Assignment #5: Decision Tree

Attention: Please prepare two files for each homework assignment: the .docx or .pdf file for your answers including figures to each question; the other .R file for your R script. File names should be "LastName_FirstName_number.docx" and "LastName_FirstName_number.R" for assignment 5. All assignments should submitted via Course Website.

1. This problem involves the `OJ` data set which is part of the `ISLR` package and also available at Blakboard.

   (a) Create a training set containing a random sample of 800 observations, and a test set containing the remaining observations.

   (b) Fit a tree to the training data, with Purchase as the response and the other variables as predictors. Use the summary() function to produce summary statistics about the tree, and describe the results obtained. What is the training error rate? How many terminal nodes does the tree have?

   (c) Create a plot of the tree, and interpret the results.

   (d) Predict the response on the test data, and produce a confusion matrix comparing the test labels to the predicted test labels. What is the test error rate?

   (e) Create a pruned tree with four terminal nodes.

   (f) Compare the test error rates between the pruned and unpruned trees. Which is higher?