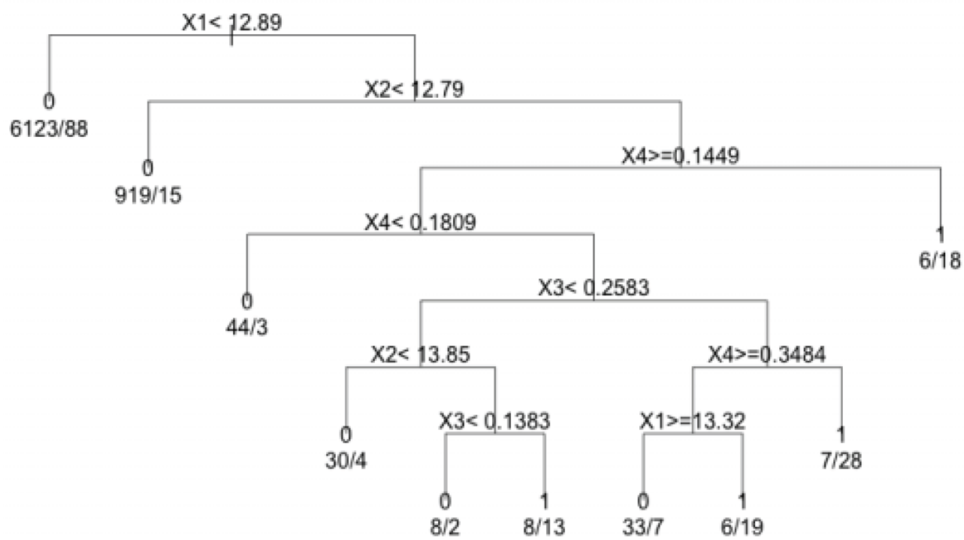
Quiz #6 – Solutions

1. How does a classification tree for a two-class response variable decide which variable (and value) to use for a particular decision (split)? Select the variable (and value) that most:

a) Increases the difference in class quantities in the parent node relative to each of the child nodes.
b) Decreases the difference in class proportions in each of the child nodes relative to the parent node.
c) Increases the difference in class quantities in each of the child nodes relative to the parent node.
d) Decreases the difference in class quantities in each of the child nodes relative to the parent node.
e) Increases the difference in class proportions in each of the child nodes relative to the parent node.

Solution: E

2. Suppose we fit a classification tree to a given data set and construct the following tree:



The question/decision rule corresponding to the first split is, "If X1<12.89 is true then go left, otherwise go right."

- If we go left, we end at a leaf node with 6123 training observations in class 0 and 88 training observations in class 1. Hence new observations that end at this leaf node would be predicted to be in class 0 with posterior probability 6123/(6123+88) = 0.98583159.
- If we go right we have another question/decision rule, "If X2<12.79 is true then go left, otherwise go right."
- And so on.

Match the following data observations with their predicted class and posterior probabilities based on this decision tree:

a) Predicted class = 1, posterior probability = 0.75
b) Predicted class = 0, posterior probability = 0.8
c) Predicted class = 1, posterior probability = 0.76

- (X1, X2, X3, X4, X5) = (13, 14, -0.3, 0.9, 0.2) → B
- (X1, X2, X3, X4, X5) = (14, 14, 0.1, 0.1, 0.1) → A
- (X1, X2, X3, X4, X5) = (13, 14, 0.3, 1.1, 0.1) → C

3. For the scenario in Question 2, which one of the following describes the characteristics of an observation that has predicted class 0 and posterior probability 44/47 = 0.936?

a) $X1 < 12.89$, $X2 < 12.79$, and $X4 < 0.1449$
b) $X1 < 12.89$, $X2 \geq 12.79$, and $0.1449 \leq X4 < 0.1809$
c) $X1 \geq 12.89$, $X2 < 12.79$, and $0.1449 \leq X4 < 0.1809$
d) $X1 \geq 12.89$, $X2 \geq 12.79$, and $0.1449 \leq X4 < 0.1809$
e) $X1 \geq 12.89$, $X2 \geq 12.79$, and $X4 \geq 0.1809$

Solution: D

4. For the scenario in Question 2, consider the bottom right question/decision rule, "$X1 \geq 13.32$," which, if true, leads to a leaf node with 33 training observations in class 0 and 7 training observations in class 1 and, if false, leads to a leaf node with 6 training observations in class 0 and 19 training observations in class 1. Suppose an alternative question/decision rule existed at this node, which, if true, leads to a leaf node with 34 training observations in class 0 and 11 training observations in class 1 and, if false, leads to a leaf node with 5 training observations in class 0 and 15 training observations in class 1. True or false: this alternative question/decision rule is preferable to the original question/decision rule? (Remember that the idea behind classification trees is to create "purer" child nodes.)

Solution: False

5. For the scenario in Question 2, X5 is not used at any of the splits. What single feature of classification trees does this best illustrate?

a) Robustness to outliers and misclassified observations in the training set.
b) Invariance under monotone transformations of quantitative predictors.
c) Automatic dimension reduction (variable selection).
d) Ease of interpretation via simple decision rules.
e) Provision of estimates of misclassification rates for test observations.

Solution: C

6. Consider bootstrap aggregation, or bagging, of classification trees. Which of the following statements about bagging in this context is the only one that is not true?

a) Bagging fits a classification tree to each of a collection of bootstrap samples of the training data and finds the proportion of trees predicting each class at each data point.
b) The structures of the trees for each bootstrap sample tend to be very similar, although predicted classifications from each tree can vary widely.
c) The predicted class for a particular data point based on the bagged trees is the class with the highest proportion of trees predicting that class.
d) It is possible to gauge the importance of each predictor when bagging classification trees.

Solution: B

7. Consider bootstrap aggregation, or bagging, of a regression tree. Which of the following statements about bagging in this context is true?

a) Bagging tends to increase mean squared prediction error by reducing bias while keeping variance unchanged.
b) Bagging tends to increase mean squared prediction error by reducing variance while keeping bias unchanged.
c) Bagging tends to decrease mean squared prediction error by reducing bias while keeping variance unchanged.
d) Bagging tends to decrease mean squared prediction error by reducing variance while keeping bias unchanged.

Solution: D

8. Consider a dataset in which there are a few input variables that dominate the major patterns in the data, but also many other input variables that would otherwise provide good splits in local regions of the input space. True or false: Random forests will most likely outperform bagging in this context.

Solution: True

9. Which of the following statements about boosting decision trees is the only one that is not true?

a) In contrast to bagged trees, which combines trees fit to bootstrap samples, boosting grows trees sequentially to residuals.
b) The number of splits in each tree should be as large as possible to allow all the variables to play a role.
c) Boosting can overfit if the total number of trees is too large.

d) Adjusting the boosting shrinkage parameter that controls the learning rate can lead to improved predictive performance.

Solution: B