STAT 760 Homework 7

Exercise 1. (Programming)– **30 pts**

Predict `Sales` using regression trees and related approaches, treating the response as a quantitative variable.

(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. What test MSE do you obtain?

(c) Use cross-validation in order to determine the optimal level of tree complexity. Does pruning the tree improve the test MSE?

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

(e) Use random forests to analyze this data. What test MSE do you obtain? Determine which variables are most important. Describe the effect of m, the number of variables considered at each split, on the error rate obtained.

The data `Carseats` can be found in Package 'ISLR'. See the details at

$$\text{https://cran.r-project.org/web/packages/ISLR/ISLR.pdf}$$

1. Please don't use tree packages.

2. Graph packages are allowed.

3. Hint: it may consider the tree data structure.