Q1

Given a test data, we could use the constructed decision tree from top down, going through nodes, to decide which regions (R1,...,RJ) the test observation belongs to. Then the predicted value for a given test observation is the mean of the training observations in that region.

Where Rj is the region the test observation belongs to.

Q2

Given a test data, we could use the constructed decision tree from top down, going through nodes, to decide which regions (R1,...,RJ) the test observation belongs to. Then the predicted label for a given test observation is the most commonly occurring class of the training observations in that region.

Q3

The pure grow strategy may produce good predictions on the training set, but is likely to overfit the data, leading to poor test set performance.

Grow-and-prune is a strategy that grow a very large tree T0, and then prune it back in order to obtain a subtree using cost complexity pruning. A tuning parameter alpha, being selected with cross-validation, is used for controlling the trade-off between the subtree's complexity and its fit to the training data. Once the optimal value of alpha is selected, we could obtain the corresponding subtree from T0.

Q4

The idea of bagging is sampling n different datasets with bootstrap from original training set. Then we construct n regression trees for n datasets. Finally, we average the results of those trees to get final output.

Merits:

reduces variance; improves accuracy

Drawbacks:

not helpful in case of high bias or underfitting; loss of interpretability; computationally expensive

Q5

Sampling n different datasets, each with m predictors, using bootstrap from original training set with x features (m = sqrt(x) if random forest). Then we construct n regression trees for n datasets. Finally, we take a majority vote of the results of those trees (most commonly occurring class among the n predictions) to get final predicted label.

Q6

Variable importance summaries are used for interpretability.

For a bag of regression trees, we record the total amount that the RSS is decreased due to splits over a given predictor, averaged over all B trees. A large value indicates an important predictor.

For a bag of classification trees, we add up the total amount that the Gini index is decreased by splits over a given predictor, averaged over all B trees.

Q7

There is around one-third of the observations (out-of-bag observations) which are not used to fit a given bagged tree. For each observation, we could predict the response using each of the trees in which that observation was OOB. The predictions are averaged or major voted so that each observation have one OOB prediction. Besides, corresponding OOB error could be computed in the same way, which serves as a valid estimate of the test error for the bagged model since the response for each observation is predicted using only the trees that were not fit using that observation.