating the training set
  - ☐ B By manipulating the input features
  - ☐ C By manipulating the class labels
  - ☐ D By manipulating the learning algorithm
  - ☒ E All of them
  - ☐ F None
  - ☐ G I do not know
- ✓ 13. Repeatedly sampling observations are taken
- ☐ A from general population
  - ☒ B original sample data set
  - ☐ C I do not know
  - ☐ D None
- ✓ 14. Random Forest differs from bagging
- ☒ A by a random sample of  $m$  predictors
  - ☐ B by bootstrapped training samples
  - ☐ C by adaptive sampling
  - ☐ D I do not know

- ✓ 15. Boosting differs from bagging
- ☐ A by a random sample of  $m$  predictors
  - ☐ B by bootstrapped training samples
  - ☒ C by adaptive sampling
  - ☐ D I do not know
- ✓ 16. Averaging many highly correlated quantities
- ☐ A lead to as large of a reduction in variance
  - ☒ B does not lead to as large of a reduction in variance
  - ☐ C lead to as large of a reduction in bias
  - ☐ D I do not know
- ✓ 17. We can perform a Random forest in R using the function
- ☒ A randomForest()
  - ☐ B rf()
  - ☐ C randomF()
  - ☐ D boot()
  - ☐ E I do not know
- ✓ 18. Random Forest works
- ☐ A for classification
  - ☐ B for regression
  - ☒ C both
  - ☐ D I do not know