Homework 4

Due on 04/19/2024

1. In this exercise, we will build tree-based models using the `College` data (see "College.csv" in Homework 2). The response variable is the out-of-state tuition (`Outstate`). Partition the dataset into two parts: training data (80%) and test data (20%).

  (a) Build a regression tree on the training data to predict the response. Create a plot of the tree.

  (b) Perform random forest on the training data. Report the variable importance and the test error.

  (c) Perform boosting on the training data. Report the variable importance and the test error.

2. This problem is based on the data "auto.csv" in Homework 3. Split the dataset into two parts: training data (70%) and test data (30%).

  (a) Build a classification tree using the training data, with `mpg_cat` as the response. Which tree size corresponds to the lowest cross-validation error? Is this the same as the tree size obtained using the 1 SE rule?

  (b) Perform boosting on the training data and report the variable importance. Report the test data performance.