Assignment2\_Ans

STAT380

2023-10-29

###  
library(tree)  
###  
library(ISLR)  
data(Carseats)  
attach(Carseats)

The data set `Carseats’ is a simulated data set containing sales of child car seats at 400 different stores of a specific departmental store over a the period of a few months. For the different activities in this assignment, we consider a categorical binary variable, that we call ‘High\_Sells’. We consider the sales amount is high, i.e., High\_sell=1, if the number of car sets that is sold is greater then 8 in the particular store.

# Problem A

### A1. Create a new binary categorical variable called High\_Sells' which is defined as follows: $ \text{High\_Sales}=1 \text{ if }Sales’>8 $ and $ =0 `Sales’ $

### A2. Add the new variable to the dataset `Carseats’.

### A1.  
  
  
### A2.

### A3. Build a classification tree where the response is High and the predictors all other variables.

### A4. Plot the tree and describe

#

### A5. Now, for the cross-validation, use a training (200 observations) and validating sets. Comment on the validation part.

#

### A6. Use the cv.tree function for the cross-validation. Compare with question 5.

#

### A7. Plot the error rate from the cross-validation as a function of both size and k.

#

### A8. Prune the tree the best size (number of nodes) found in question 7.

### A9. Predict with the pruned tree. Conclude.

#

### A10. Compare the classification performance of the tree and the pruned tree.

#

# Problem B (Fitting Regression Trees)

### We will use the regression trees for the Boston Housing data

## B1. Train a regression tree on the half of the observations of the Boston housing data. Comment.

##

## B2. Plot the tree.

##

## B3. Run the cross-validation for the tree and plot the errors.

##

## B4. Prune the regression tree to the optimal number of nodes.

##

## B5. Plot the predicted values vs the test values and calculate the MSE. comment.

##