Assignment #6: Tree-Based Methods

Problem 1

In the lab, a classification tree was applied to the Carseats data set after converting Sales into a binary response variable. This question will seek to predict Sales using regression trees and related approaches, treating the response as a quantitative variable (that is, without the conversion).

(a) Split the data set into a training set and a test set.

(b) Fit a regression tree to the training set. Plot the tree, and interpret the results. Then compute the test MSE.

(c) Prune the tree obtained in (b). Use cross validation to determine the optimal level of tree complexity. Plot the pruned tree and interpret the results. Compute the test MSE of the pruned tree. Does pruning improve the test error?

(d) Use the bagging approach to analyze the data. What test MSE do you obtain? Determine which variables are most important.

(e) Use random forests to analyze the data. What test MSE do you obtain? Determine which variables are most important.

Problem 2

In the lab, we applied random forests to the Boston data using mtry=6 and ntree=100.

(a) Consider a more comprehensive range of values for mtry: 1, 2,…,13. Given each value of mtry, find the test error resulting from random forests on the Boston data (using ntree=100). Create a plot displaying the test error rate vs. the value of mtry. Comment on the results in the plot.

(b) Similarly, consider a range of values for ntree (between 5 to 200). Given each value of ntree, find the test error resulting from random forests (using mtry=6). Create a plot displaying the test error vs. the value of ntree. Comment on the results in the plot.

Submit through link: eCampus -> Assignments

Deadline: Nov 7, Tuesday @11:59pm