Assignment2 Answerts

STAT380

2023-10-29

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

## Warning: package 'rattle' was built under R version 4.0.5

## Loading required package: tibble

## Loading required package: bitops

## Rattle: A free graphical interface for data science with R.  
## Version 5.5.1 Copyright (c) 2006-2021 Togaware Pty Ltd.  
## Type 'rattle()' to shake, rattle, and roll your data.

library(rpart.plot)

## Loading required package: rpart

library(RColorBrewer)  
library(partykit)

## Loading required package: grid

## Loading required package: libcoin

## Loading required package: mvtnorm

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 period of a few months. For the different activities in this assignment, we consider a categorical binary variable, that we call ‘High\_Sales’. We consider the sales amount is high, i.e., High\_Sales=1, if the number of car sets that are sold is greater than 8 in the particular store.

# Problem A

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

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

data(Carseats)  
### A1.  
High\_Sales<-ifelse(Carseats$Sales>8, 'High', 'Low')  
  
### A2.  
Carseats$High\_Sales=as.factor(High\_Sales)

### A3. Build a classification tree where the response is ‘High\_Sales’ and the predictors are all the other variables except ‘Sales’ and ‘High\_Sales’.

names(Carseats)

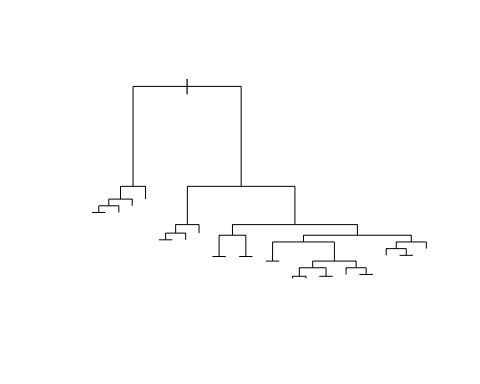
## [1] "Sales" "CompPrice" "Income" "Advertising" "Population"   
## [6] "Price" "ShelveLoc" "Age" "Education" "Urban"   
## [11] "US" "High\_Sales"

### A3.  
fit = rpart(High\_Sales~CompPrice+Income+Advertising+Population+Price+ShelveLoc+Age+Education+Urban+US,data=Carseats,method="class",minsplit=20,cp=-1)

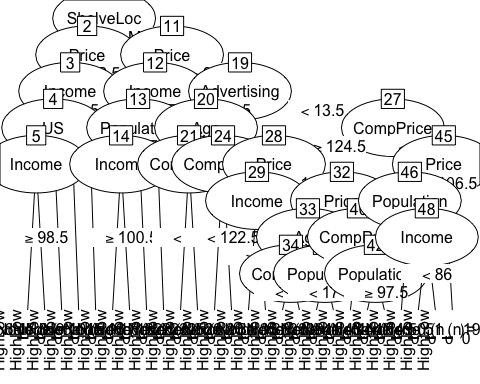
### A4.1 Plot the fitted tree and

#### A4.2 describe the different classification region.

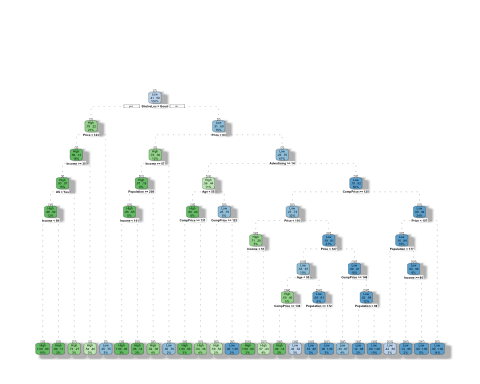
# A4.1 Plot the regression tree  
plot(fit)



plot(as.party(fit)) # Basic Plot



fancyRpartPlot(fit, caption = NULL) # More Nicer plot but need libraries `RColorBrewer', `rattle' and `rpart.plot'



#A4.2 describe the different classification region.

## Training and Testing Set

### A5.1. Now, Create a training set and a testing set , use a training set (70% in the training Set and 30% in test set).

#### A5.1:  
set.seed(234)  
 #inTrain = createDataPartition(Carseats$Sales, p = 0.75, list = FALSE)  
 Total\_data\_size=as.integer(nrow(Carseats))  
 inTrain = sample(1:Total\_data\_size, round(Total\_data\_size\*.70))  
 Training\_Set = Carseats[inTrain, ]  
 dim(Training\_Set)

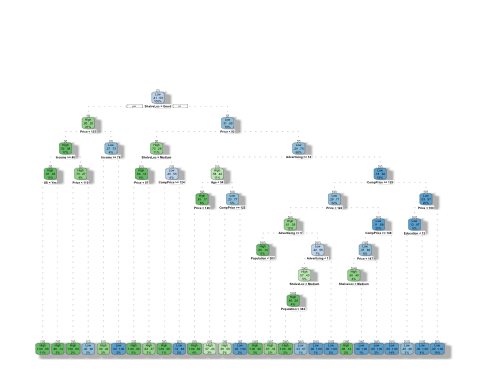
## [1] 280 12

Testing\_Set<-Carseats[-inTrain, ]  
 dim(Testing\_Set)

## [1] 120 12

### A5.2. Fit the Classification Tree in the Training set and Plot the tree

### A5.2.  
fit\_train = rpart(High\_Sales~CompPrice+Income+Advertising+Population+Price+ShelveLoc+Age+Education+Urban+US,data=Training\_Set,method="class",minsplit=10,cp=-1)  
  
  
  
# plot mytree  
fancyRpartPlot(fit\_train, caption = NULL)



### A6. Comment on the Error Rate in the validation part.

Identify an optimal value for complexiety parameter `cp’. Note: cp = “value” is an assigned numeric value that will determine how tall a tree is to be growen. The smaller value (closer to 0) leads to the larger the trees. The default value is 0.01.

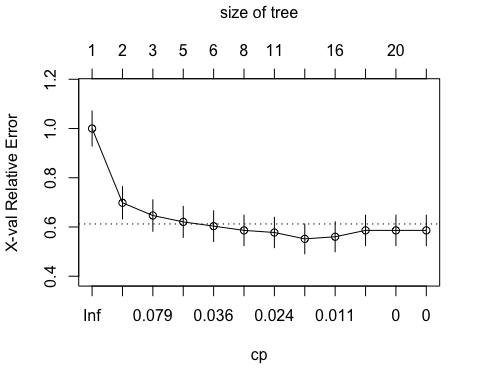
summary(fit\_train)

## Call:  
## rpart(formula = High\_Sales ~ CompPrice + Income + Advertising +   
## Population + Price + ShelveLoc + Age + Education + Urban +   
## US, data = Training\_Set, method = "class", minsplit = 10,   
## cp = -1)  
## n= 280   
##   
## CP nsplit rel error xerror xstd  
## 1 0.301724138 0 1.0000000 1.0000000 0.07105815  
## 2 0.120689655 1 0.6982759 0.6982759 0.06540818  
## 3 0.051724138 2 0.5775862 0.6465517 0.06388083  
## 4 0.043103448 4 0.4741379 0.6206897 0.06304649  
## 5 0.030172414 5 0.4310345 0.6034483 0.06246283  
## 6 0.025862069 7 0.3706897 0.5862069 0.06185651  
## 7 0.021551724 10 0.2931034 0.5775862 0.06154463  
## 8 0.014367816 12 0.2500000 0.5517241 0.06057311  
## 9 0.008620690 15 0.2068966 0.5603448 0.06090303  
## 10 0.004310345 17 0.1896552 0.5862069 0.06185651  
## 11 0.000000000 19 0.1810345 0.5862069 0.06185651  
## 12 -1.000000000 26 0.1810345 0.5862069 0.06185651  
##   
## Variable importance  
## Price ShelveLoc CompPrice Advertising Age Income   
## 27 18 15 12 8 7   
## Population Education US Urban   
## 7 4 1 1   
##   
## Node number 1: 280 observations, complexity param=0.3017241  
## predicted class=Low expected loss=0.4142857 P(node) =1  
## class counts: 116 164  
## probabilities: 0.414 0.586   
## left son=2 (59 obs) right son=3 (221 obs)  
## Primary splits:  
## ShelveLoc splits as RLR, improve=21.853040, (0 missing)  
## Price < 92.5 to the left, improve=15.953800, (0 missing)  
## Income < 60.5 to the right, improve=10.590280, (0 missing)  
## Advertising < 7.5 to the right, improve=10.046180, (0 missing)  
## Age < 61.5 to the left, improve= 4.865515, (0 missing)  
##   
## Node number 2: 59 observations, complexity param=0.04310345  
## predicted class=High expected loss=0.2033898 P(node) =0.2107143  
## class counts: 47 12  
## probabilities: 0.797 0.203   
## left son=4 (48 obs) right son=5 (11 obs)  
## Primary splits:  
## Price < 135 to the left, improve=7.421674, (0 missing)  
## Income < 42.5 to the right, improve=3.010243, (0 missing)  
## US splits as RL, improve=2.406435, (0 missing)  
## Advertising < 2.5 to the right, improve=2.055987, (0 missing)  
## Population < 342 to the right, improve=1.423772, (0 missing)  
## Surrogate splits:  
## CompPrice < 150 to the left, agree=0.847, adj=0.182, (0 split)  
## Age < 77.5 to the left, agree=0.847, adj=0.182, (0 split)  
##   
## Node number 3: 221 observations, complexity param=0.1206897  
## predicted class=Low expected loss=0.3122172 P(node) =0.7892857  
## class counts: 69 152  
## probabilities: 0.312 0.688   
## left son=6 (32 obs) right son=7 (189 obs)  
## Primary splits:  
## Price < 92.5 to the left, improve=12.368060, (0 missing)  
## Income < 60.5 to the right, improve= 8.396198, (0 missing)  
## Advertising < 7.5 to the right, improve= 7.592557, (0 missing)  
## ShelveLoc splits as R-L, improve= 5.560181, (0 missing)  
## Age < 50.5 to the left, improve= 3.968117, (0 missing)  
## Surrogate splits:  
## CompPrice < 103.5 to the left, agree=0.882, adj=0.188, (0 split)  
##   
## Node number 4: 48 observations, complexity param=0.004310345  
## predicted class=High expected loss=0.08333333 P(node) =0.1714286  
## class counts: 44 4  
## probabilities: 0.917 0.083   
## left son=8 (37 obs) right son=9 (11 obs)  
## Primary splits:  
## Income < 40 to the right, improve=1.0237510, (0 missing)  
## US splits as RL, improve=0.8888889, (0 missing)  
## Advertising < 0.5 to the right, improve=0.6778711, (0 missing)  
## Price < 109.5 to the left, improve=0.5641026, (0 missing)  
## Education < 14.5 to the right, improve=0.4367816, (0 missing)  
## Surrogate splits:  
## Age < 28.5 to the right, agree=0.812, adj=0.182, (0 split)  
## Population < 50 to the right, agree=0.792, adj=0.091, (0 split)  
##   
## Node number 5: 11 observations, complexity param=0.00862069  
## predicted class=Low expected loss=0.2727273 P(node) =0.03928571  
## class counts: 3 8  
## probabilities: 0.273 0.727   
## left son=10 (5 obs) right son=11 (6 obs)  
## Primary splits:  
## Income < 75.5 to the right, improve=1.9636360, (0 missing)  
## Advertising < 3 to the right, improve=1.3636360, (0 missing)  
## Population < 324.5 to the right, improve=1.2803030, (0 missing)  
## US splits as RL, improve=0.9350649, (0 missing)  
## Price < 142.5 to the left, improve=0.6493506, (0 missing)  
## Surrogate splits:  
## Price < 156.5 to the left, agree=0.727, adj=0.4, (0 split)  
## Age < 46.5 to the left, agree=0.727, adj=0.4, (0 split)  
## CompPrice < 139.5 to the left, agree=0.636, adj=0.2, (0 split)  
## Advertising < 1 to the right, agree=0.636, adj=0.2, (0 split)  
## Population < 292.5 to the right, agree=0.636, adj=0.2, (0 split)  
##   
## Node number 6: 32 observations, complexity param=0.02155172  
## predicted class=High expected loss=0.28125 P(node) =0.1142857  
## class counts: 23 9  
## probabilities: 0.719 0.281   
## left son=12 (21 obs) right son=13 (11 obs)  
## Primary splits:  
## ShelveLoc splits as R-L, improve=2.3400970, (0 missing)  
## Income < 57 to the right, improve=2.1939100, (0 missing)  
## Age < 66.5 to the left, improve=1.0208330, (0 missing)  
## Urban splits as LR, improve=0.9556818, (0 missing)  
## Advertising < 9.5 to the right, improve=0.9533730, (0 missing)  
## Surrogate splits:  
## CompPrice < 122 to the left, agree=0.781, adj=0.364, (0 split)  
## Age < 76.5 to the left, agree=0.750, adj=0.273, (0 split)  
## Income < 57 to the right, agree=0.719, adj=0.182, (0 split)  
##   
## Node number 7: 189 observations, complexity param=0.05172414  
## predicted class=Low expected loss=0.2433862 P(node) =0.675  
## class counts: 46 143  
## probabilities: 0.243 0.757   
## left son=14 (31 obs) right son=15 (158 obs)  
## Primary splits:  
## Advertising < 13.5 to the right, improve=8.435742, (0 missing)  
## CompPrice < 124.5 to the right, improve=6.074711, (0 missing)  
## Age < 49.5 to the left, improve=5.253097, (0 missing)  
## Price < 124.5 to the left, improve=4.960317, (0 missing)  
## Income < 60.5 to the right, improve=4.550008, (0 missing)  
## Surrogate splits:  
## Population < 500 to the right, agree=0.841, adj=0.032, (0 split)  
##   
## Node number 8: 37 observations, complexity param=0  
## predicted class=High expected loss=0.02702703 P(node) =0.1321429  
## class counts: 36 1  
## probabilities: 0.973 0.027   
##   
## Node number 9: 11 observations, complexity param=0.004310345  
## predicted class=High expected loss=0.2727273 P(node) =0.03928571  
## class counts: 8 3  
## probabilities: 0.727 0.273   
## left son=18 (6 obs) right son=19 (5 obs)  
## Primary splits:  
## Price < 109.5 to the left, improve=1.9636360, (0 missing)  
## Education < 12.5 to the left, improve=1.3636360, (0 missing)  
## Age < 58 to the left, improve=1.2803030, (0 missing)  
## Income < 30.5 to the left, improve=0.9350649, (0 missing)  
## Advertising < 10.5 to the right, improve=0.9350649, (0 missing)  
## Surrogate splits:  
## Income < 36.5 to the left, agree=0.727, adj=0.4, (0 split)  
## Advertising < 13.5 to the left, agree=0.727, adj=0.4, (0 split)  
## Population < 261.5 to the left, agree=0.727, adj=0.4, (0 split)  
## Age < 32.5 to the right, agree=0.727, adj=0.4, (0 split)  
## Education < 11.5 to the left, agree=0.727, adj=0.4, (0 split)  
##   
## Node number 10: 5 observations  
## predicted class=High expected loss=0.4 P(node) =0.01785714  
## class counts: 3 2  
## probabilities: 0.600 0.400   
##   
## Node number 11: 6 observations  
## predicted class=Low expected loss=0 P(node) =0.02142857  
## class counts: 0 6  
## probabilities: 0.000 1.000   
##   
## Node number 12: 21 observations, complexity param=0  
## predicted class=High expected loss=0.1428571 P(node) =0.075  
## class counts: 18 3  
## probabilities: 0.857 0.143   
##   
## Node number 13: 11 observations, complexity param=0.02155172  
## predicted class=Low expected loss=0.4545455 P(node) =0.03928571  
## class counts: 5 6  
## probabilities: 0.455 0.545   
## left son=26 (4 obs) right son=27 (7 obs)  
## Primary splits:  
## CompPrice < 124 to the right, improve=3.740260, (0 missing)  
## Income < 61.5 to the right, improve=2.597403, (0 missing)  
## Population < 396 to the left, improve=1.704545, (0 missing)  
## Advertising < 9.5 to the right, improve=1.097403, (0 missing)  
## Price < 83.5 to the right, improve=1.097403, (0 missing)  
## Surrogate splits:  
## Advertising < 9.5 to the right, agree=0.818, adj=0.50, (0 split)  
## Income < 61.5 to the right, agree=0.727, adj=0.25, (0 split)  
## Price < 85.5 to the right, agree=0.727, adj=0.25, (0 split)  
##   
## Node number 14: 31 observations, complexity param=0.05172414  
## predicted class=High expected loss=0.4193548 P(node) =0.1107143  
## class counts: 18 13  
## probabilities: 0.581 0.419   
## left son=28 (18 obs) right son=29 (13 obs)  
## Primary splits:  
## Age < 54 to the left, improve=5.481390, (0 missing)  
## Income < 99.5 to the right, improve=3.792426, (0 missing)  
## CompPrice < 121 to the right, improve=3.336090, (0 missing)  
## Price < 126 to the left, improve=3.336090, (0 missing)  
## Advertising < 21.5 to the right, improve=1.615293, (0 missing)  
## Surrogate splits:  
## CompPrice < 129.5 to the right, agree=0.677, adj=0.231, (0 split)  
## Advertising < 17.5 to the right, agree=0.677, adj=0.231, (0 split)  
## Income < 51.5 to the right, agree=0.645, adj=0.154, (0 split)  
## Price < 112.5 to the right, agree=0.645, adj=0.154, (0 split)  
## Population < 151.5 to the right, agree=0.613, adj=0.077, (0 split)  
##   
## Node number 15: 158 observations, complexity param=0.03017241  
## predicted class=Low expected loss=0.1772152 P(node) =0.5642857  
## class counts: 28 130  
## probabilities: 0.177 0.823   
## left son=30 (89 obs) right son=31 (69 obs)  
## Primary splits:  
## CompPrice < 124.5 to the right, improve=5.382903, (0 missing)  
## Price < 126.5 to the left, improve=2.710202, (0 missing)  
## ShelveLoc splits as R-L, improve=2.533525, (0 missing)  
## Income < 60.5 to the right, improve=1.692874, (0 missing)  
## Age < 49.5 to the left, improve=1.570408, (0 missing)  
## Surrogate splits:  
## Price < 115.5 to the right, agree=0.722, adj=0.362, (0 split)  
## Age < 50.5 to the left, agree=0.658, adj=0.217, (0 split)  
## Population < 405.5 to the left, agree=0.614, adj=0.116, (0 split)  
## Income < 22.5 to the right, agree=0.582, adj=0.043, (0 split)  
## Education < 11.5 to the right, agree=0.582, adj=0.043, (0 split)  
##   
## Node number 18: 6 observations  
## predicted class=High expected loss=0 P(node) =0.02142857  
## class counts: 6 0  
## probabilities: 1.000 0.000   
##   
## Node number 19: 5 observations  
## predicted class=Low expected loss=0.4 P(node) =0.01785714  
## class counts: 2 3  
## probabilities: 0.400 0.600   
##   
## Node number 26: 4 observations  
## predicted class=High expected loss=0 P(node) =0.01428571  
## class counts: 4 0  
## probabilities: 1.000 0.000   
##   
## Node number 27: 7 observations  
## predicted class=Low expected loss=0.1428571 P(node) =0.025  
## class counts: 1 6  
## probabilities: 0.143 0.857   
##   
## Node number 28: 18 observations, complexity param=0  
## predicted class=High expected loss=0.1666667 P(node) =0.06428571  
## class counts: 15 3  
## probabilities: 0.833 0.167   
##   
## Node number 29: 13 observations, complexity param=0  
## predicted class=Low expected loss=0.2307692 P(node) =0.04642857  
## class counts: 3 10  
## probabilities: 0.231 0.769   
##   
## Node number 30: 89 observations, complexity param=0.03017241  
## predicted class=Low expected loss=0.2921348 P(node) =0.3178571  
## class counts: 26 63  
## probabilities: 0.292 0.708   
## left son=60 (33 obs) right son=61 (56 obs)  
## Primary splits:  
## Price < 121.5 to the left, improve=10.337130, (0 missing)  
## ShelveLoc splits as R-L, improve= 4.420239, (0 missing)  
## CompPrice < 143.5 to the right, improve= 2.108134, (0 missing)  
## Income < 60.5 to the right, improve= 1.342253, (0 missing)  
## Advertising < 6.5 to the right, improve= 1.131840, (0 missing)  
## Surrogate splits:  
## Population < 75 to the left, agree=0.697, adj=0.182, (0 split)  
## CompPrice < 129.5 to the left, agree=0.685, adj=0.152, (0 split)  
## Income < 25.5 to the left, agree=0.652, adj=0.061, (0 split)  
## Age < 77.5 to the right, agree=0.652, adj=0.061, (0 split)  
##   
## Node number 31: 69 observations, complexity param=0  
## predicted class=Low expected loss=0.02898551 P(node) =0.2464286  
## class counts: 2 67  
## probabilities: 0.029 0.971   
##   
## Node number 60: 33 observations, complexity param=0.02586207  
## predicted class=High expected loss=0.3939394 P(node) =0.1178571  
## class counts: 20 13  
## probabilities: 0.606 0.394   
## left son=120 (14 obs) right son=121 (19 obs)  
## Primary splits:  
## Advertising < 5 to the right, improve=3.065846, (0 missing)  
## ShelveLoc splits as R-L, improve=2.831650, (0 missing)  
## Income < 63.5 to the left, improve=2.472960, (0 missing)  
## Price < 105 to the left, improve=1.979798, (0 missing)  
## Age < 44.5 to the left, improve=1.948052, (0 missing)  
## Surrogate splits:  
## US splits as RL, agree=0.879, adj=0.714, (0 split)  
## Education < 11.5 to the left, agree=0.697, adj=0.286, (0 split)  
## Income < 62.5 to the left, agree=0.606, adj=0.071, (0 split)  
## Population < 36 to the left, agree=0.606, adj=0.071, (0 split)  
## Price < 96.5 to the left, agree=0.606, adj=0.071, (0 split)  
##   
## Node number 61: 56 observations, complexity param=0.01436782  
## predicted class=Low expected loss=0.1071429 P(node) =0.2  
## class counts: 6 50  
## probabilities: 0.107 0.893   
## left son=122 (16 obs) right son=123 (40 obs)  
## Primary splits:  
## CompPrice < 147.5 to the right, improve=3.2142860, (0 missing)  
## Income < 77.5 to the right, improve=0.9142857, (0 missing)  
## Population < 398 to the right, improve=0.7506494, (0 missing)  
## ShelveLoc splits as R-L, improve=0.6602317, (0 missing)  
## Price < 144.5 to the left, improve=0.5604396, (0 missing)  
## Surrogate splits:  
## Price < 158 to the right, agree=0.786, adj=0.250, (0 split)  
## Population < 367.5 to the right, agree=0.750, adj=0.125, (0 split)  
##   
## Node number 120: 14 observations, complexity param=0  
## predicted class=High expected loss=0.1428571 P(node) =0.05  
## class counts: 12 2  
## probabilities: 0.857 0.143   
##   
## Node number 121: 19 observations, complexity param=0.02586207  
## predicted class=Low expected loss=0.4210526 P(node) =0.06785714  
## class counts: 8 11  
## probabilities: 0.421 0.579   
## left son=242 (14 obs) right son=243 (5 obs)  
## Primary splits:  
## Advertising < 1 to the left, improve=2.406015, (0 missing)  
## ShelveLoc splits as R-L, improve=2.406015, (0 missing)  
## CompPrice < 144.5 to the right, improve=2.388158, (0 missing)  
## Population < 53.5 to the right, improve=1.796491, (0 missing)  
## US splits as LR, improve=1.796491, (0 missing)  
## Surrogate splits:  
## Income < 27.5 to the right, agree=0.842, adj=0.4, (0 split)  
## Age < 61.5 to the left, agree=0.789, adj=0.2, (0 split)  
##   
## Node number 122: 16 observations, complexity param=0.01436782  
## predicted class=Low expected loss=0.375 P(node) =0.05714286  
## class counts: 6 10  
## probabilities: 0.375 0.625   
## left son=244 (10 obs) right son=245 (6 obs)  
## Primary splits:  
## Price < 147 to the left, improve=2.700000, (0 missing)  
## Income < 76.5 to the right, improve=2.627273, (0 missing)  
## CompPrice < 156 to the left, improve=2.045455, (0 missing)  
## Education < 15.5 to the left, improve=1.633333, (0 missing)  
## ShelveLoc splits as R-L, improve=1.500000, (0 missing)  
## Surrogate splits:  
## Education < 16.5 to the left, agree=0.812, adj=0.500, (0 split)  
## CompPrice < 158 to the left, agree=0.750, adj=0.333, (0 split)  
## Income < 38 to the right, agree=0.750, adj=0.333, (0 split)  
## Advertising < 10.5 to the left, agree=0.750, adj=0.333, (0 split)  
## Age < 26.5 to the right, agree=0.750, adj=0.333, (0 split)  
##   
## Node number 123: 40 observations  
## predicted class=Low expected loss=0 P(node) =0.1428571  
## class counts: 0 40  
## probabilities: 0.000 1.000   
##   
## Node number 242: 14 observations, complexity param=0.02586207  
## predicted class=High expected loss=0.4285714 P(node) =0.05  
## class counts: 8 6  
## probabilities: 0.571 0.429   
## left son=484 (10 obs) right son=485 (4 obs)  
## Primary splits:  
## ShelveLoc splits as R-L, improve=3.657143, (0 missing)  
## Population < 53.5 to the right, improve=2.493506, (0 missing)  
## Income < 63.5 to the left, improve=2.057143, (0 missing)  
## CompPrice < 141.5 to the right, improve=1.402597, (0 missing)  
## Price < 105 to the left, improve=1.402597, (0 missing)  
## Surrogate splits:  
## Population < 53.5 to the right, agree=0.929, adj=0.75, (0 split)  
## Price < 119.5 to the left, agree=0.786, adj=0.25, (0 split)  
## Urban splits as RL, agree=0.786, adj=0.25, (0 split)  
##   
## Node number 243: 5 observations  
## predicted class=Low expected loss=0 P(node) =0.01785714  
## class counts: 0 5  
## probabilities: 0.000 1.000   
##   
## Node number 244: 10 observations, complexity param=0.01436782  
## predicted class=High expected loss=0.4 P(node) =0.03571429  
## class counts: 6 4  
## probabilities: 0.600 0.400   
## left son=488 (7 obs) right son=489 (3 obs)  
## Primary splits:  
## ShelveLoc splits as R-L, improve=3.085714, (0 missing)  
## CompPrice < 152.5 to the left, improve=1.633333, (0 missing)  
## Income < 74 to the right, improve=0.800000, (0 missing)  
## Price < 131.5 to the right, improve=0.800000, (0 missing)  
## Education < 15.5 to the left, improve=0.800000, (0 missing)  
## Surrogate splits:  
## CompPrice < 152.5 to the left, agree=0.9, adj=0.667, (0 split)  
## Income < 58.5 to the right, agree=0.8, adj=0.333, (0 split)  
## Price < 131.5 to the right, agree=0.8, adj=0.333, (0 split)  
## Education < 15.5 to the left, agree=0.8, adj=0.333, (0 split)  
##   
## Node number 245: 6 observations  
## predicted class=Low expected loss=0 P(node) =0.02142857  
## class counts: 0 6  
## probabilities: 0.000 1.000   
##   
## Node number 484: 10 observations, complexity param=0.00862069  
## predicted class=High expected loss=0.2 P(node) =0.03571429  
## class counts: 8 2  
## probabilities: 0.800 0.200   
## left son=968 (7 obs) right son=969 (3 obs)  
## Primary splits:  
## Population < 382 to the left, improve=1.8666670, (0 missing)  
## Price < 114 to the left, improve=0.8000000, (0 missing)  
## Income < 67 to the left, improve=0.5333333, (0 missing)  
## Age < 43 to the left, improve=0.5333333, (0 missing)  
## CompPrice < 141 to the right, improve=0.3428571, (0 missing)  
## Surrogate splits:  
## Price < 114 to the left, agree=0.8, adj=0.333, (0 split)  
##   
## Node number 485: 4 observations  
## predicted class=Low expected loss=0 P(node) =0.01428571  
## class counts: 0 4  
## probabilities: 0.000 1.000   
##   
## Node number 488: 7 observations  
## predicted class=High expected loss=0.1428571 P(node) =0.025  
## class counts: 6 1  
## probabilities: 0.857 0.143   
##   
## Node number 489: 3 observations  
## predicted class=Low expected loss=0 P(node) =0.01071429  
## class counts: 0 3  
## probabilities: 0.000 1.000   
##   
## Node number 968: 7 observations  
## predicted class=High expected loss=0 P(node) =0.025  
## class counts: 7 0  
## probabilities: 1.000 0.000   
##   
## Node number 969: 3 observations  
## predicted class=Low expected loss=0.3333333 P(node) =0.01071429  
## class counts: 1 2  
## probabilities: 0.333 0.667

printcp(fit\_train)

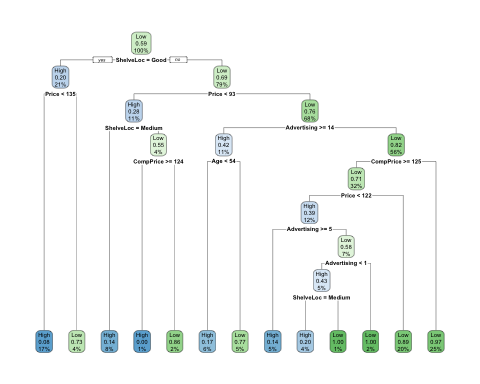
##   
## Classification tree:  
## rpart(formula = High\_Sales ~ CompPrice + Income + Advertising +   
## Population + Price + ShelveLoc + Age + Education + Urban +   
## US, data = Training\_Set, method = "class", minsplit = 10,   
## cp = -1)  
##   
## Variables actually used in tree construction:  
## [1] Advertising Age CompPrice Education Income Population   
## [7] Price ShelveLoc US   
##   
## Root node error: 116/280 = 0.41429  
##   
## n= 280   
##   
## CP nsplit rel error xerror xstd  
## 1 0.3017241 0 1.00000 1.00000 0.071058  
## 2 0.1206897 1 0.69828 0.69828 0.065408  
## 3 0.0517241 2 0.57759 0.64655 0.063881  
## 4 0.0431034 4 0.47414 0.62069 0.063046  
## 5 0.0301724 5 0.43103 0.60345 0.062463  
## 6 0.0258621 7 0.37069 0.58621 0.061857  
## 7 0.0215517 10 0.29310 0.57759 0.061545  
## 8 0.0143678 12 0.25000 0.55172 0.060573  
## 9 0.0086207 15 0.20690 0.56034 0.060903  
## 10 0.0043103 17 0.18966 0.58621 0.061857  
## 11 0.0000000 19 0.18103 0.58621 0.061857  
## 12 -1.0000000 26 0.18103 0.58621 0.061857

plotcp(fit\_train)



### A7. Prune the tree According to the optimal value of cp that you have obtainied in A6.

### A8.  
  
bestcp <-fit\_train$cptable[which.min(fit\_train$cptable[,"xerror"]),"CP"]  
pruned.tree <- prune(fit\_train, cp = bestcp)  
rpart.plot(pruned.tree)

 ###

### A8.1 Predict on the Testing set with the pruned tree and

### A8.2 Predict on the Testing set with the pruned tree andPredict on the Testing set with the entire tree fitted using the training set

### A8.1  
# Alternate specification   
pred\_test.prune = predict(pruned.tree, Testing\_Set, type="class")  
  
### A8.1  
pred\_test.full\_tree=predict(fit\_train, Testing\_Set, type="class")

### A9.1 Create A classification Tables of the errors using the Predicted values from the pruned tree

### A9.2 Create A classification Tables of the errors using the Predicted values from entire tree fitted using the training set

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

#A9.1  
table(pred\_test.prune,Testing\_Set$High\_Sales )

##   
## pred\_test.prune High Low  
## High 31 14  
## Low 17 58

#A9.2  
table(pred\_test.full\_tree,Testing\_Set$High\_Sales )

##   
## pred\_test.full\_tree High Low  
## High 34 15  
## Low 14 57

## Write a Conslusion of your finding

# Problem B (Fitting Regression Trees)

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

Load the data from the github course page using: BostonH<-read.csv(url(“<https://raw.githubusercontent.com/subhadippal2019/STAT380UAEU/main/BostonHousing.csv>”))

BostonH<-read.csv(url("https://raw.githubusercontent.com/subhadippal2019/STAT380UAEU/main/BostonHousing.csv"))

### B1 Split the data in Training and Testing Set. use a 60%/40% split fopr the trainig and Testing

#### A5.1:  
set.seed(234)  
 #inTrain = createDataPartition(Carseats$Sales, p = 0.6, list = FALSE)  
 Total\_data\_size=as.integer(nrow(BostonH))  
 inTrain = sample(1:Total\_data\_size, round(Total\_data\_size\*.60))  
 Training\_Set = BostonH[inTrain, ]  
 dim(Training\_Set)

## [1] 304 14

Testing\_Set<-BostonH[-inTrain, ]  
 dim(Testing\_Set)

## [1] 202 14

names(BostonH)

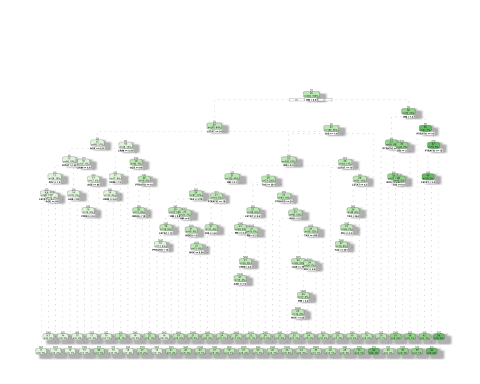
## [1] "CRIM" "ZN" "INDUS" "CHAS" "NOX" "RM"   
## [7] "AGE" "DIS" "RAD" "TAX" "PTRATIO" "LSTAT"   
## [13] "MEDV" "CAT..MEDV"

### B2.

Fit a regression tree on the Training Set using the MEDV', the median price of houses in a region, as the response variable while all the other variables EXCEPT theCAT..MEDV’ as the covariates. Display/plot the fitted tree.

reg\_fit\_train = rpart(MEDV~CRIM+ZN+INDUS+CHAS+NOX+RM+AGE+DIS+RAD+TAX+PTRATIO+LSTAT,data=Training\_Set,method="anova",minsplit=10,cp=-1)  
  
# plot fitted tree  
fancyRpartPlot(reg\_fit\_train, caption = NULL)

## Warning: labs do not fit even at cex 0.15, there may be some overplotting



### B3. Comment on the Error Rate in the validation part.

Identify an optimal value for complexiety parameter `cp’.

summary(reg\_fit\_train)

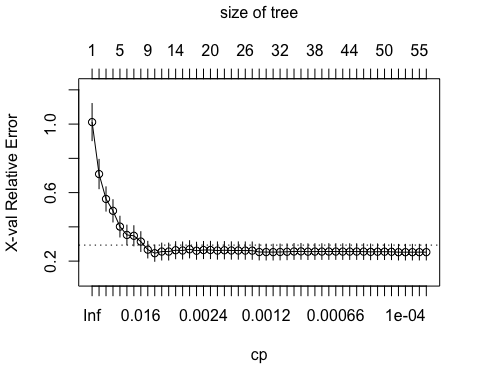
## Call:  
## rpart(formula = MEDV ~ CRIM + ZN + INDUS + CHAS + NOX + RM +   
## AGE + DIS + RAD + TAX + PTRATIO + LSTAT, data = Training\_Set,   
## method = "anova", minsplit = 10, cp = -1)  
## n= 304   
##   
## CP nsplit rel error xerror xstd  
## 1 4.694095e-01 0 1.00000000 1.0116921 0.10837895  
## 2 1.462492e-01 1 0.53059055 0.7084272 0.08551202  
## 3 1.003575e-01 2 0.38434131 0.5627995 0.07072420  
## 4 6.912512e-02 3 0.28398383 0.4936381 0.06499579  
## 5 4.108818e-02 4 0.21485871 0.4014709 0.06044461  
## 6 2.135148e-02 5 0.17377053 0.3527239 0.05726232  
## 7 2.121130e-02 6 0.15241905 0.3478926 0.05864717  
## 8 1.199707e-02 7 0.13120775 0.3141176 0.05768518  
## 9 6.738840e-03 8 0.11921068 0.2668952 0.04835341  
## 10 6.523295e-03 9 0.11247184 0.2462870 0.04721017  
## 11 4.392687e-03 10 0.10594855 0.2557983 0.04923260  
## 12 3.613822e-03 11 0.10155586 0.2557230 0.04922291  
## 13 3.268399e-03 13 0.09432821 0.2636671 0.04971929  
## 14 3.220896e-03 14 0.09105981 0.2620812 0.04968381  
## 15 3.182450e-03 15 0.08783892 0.2690708 0.05006567  
## 16 2.472982e-03 17 0.08147402 0.2606376 0.04965900  
## 17 2.345312e-03 18 0.07900104 0.2647593 0.04990972  
## 18 2.031008e-03 19 0.07665572 0.2650563 0.05033385  
## 19 1.833160e-03 20 0.07462472 0.2617809 0.05030700  
## 20 1.763265e-03 21 0.07279156 0.2634976 0.05051524  
## 21 1.524593e-03 22 0.07102829 0.2621032 0.05045542  
## 22 1.456666e-03 24 0.06797910 0.2619675 0.05044618  
## 23 1.380497e-03 25 0.06652244 0.2615355 0.05045238  
## 24 1.334825e-03 26 0.06514194 0.2612389 0.05044247  
## 25 1.240845e-03 27 0.06380712 0.2531459 0.04615447  
## 26 1.183344e-03 28 0.06256627 0.2528892 0.04613778  
## 27 1.149371e-03 30 0.06019958 0.2526916 0.04607407  
## 28 1.101504e-03 31 0.05905021 0.2533379 0.04609731  
## 29 1.024189e-03 33 0.05684720 0.2544382 0.04611685  
## 30 9.805197e-04 34 0.05582301 0.2562454 0.04613944  
## 31 8.357757e-04 35 0.05484249 0.2568518 0.04614560  
## 32 7.468992e-04 36 0.05400672 0.2550853 0.04571234  
## 33 7.291792e-04 37 0.05325982 0.2556557 0.04570651  
## 34 7.147259e-04 39 0.05180146 0.2556557 0.04570651  
## 35 6.670202e-04 40 0.05108674 0.2560000 0.04570175  
## 36 6.602880e-04 41 0.05041972 0.2557674 0.04568031  
## 37 5.214363e-04 42 0.04975943 0.2555897 0.04563482  
## 38 4.694949e-04 43 0.04923799 0.2555648 0.04562457  
## 39 4.601024e-04 44 0.04876850 0.2556363 0.04562333  
## 40 4.522232e-04 45 0.04830839 0.2554967 0.04562489  
## 41 4.066649e-04 46 0.04785617 0.2538991 0.04559449  
## 42 3.206623e-04 47 0.04744951 0.2548707 0.04559948  
## 43 3.173233e-04 49 0.04680818 0.2549404 0.04559495  
## 44 2.843224e-04 50 0.04649086 0.2544605 0.04560016  
## 45 1.134724e-04 51 0.04620654 0.2521131 0.04557289  
## 46 9.268411e-05 52 0.04609306 0.2522639 0.04557408  
## 47 8.508791e-05 53 0.04600038 0.2522639 0.04557408  
## 48 5.324232e-05 54 0.04591529 0.2524097 0.04557161  
## 49 -1.000000e+00 55 0.04586205 0.2524097 0.04557161  
##   
## Variable importance  
## RM LSTAT DIS INDUS NOX CRIM PTRATIO AGE ZN TAX   
## 31 21 9 8 7 7 6 5 2 2   
## RAD CHAS   
## 1 1   
##   
## Node number 1: 304 observations, complexity param=0.4694095  
## mean=23.21447, MSE=86.59926   
## left son=2 (255 obs) right son=3 (49 obs)  
## Primary splits:  
## RM < 6.92 to the left, improve=0.4694095, (0 missing)  
## LSTAT < 4.65 to the right, improve=0.4648085, (0 missing)  
## INDUS < 6.66 to the right, improve=0.2427035, (0 missing)  
## PTRATIO < 18.75 to the right, improve=0.2282768, (0 missing)  
## NOX < 0.6695 to the right, improve=0.1909624, (0 missing)  
## Surrogate splits:  
## LSTAT < 4.6 to the right, agree=0.891, adj=0.327, (0 split)  
## PTRATIO < 14.55 to the right, agree=0.862, adj=0.143, (0 split)  
## INDUS < 1.605 to the right, agree=0.855, adj=0.102, (0 split)  
## ZN < 85 to the left, agree=0.852, adj=0.082, (0 split)  
## NOX < 0.4045 to the right, agree=0.845, adj=0.041, (0 split)  
##   
## Node number 2: 255 observations, complexity param=0.1462492  
## mean=20.41961, MSE=39.00989   
## left son=4 (95 obs) right son=5 (160 obs)  
## Primary splits:  
## LSTAT < 14.395 to the right, improve=0.3870495, (0 missing)  
## NOX < 0.6695 to the right, improve=0.2362884, (0 missing)  
## CRIM < 8.37969 to the right, improve=0.2147705, (0 missing)  
## AGE < 90.15 to the right, improve=0.1819709, (0 missing)  
## DIS < 2.57205 to the left, improve=0.1813514, (0 missing)  
## Surrogate splits:  
## AGE < 88.2 to the right, agree=0.824, adj=0.526, (0 split)  
## NOX < 0.565 to the right, agree=0.800, adj=0.463, (0 split)  
## DIS < 2.6182 to the left, agree=0.796, adj=0.453, (0 split)  
## INDUS < 16.57 to the right, agree=0.784, adj=0.421, (0 split)  
## CRIM < 4.067905 to the right, agree=0.780, adj=0.411, (0 split)  
##   
## Node number 3: 49 observations, complexity param=0.1003575  
## mean=37.75918, MSE=82.05915   
## left son=6 (28 obs) right son=7 (21 obs)  
## Primary splits:  
## RM < 7.435 to the left, improve=0.6570743, (0 missing)  
## LSTAT < 4.68 to the right, improve=0.3856238, (0 missing)  
## PTRATIO < 19.45 to the right, improve=0.3254822, (0 missing)  
## CRIM < 3.870675 to the right, improve=0.2717869, (0 missing)  
## RAD < 16 to the right, improve=0.2717869, (0 missing)  
## Surrogate splits:  
## LSTAT < 3.99 to the right, agree=0.816, adj=0.571, (0 split)  
## CRIM < 0.11276 to the left, agree=0.633, adj=0.143, (0 split)  
## CHAS < 0.5 to the left, agree=0.633, adj=0.143, (0 split)  
## PTRATIO < 14.75 to the right, agree=0.633, adj=0.143, (0 split)  
## INDUS < 1.295 to the right, agree=0.612, adj=0.095, (0 split)  
##   
## Node number 4: 95 observations, complexity param=0.0212113  
## mean=15.37684, MSE=18.22536   
## left son=8 (56 obs) right son=9 (39 obs)  
## Primary splits:  
## NOX < 0.607 to the right, improve=0.3225190, (0 missing)  
## DIS < 2.0037 to the left, improve=0.2884914, (0 missing)  
## CRIM < 7.006285 to the right, improve=0.2547434, (0 missing)  
## LSTAT < 21.785 to the right, improve=0.2457841, (0 missing)  
## TAX < 567.5 to the right, improve=0.2022054, (0 missing)  
## Surrogate splits:  
## DIS < 2.38405 to the left, agree=0.863, adj=0.667, (0 split)  
## TAX < 397 to the right, agree=0.853, adj=0.641, (0 split)  
## INDUS < 16.01 to the right, agree=0.832, adj=0.590, (0 split)  
## CRIM < 1.40092 to the right, agree=0.800, adj=0.513, (0 split)  
## AGE < 88.8 to the right, agree=0.726, adj=0.333, (0 split)  
##   
## Node number 5: 160 observations, complexity param=0.06912512  
## mean=23.41375, MSE=27.28706   
## left son=10 (156 obs) right son=11 (4 obs)  
## Primary splits:  
## DIS < 1.38485 to the right, improve=0.41681840, (0 missing)  
## LSTAT < 4.52 to the right, improve=0.29115770, (0 missing)  
## RM < 6.543 to the left, improve=0.26486240, (0 missing)  
## CHAS < 0.5 to the left, improve=0.10593260, (0 missing)  
## CRIM < 4.866945 to the left, improve=0.08719906, (0 missing)  
## Surrogate splits:  
## LSTAT < 3.745 to the right, agree=0.981, adj=0.25, (0 split)  
##   
## Node number 6: 28 observations, complexity param=0.02135148  
## mean=31.4, MSE=34.02929   
## left son=12 (4 obs) right son=13 (24 obs)  
## Primary splits:  
## PTRATIO < 19.45 to the right, improve=0.5899361, (0 missing)  
## TAX < 534.5 to the right, improve=0.5191450, (0 missing)  
## RAD < 16 to the right, improve=0.5191450, (0 missing)  
## CRIM < 3.47737 to the right, improve=0.5191450, (0 missing)  
## LSTAT < 8.905 to the right, improve=0.3771292, (0 missing)  
## Surrogate splits:  
## CRIM < 3.47737 to the right, agree=0.964, adj=0.75, (0 split)  
## INDUS < 9.5 to the right, agree=0.964, adj=0.75, (0 split)  
## RAD < 16 to the right, agree=0.964, adj=0.75, (0 split)  
## TAX < 534.5 to the right, agree=0.964, adj=0.75, (0 split)  
## NOX < 0.659 to the right, agree=0.929, adj=0.50, (0 split)  
##   
## Node number 7: 21 observations, complexity param=0.006523295  
## mean=46.2381, MSE=20.28807   
## left son=14 (3 obs) right son=15 (18 obs)  
## Primary splits:  
## PTRATIO < 17.9 to the right, improve=0.4030832, (0 missing)  
## DIS < 3.20745 to the right, improve=0.2661019, (0 missing)  
## NOX < 0.541 to the left, improve=0.1910577, (0 missing)  
## INDUS < 2.785 to the right, improve=0.1641292, (0 missing)  
## CRIM < 0.576815 to the left, improve=0.1641292, (0 missing)  
## Surrogate splits:  
## RAD < 2.5 to the left, agree=0.905, adj=0.333, (0 split)  
##   
## Node number 8: 56 observations, complexity param=0.01199707  
## mean=13.35357, MSE=13.03713   
## left son=16 (29 obs) right son=17 (27 obs)  
## Primary splits:  
## LSTAT < 20.07 to the right, improve=0.4326064, (0 missing)  
## CRIM < 8.91589 to the right, improve=0.2840338, (0 missing)  
## DIS < 2.0037 to the left, improve=0.2061175, (0 missing)  
## RM < 5.441 to the left, improve=0.1625607, (0 missing)  
## TAX < 551.5 to the right, improve=0.1315007, (0 missing)  
## Surrogate splits:  
## CRIM < 8.91589 to the right, agree=0.750, adj=0.481, (0 split)  
## DIS < 1.61755 to the left, agree=0.750, adj=0.481, (0 split)  
## RM < 5.6275 to the left, agree=0.696, adj=0.370, (0 split)  
## AGE < 98.75 to the right, agree=0.643, adj=0.259, (0 split)  
## NOX < 0.7065 to the left, agree=0.625, adj=0.222, (0 split)  
##   
## Node number 9: 39 observations, complexity param=0.00318245  
## mean=18.28205, MSE=11.35686   
## left son=18 (17 obs) right son=19 (22 obs)  
## Primary splits:  
## CRIM < 0.55381 to the right, improve=0.15302980, (0 missing)  
## PTRATIO < 20.95 to the right, improve=0.14718120, (0 missing)  
## NOX < 0.531 to the right, improve=0.11793450, (0 missing)  
## AGE < 58.85 to the right, improve=0.09147999, (0 missing)  
## TAX < 300 to the right, improve=0.07776077, (0 missing)  
## Surrogate splits:  
## PTRATIO < 19.7 to the right, agree=0.897, adj=0.765, (0 split)  
## RAD < 16 to the right, agree=0.846, adj=0.647, (0 split)  
## TAX < 567.5 to the right, agree=0.846, adj=0.647, (0 split)  
## NOX < 0.531 to the right, agree=0.795, adj=0.529, (0 split)  
## INDUS < 16.01 to the right, agree=0.769, adj=0.471, (0 split)  
##   
## Node number 10: 156 observations, complexity param=0.04108818  
## mean=22.87372, MSE=13.97322   
## left son=20 (122 obs) right son=21 (34 obs)  
## Primary splits:  
## RM < 6.543 to the left, improve=0.4962307, (0 missing)  
## LSTAT < 7.76 to the right, improve=0.3414905, (0 missing)  
## NOX < 0.5125 to the right, improve=0.2103137, (0 missing)  
## PTRATIO < 18.1 to the right, improve=0.1445624, (0 missing)  
## INDUS < 6.66 to the right, improve=0.1400971, (0 missing)  
## Surrogate splits:  
## LSTAT < 5.055 to the right, agree=0.859, adj=0.353, (0 split)  
## ZN < 31.5 to the left, agree=0.827, adj=0.206, (0 split)  
## CRIM < 0.01837 to the right, agree=0.801, adj=0.088, (0 split)  
## INDUS < 3.095 to the right, agree=0.795, adj=0.059, (0 split)  
## CHAS < 0.5 to the left, agree=0.788, adj=0.029, (0 split)  
##   
## Node number 11: 4 observations  
## mean=44.475, MSE=91.57688   
##   
## Node number 12: 4 observations  
## mean=20.425, MSE=17.33187   
##   
## Node number 13: 24 observations, complexity param=0.003613822  
## mean=33.22917, MSE=13.39123   
## left son=26 (21 obs) right son=27 (3 obs)  
## Primary splits:  
## DIS < 2.00445 to the right, improve=0.2360733, (0 missing)  
## AGE < 90.05 to the left, improve=0.2154123, (0 missing)  
## TAX < 378 to the left, improve=0.1597127, (0 missing)  
## LSTAT < 4.66 to the right, improve=0.1581907, (0 missing)  
## CRIM < 0.53095 to the left, improve=0.1318154, (0 missing)  
## Surrogate splits:  
## CRIM < 0.53095 to the left, agree=0.958, adj=0.667, (0 split)  
## NOX < 0.556 to the left, agree=0.958, adj=0.667, (0 split)  
## AGE < 95.95 to the left, agree=0.958, adj=0.667, (0 split)  
## PTRATIO < 14.75 to the right, agree=0.917, adj=0.333, (0 split)  
##   
## Node number 14: 3 observations  
## mean=39.23333, MSE=12.46889   
##   
## Node number 15: 18 observations, complexity param=0.003268399  
## mean=47.40556, MSE=12.05052   
## left son=30 (10 obs) right son=31 (8 obs)  
## Primary splits:  
## PTRATIO < 14.8 to the right, improve=0.3966837, (0 missing)  
## RAD < 6.5 to the right, improve=0.2709784, (0 missing)  
## DIS < 3.20745 to the right, improve=0.2137687, (0 missing)  
## NOX < 0.541 to the left, improve=0.1769284, (0 missing)  
## LSTAT < 3.145 to the left, improve=0.1694421, (0 missing)  
## Surrogate splits:  
## NOX < 0.541 to the left, agree=0.833, adj=0.625, (0 split)  
## DIS < 2.658 to the right, agree=0.833, adj=0.625, (0 split)  
## CRIM < 0.576815 to the left, agree=0.778, adj=0.500, (0 split)  
## AGE < 89 to the left, agree=0.778, adj=0.500, (0 split)  
## RAD < 6.5 to the right, agree=0.778, adj=0.500, (0 split)  
##   
## Node number 16: 29 observations, complexity param=0.003220896  
## mean=11.06207, MSE=6.914768   
## left son=32 (20 obs) right son=33 (9 obs)  
## Primary splits:  
## DIS < 1.4226 to the right, improve=0.4228525, (0 missing)  
## TAX < 551.5 to the right, improve=0.2128475, (0 missing)  
## NOX < 0.8055 to the left, improve=0.1549700, (0 missing)  
## PTRATIO < 17.4 to the right, improve=0.1549700, (0 missing)  
## LSTAT < 29.485 to the right, improve=0.1451968, (0 missing)  
## Surrogate splits:  
## NOX < 0.8055 to the left, agree=0.828, adj=0.444, (0 split)  
## TAX < 420 to the right, agree=0.828, adj=0.444, (0 split)  
## PTRATIO < 17.4 to the right, agree=0.828, adj=0.444, (0 split)  
## CRIM < 5.57391 to the right, agree=0.724, adj=0.111, (0 split)  
## INDUS < 18.84 to the left, agree=0.724, adj=0.111, (0 split)  
##   
## Node number 17: 27 observations, complexity param=0.002031008  
## mean=15.81481, MSE=7.915336   
## left son=34 (10 obs) right son=35 (17 obs)  
## Primary splits:  
## CRIM < 6.88166 to the right, improve=0.25018780, (0 missing)  
## RM < 6.168 to the right, improve=0.23562120, (0 missing)  
## DIS < 2.0037 to the left, improve=0.07191341, (0 missing)  
## AGE < 90.4 to the right, improve=0.05195677, (0 missing)  
## INDUS < 18.84 to the left, improve=0.04704191, (0 missing)  
## Surrogate splits:  
## RM < 6.168 to the right, agree=0.741, adj=0.3, (0 split)  
## INDUS < 18.84 to the left, agree=0.704, adj=0.2, (0 split)  
## RAD < 14.5 to the right, agree=0.704, adj=0.2, (0 split)  
## TAX < 551.5 to the right, agree=0.704, adj=0.2, (0 split)  
## AGE < 84.05 to the left, agree=0.667, adj=0.1, (0 split)  
##   
## Node number 18: 17 observations, complexity param=0.00318245  
## mean=16.78235, MSE=14.94734   
## left son=36 (13 obs) right son=37 (4 obs)  
## Primary splits:  
## CRIM < 12.66115 to the left, improve=0.3926883, (0 missing)  
## NOX < 0.5905 to the left, improve=0.2470192, (0 missing)  
## RM < 5.9765 to the left, improve=0.1849005, (0 missing)  
## AGE < 97.7 to the left, improve=0.1453654, (0 missing)  
## DIS < 3.76795 to the right, improve=0.1426743, (0 missing)  
## Surrogate splits:  
## DIS < 1.77575 to the right, agree=0.882, adj=0.50, (0 split)  
## NOX < 0.5905 to the left, agree=0.824, adj=0.25, (0 split)  
## AGE < 75.4 to the right, agree=0.824, adj=0.25, (0 split)  
##   
## Node number 19: 22 observations, complexity param=0.001240845  
## mean=19.44091, MSE=5.501508   
## left son=38 (7 obs) right son=39 (15 obs)  
## Primary splits:  
## AGE < 94.1 to the right, improve=0.2698988, (0 missing)  
## DIS < 5.57015 to the right, improve=0.2359684, (0 missing)  
## RAD < 3.5 to the left, improve=0.1657951, (0 missing)  
## TAX < 255 to the left, improve=0.1618466, (0 missing)  
## NOX < 0.4685 to the left, improve=0.1444265, (0 missing)  
## Surrogate splits:  
## LSTAT < 22.275 to the right, agree=0.773, adj=0.286, (0 split)  
## CRIM < 0.381565 to the right, agree=0.727, adj=0.143, (0 split)  
## RM < 5.4015 to the left, agree=0.727, adj=0.143, (0 split)  
## DIS < 2.0466 to the left, agree=0.727, adj=0.143, (0 split)  
##   
## Node number 20: 122 observations, complexity param=0.00673884  
## mean=21.48361, MSE=6.049403   
## left son=40 (51 obs) right son=41 (71 obs)  
## Primary splits:  
## RM < 6.062 to the left, improve=0.24038120, (0 missing)  
## LSTAT < 7.76 to the right, improve=0.19410820, (0 missing)  
## NOX < 0.5125 to the right, improve=0.14275900, (0 missing)  
## AGE < 18.45 to the right, improve=0.10822090, (0 missing)  
## TAX < 302 to the right, improve=0.09300501, (0 missing)  
## Surrogate splits:  
## LSTAT < 11.395 to the right, agree=0.648, adj=0.157, (0 split)  
## NOX < 0.522 to the right, agree=0.639, adj=0.137, (0 split)  
## DIS < 2.11615 to the left, agree=0.623, adj=0.098, (0 split)  
## PTRATIO < 20.95 to the right, agree=0.623, adj=0.098, (0 split)  
## TAX < 228.5 to the left, agree=0.615, adj=0.078, (0 split)  
##   
## Node number 21: 34 observations, complexity param=0.004392687  
## mean=27.86176, MSE=10.59119   
## left son=42 (3 obs) right son=43 (31 obs)  
## Primary splits:  
## LSTAT < 10.045 to the right, improve=0.3211401, (0 missing)  
## NOX < 0.524 to the right, improve=0.2606937, (0 missing)  
## AGE < 41.65 to the right, improve=0.2286956, (0 missing)  
## TAX < 278 to the right, improve=0.1697309, (0 missing)  
## INDUS < 6.66 to the right, improve=0.1561881, (0 missing)  
##   
## Node number 26: 21 observations, complexity param=0.003613822  
## mean=32.55714, MSE=10.87578   
## left son=52 (6 obs) right son=53 (15 obs)  
## Primary splits:  
## NOX < 0.4885 to the right, improve=0.5009153, (0 missing)  
## AGE < 62.8 to the right, improve=0.2648085, (0 missing)  
## DIS < 3.45845 to the left, improve=0.2546765, (0 missing)  
## RM < 7.0835 to the left, improve=0.2462314, (0 missing)  
## TAX < 248 to the right, improve=0.2347100, (0 missing)  
## Surrogate splits:  
## AGE < 62.8 to the right, agree=0.952, adj=0.833, (0 split)  
## CRIM < 0.23139 to the right, agree=0.905, adj=0.667, (0 split)  
## DIS < 3.45845 to the left, agree=0.857, adj=0.500, (0 split)  
## RAD < 7.5 to the right, agree=0.857, adj=0.500, (0 split)  
## INDUS < 3.37 to the right, agree=0.810, adj=0.333, (0 split)  
##   
## Node number 27: 3 observations  
## mean=37.93333, MSE=5.708889   
##   
## Node number 30: 10 observations, complexity param=0.002472982  
## mean=45.45, MSE=12.9605   
## left son=60 (4 obs) right son=61 (6 obs)  
## Primary splits:  
## LSTAT < 3.445 to the left, improve=0.50232760, (0 missing)  
## DIS < 3.20745 to the right, improve=0.26234470, (0 missing)  
## AGE < 71.85 to the right, improve=0.19061120, (0 missing)  
## CRIM < 0.356805 to the right, improve=0.18261530, (0 missing)  
## INDUS < 3.54 to the right, improve=0.09370046, (0 missing)  
## Surrogate splits:  
## RM < 7.6655 to the left, agree=0.8, adj=0.50, (0 split)  
## DIS < 3.74515 to the right, agree=0.8, adj=0.50, (0 split)  
## CRIM < 0.356805 to the right, agree=0.7, adj=0.25, (0 split)  
## ZN < 10 to the right, agree=0.7, adj=0.25, (0 split)  
## NOX < 0.46545 to the left, agree=0.7, adj=0.25, (0 split)  
##   
## Node number 31: 8 observations  
## mean=49.85, MSE=0.1575   
##   
## Node number 32: 20 observations, complexity param=0.00183316  
## mean=9.915, MSE=5.395275   
## left son=64 (8 obs) right son=65 (12 obs)  
## Primary splits:  
## LSTAT < 24.975 to the right, improve=0.4472440, (0 missing)  
## DIS < 1.97805 to the left, improve=0.2849662, (0 missing)  
## RM < 5.475 to the left, improve=0.2616754, (0 missing)  
## AGE < 97.3 to the right, improve=0.1774284, (0 missing)  
## NOX < 0.72 to the left, improve=0.1597961, (0 missing)  
## Surrogate splits:  
## RM < 5.3455 to the left, agree=0.85, adj=0.625, (0 split)  
## DIS < 1.5098 to the left, agree=0.80, adj=0.500, (0 split)  
## CRIM < 13.98975 to the right, agree=0.70, adj=0.250, (0 split)  
## AGE < 91.8 to the left, agree=0.70, adj=0.250, (0 split)  
##   
## Node number 33: 9 observations  
## mean=13.61111, MSE=0.8698765   
##   
## Node number 34: 10 observations, complexity param=0.0007147259  
## mean=13.98, MSE=4.9296   
## left son=68 (4 obs) right son=69 (6 obs)  
## Primary splits:  
## AGE < 95 to the left, improve=0.38169430, (0 missing)  
## CRIM < 10.6602 to the right, improve=0.21703300, (0 missing)  
## DIS < 2.03385 to the left, improve=0.11716980, (0 missing)  
## RM < 6.4085 to the right, improve=0.06732508, (0 missing)  
## NOX < 0.6965 to the left, improve=0.06732508, (0 missing)  
## Surrogate splits:  
## CRIM < 12.7541 to the right, agree=0.8, adj=0.50, (0 split)  
## LSTAT < 16.67 to the left, agree=0.8, adj=0.50, (0 split)  
## NOX < 0.6965 to the left, agree=0.7, adj=0.25, (0 split)  
## RM < 6.189 to the right, agree=0.7, adj=0.25, (0 split)  
## DIS < 1.8947 to the left, agree=0.7, adj=0.25, (0 split)  
##   
## Node number 35: 17 observations, complexity param=0.001334825  
## mean=16.89412, MSE=6.526436   
## left son=70 (14 obs) right son=71 (3 obs)  
## Primary splits:  
## AGE < 90.4 to the right, improve=0.31672850, (0 missing)  
## RM < 5.673 to the left, improve=0.21530950, (0 missing)  
## DIS < 2.1199 to the left, improve=0.18145430, (0 missing)  
## LSTAT < 15.965 to the right, improve=0.12491510, (0 missing)  
## CRIM < 5.76921 to the right, improve=0.03930476, (0 missing)  
## Surrogate splits:  
## LSTAT < 14.865 to the right, agree=0.882, adj=0.333, (0 split)  
##   
## Node number 36: 13 observations, complexity param=0.001149371  
## mean=15.43846, MSE=7.548521   
## left son=72 (4 obs) right son=73 (9 obs)  
## Primary splits:  
## CRIM < 6.87377 to the right, improve=0.30834920, (0 missing)  
## RM < 5.9765 to the left, improve=0.29938390, (0 missing)  
## DIS < 2.25705 to the left, improve=0.24941310, (0 missing)  
## LSTAT < 15.335 to the right, improve=0.20326720, (0 missing)  
## PTRATIO < 20.6 to the right, improve=0.06984577, (0 missing)  
## Surrogate splits:  
## DIS < 2.25705 to the left, agree=0.923, adj=0.75, (0 split)  
## NOX < 0.5835 to the right, agree=0.846, adj=0.50, (0 split)  
## RM < 6.2575 to the right, agree=0.846, adj=0.50, (0 split)  
## RAD < 14.5 to the right, agree=0.769, adj=0.25, (0 split)  
## TAX < 534.5 to the right, agree=0.769, adj=0.25, (0 split)  
##   
## Node number 37: 4 observations  
## mean=21.15, MSE=14.0475   
##   
## Node number 38: 7 observations  
## mean=17.65714, MSE=3.865306   
##   
## Node number 39: 15 observations, complexity param=0.0007291792  
## mean=20.27333, MSE=4.087289   
## left son=78 (11 obs) right son=79 (4 obs)  
## Primary splits:  
## PTRATIO < 17.65 to the right, improve=0.1810827, (0 missing)  
## TAX < 347.5 to the right, improve=0.1739616, (0 missing)  
## DIS < 5.57015 to the right, improve=0.1712654, (0 missing)  
## NOX < 0.4685 to the left, improve=0.1712654, (0 missing)  
## INDUS < 10.245 to the left, improve=0.1663858, (0 missing)  
## Surrogate splits:  
## CRIM < 0.155025 to the right, agree=0.8, adj=0.25, (0 split)  
## INDUS < 6.555 to the right, agree=0.8, adj=0.25, (0 split)  
##   
## Node number 40: 51 observations, complexity param=0.001183344  
## mean=20.06078, MSE=4.695717   
## left son=80 (37 obs) right son=81 (14 obs)  
## Primary splits:  
## TAX < 278 to the right, improve=0.12805880, (0 missing)  
## LSTAT < 13.305 to the left, improve=0.08235965, (0 missing)  
## RM < 5.7595 to the left, improve=0.06247063, (0 missing)  
## DIS < 1.68125 to the left, improve=0.05942835, (0 missing)  
## NOX < 0.541 to the right, improve=0.03976431, (0 missing)  
## Surrogate splits:  
## RAD < 3.5 to the right, agree=0.824, adj=0.357, (0 split)  
## CRIM < 0.05643 to the right, agree=0.765, adj=0.143, (0 split)  
## INDUS < 5.8 to the right, agree=0.765, adj=0.143, (0 split)  
## RM < 6.034 to the left, agree=0.765, adj=0.143, (0 split)  
## LSTAT < 13.63 to the left, agree=0.765, adj=0.143, (0 split)  
##   
## Node number 41: 71 observations, complexity param=0.002345312  
## mean=22.50563, MSE=4.523067   
## left son=82 (28 obs) right son=83 (43 obs)  
## Primary splits:  
## TAX < 360.5 to the right, improve=0.1922636, (0 missing)  
## NOX < 0.5125 to the right, improve=0.1911894, (0 missing)  
## PTRATIO < 19.95 to the right, improve=0.1765554, (0 missing)  
## LSTAT < 7.995 to the right, improve=0.1701744, (0 missing)  
## AGE < 30.35 to the right, improve=0.1457954, (0 missing)  
## Surrogate splits:  
## NOX < 0.5165 to the right, agree=0.873, adj=0.679, (0 split)  
## INDUS < 11.82 to the right, agree=0.831, adj=0.571, (0 split)  
## DIS < 3.02775 to the left, agree=0.831, adj=0.571, (0 split)  
## CRIM < 1.066625 to the right, agree=0.803, adj=0.500, (0 split)  
## AGE < 71.05 to the right, agree=0.803, adj=0.500, (0 split)  
##   
## Node number 42: 3 observations  
## mean=21.93333, MSE=0.5422222   
##   
## Node number 43: 31 observations, complexity param=0.001763265  
## mean=28.43548, MSE=7.833257   
## left son=86 (25 obs) right son=87 (6 obs)  
## Primary splits:  
## LSTAT < 4.52 to the right, improve=0.1911619, (0 missing)  
## AGE < 41.65 to the right, improve=0.1579540, (0 missing)  
## NOX < 0.524 to the right, improve=0.1461511, (0 missing)  
## RM < 6.6385 to the left, improve=0.1436286, (0 missing)  
## TAX < 270.5 to the right, improve=0.1124914, (0 missing)  
## Surrogate splits:  
## ZN < 72.5 to the left, agree=0.871, adj=0.333, (0 split)  
## INDUS < 2.1 to the right, agree=0.839, adj=0.167, (0 split)  
## DIS < 8.0763 to the left, agree=0.839, adj=0.167, (0 split)  
##   
## Node number 52: 6 observations  
## mean=28.86667, MSE=8.662222   
##   
## Node number 53: 15 observations, complexity param=0.0008357757  
## mean=34.03333, MSE=4.134222   
## left son=106 (6 obs) right son=107 (9 obs)  
## Primary splits:  
## DIS < 6.89355 to the right, improve=0.3548072, (0 missing)  
## NOX < 0.42655 to the left, improve=0.3548072, (0 missing)  
## ZN < 57.5 to the right, improve=0.2993765, (0 missing)  
## AGE < 23.55 to the left, improve=0.2797297, (0 missing)  
## RM < 7.145 to the left, improve=0.2386632, (0 missing)  
## Surrogate splits:  
## NOX < 0.42655 to the left, agree=1.000, adj=1.000, (0 split)  
## ZN < 57.5 to the right, agree=0.933, adj=0.833, (0 split)  
## INDUS < 1.85 to the left, agree=0.867, adj=0.667, (0 split)  
## AGE < 23.55 to the left, agree=0.867, adj=0.667, (0 split)  
## CRIM < 0.0234 to the left, agree=0.800, adj=0.500, (0 split)  
##   
## Node number 60: 4 observations  
## mean=42.325, MSE=9.756875   
##   
## Node number 61: 6 observations  
## mean=47.53333, MSE=4.245556   
##   
## Node number 64: 8 observations  
## mean=8.0125, MSE=1.521094   
##   
## Node number 65: 12 observations, complexity param=0.0009805197  
## mean=11.18333, MSE=3.956389   
## left son=130 (6 obs) right son=131 (6 obs)  
## Primary splits:  
## AGE < 95.2 to the right, improve=0.5437057, (0 missing)  
## RM < 6.03 to the left, improve=0.2444008, (0 missing)  
## DIS < 1.97805 to the left, improve=0.2387372, (0 missing)  
## CRIM < 11.98135 to the right, improve=0.1100892, (0 missing)  
## LSTAT < 22.495 to the left, improve=0.1050575, (0 missing)  
## Surrogate splits:  
## NOX < 0.72 to the left, agree=0.750, adj=0.500, (0 split)  
## DIS < 1.8099 to the left, agree=0.750, adj=0.500, (0 split)  
## CRIM < 11.98135 to the right, agree=0.667, adj=0.333, (0 split)  
## RM < 5.5815 to the left, agree=0.667, adj=0.333, (0 split)  
## LSTAT < 21.595 to the left, agree=0.667, adj=0.333, (0 split)  
##   
## Node number 68: 4 observations  
## mean=12.3, MSE=4.565   
##   
## Node number 69: 6 observations  
## mean=15.1, MSE=2.036667   
##   
## Node number 70: 14 observations, complexity param=0.000660288  
## mean=16.22857, MSE=3.952041   
## left son=140 (7 obs) right son=141 (7 obs)  
## Primary splits:  
## CRIM < 2.152115 to the right, improve=0.3141751, (0 missing)  
## LSTAT < 15.965 to the right, improve=0.3087026, (0 missing)  
## RM < 5.673 to the left, improve=0.2478761, (0 missing)  
## AGE < 97.15 to the left, improve=0.1979957, (0 missing)  
## DIS < 1.67275 to the right, improve=0.1297937, (0 missing)  
## Surrogate splits:  
## INDUS < 18.84 to the left, agree=0.857, adj=0.714, (0 split)  
## NOX < 0.6585 to the right, agree=0.857, adj=0.714, (0 split)  
## RAD < 4.5 to the right, agree=0.857, adj=0.714, (0 split)  
## TAX < 551.5 to the right, agree=0.857, adj=0.714, (0 split)  
## PTRATIO < 20.7 to the left, agree=0.857, adj=0.714, (0 split)  
##   
## Node number 71: 3 observations  
## mean=20, MSE=6.826667   
##   
## Node number 72: 4 observations  
## mean=13.15, MSE=2.9625   
##   
## Node number 73: 9 observations  
## mean=16.45556, MSE=6.224691   
##   
## Node number 78: 11 observations, complexity param=0.0007291792  
## mean=19.75455, MSE=4.207934   
## left son=156 (6 obs) right son=157 (5 obs)  
## Primary splits:  
## INDUS < 10.245 to the left, improve=0.5895992, (0 missing)  
## AGE < 72.8 to the right, improve=0.3372908, (0 missing)  
## TAX < 290.5 to the right, improve=0.2372561, (0 missing)  
## RM < 5.8355 to the right, improve=0.1221777, (0 missing)  
## CRIM < 0.24063 to the left, improve=0.1188074, (0 missing)  
## Surrogate splits:  
## TAX < 290.5 to the right, agree=0.909, adj=0.8, (0 split)  
## PTRATIO < 19.15 to the right, agree=0.818, adj=0.6, (0 split)  
## AGE < 86.95 to the left, agree=0.727, adj=0.4, (0 split)  
## DIS < 2.2434 to the right, agree=0.727, adj=0.4, (0 split)  
## RAD < 2.5 to the right, agree=0.727, adj=0.4, (0 split)  
##   
## Node number 79: 4 observations  
## mean=21.7, MSE=0.98   
##   
## Node number 80: 37 observations, complexity param=0.001101504  
## mean=19.58378, MSE=4.144602   
## left son=160 (16 obs) right son=161 (21 obs)  
## Primary splits:  
## RM < 5.7595 to the left, improve=0.14156160, (0 missing)  
## RAD < 7 to the left, improve=0.11703670, (0 missing)  
## PTRATIO < 19.9 to the left, improve=0.10270720, (0 missing)  
## LSTAT < 13.305 to the left, improve=0.08930536, (0 missing)  
## TAX < 549 to the left, improve=0.06829883, (0 missing)  
## Surrogate splits:  
## DIS < 2.10675 to the left, agree=0.757, adj=0.438, (0 split)  
## NOX < 0.6115 to the right, agree=0.730, adj=0.375, (0 split)  
## LSTAT < 10.905 to the right, agree=0.730, adj=0.375, (0 split)  
## CRIM < 0.67778 to the right, agree=0.703, adj=0.313, (0 split)  
## AGE < 85.55 to the right, agree=0.703, adj=0.313, (0 split)  
##   
## Node number 81: 14 observations, complexity param=0.001183344  
## mean=21.32143, MSE=3.961684   
## left son=162 (11 obs) right son=163 (3 obs)  
## Primary splits:  
## PTRATIO < 17.85 to the right, improve=0.5704306, (0 missing)  
## INDUS < 4.84 to the right, improve=0.4577276, (0 missing)  
## DIS < 4.4281 to the right, improve=0.2857474, (0 missing)  
## AGE < 62.15 to the left, improve=0.2343978, (0 missing)  
## TAX < 238 to the left, improve=0.2282592, (0 missing)  
## Surrogate splits:  
## INDUS < 4.23 to the right, agree=0.929, adj=0.667, (0 split)  
## RM < 5.643 to the right, agree=0.929, adj=0.667, (0 split)  
## AGE < 62.15 to the left, agree=0.857, adj=0.333, (0 split)  
## DIS < 3.4527 to the right, agree=0.857, adj=0.333, (0 split)  
##   
## Node number 82: 28 observations, complexity param=0.001456666  
## mean=21.35, MSE=4.378929   
## left son=164 (23 obs) right son=165 (5 obs)  
## Primary splits:  
## LSTAT < 8.22 to the right, improve=0.3127676, (0 missing)  
## NOX < 0.5945 to the left, improve=0.2381227, (0 missing)  
## INDUS < 18.84 to the left, improve=0.1764822, (0 missing)  
## PTRATIO < 15.3 to the right, improve=0.1417907, (0 missing)  
## DIS < 2.1112 to the right, improve=0.1338084, (0 missing)  
## Surrogate splits:  
## AGE < 24.7 to the right, agree=0.893, adj=0.4, (0 split)  
## DIS < 1.93295 to the right, agree=0.857, adj=0.2, (0 split)  
##   
## Node number 83: 43 observations, complexity param=0.001024189  
## mean=23.25814, MSE=3.181038   
## left son=166 (3 obs) right son=167 (40 obs)  
## Primary splits:  
## PTRATIO < 19.95 to the right, improve=0.1971200, (0 missing)  
## AGE < 69.2 to the right, improve=0.1820901, (0 missing)  
## NOX < 0.5125 to the right, improve=0.1651464, (0 missing)  
## RAD < 2.5 to the left, improve=0.1109991, (0 missing)  
## CRIM < 0.370925 to the right, improve=0.1064464, (0 missing)  
## Surrogate splits:  
## NOX < 0.5125 to the right, agree=0.953, adj=0.333, (0 split)  
##   
## Node number 86: 25 observations, complexity param=0.001524593  
## mean=27.836, MSE=7.372704   
## left son=172 (22 obs) right son=173 (3 obs)  
## Primary splits:  
## TAX < 364 to the left, improve=0.2134755, (0 missing)  
## DIS < 6.8985 to the right, improve=0.1818963, (0 missing)  
## RM < 6.6385 to the left, improve=0.1720732, (0 missing)  
## INDUS < 4.27 to the right, improve=0.1476627, (0 missing)  
## NOX < 0.436 to the left, improve=0.1340881, (0 missing)  
## Surrogate splits:  
## PTRATIO < 15.25 to the right, agree=0.92, adj=0.333, (0 split)  
## LSTAT < 4.695 to the right, agree=0.92, adj=0.333, (0 split)  
##   
## Node number 87: 6 observations  
## mean=30.93333, MSE=2.015556   
##   
## Node number 106: 6 observations  
## mean=32.55, MSE=3.189167   
##   
## Node number 107: 9 observations  
## mean=35.02222, MSE=2.319506   
##   
## Node number 130: 6 observations  
## mean=9.716667, MSE=2.224722   
##   
## Node number 131: 6 observations  
## mean=12.65, MSE=1.385833   
##   
## Node number 140: 7 observations  
## mean=15.11429, MSE=3.278367   
##   
## Node number 141: 7 observations  
## mean=17.34286, MSE=2.142449   
##   
## Node number 156: 6 observations  
## mean=18.31667, MSE=1.218056   
##   
## Node number 157: 5 observations  
## mean=21.48, MSE=2.3376   
##   
## Node number 160: 16 observations, complexity param=0.001101504  
## mean=18.70625, MSE=5.145586   
## left son=320 (11 obs) right son=321 (5 obs)  
## Primary splits:  
## LSTAT < 13.215 to the left, improve=0.4407696, (0 missing)  
## NOX < 0.6055 to the left, improve=0.2778884, (0 missing)  
## TAX < 417.5 to the left, improve=0.2430527, (0 missing)  
## RAD < 5.5 to the left, improve=0.2311448, (0 missing)  
## CRIM < 3.15685 to the left, improve=0.2300462, (0 missing)  
## Surrogate splits:  
## CRIM < 3.687585 to the left, agree=0.812, adj=0.4, (0 split)  
## INDUS < 9.955 to the left, agree=0.812, adj=0.4, (0 split)  
## NOX < 0.5455 to the left, agree=0.812, adj=0.4, (0 split)  
## RM < 4.968 to the right, agree=0.812, adj=0.4, (0 split)  
## DIS < 2.0274 to the right, agree=0.812, adj=0.4, (0 split)  
##   
## Node number 161: 21 observations, complexity param=0.0004694949  
## mean=20.25238, MSE=2.348209   
## left son=322 (5 obs) right son=323 (16 obs)  
## Primary splits:  
## RM < 5.996 to the right, improve=0.2506471, (0 missing)  
## CRIM < 0.4432 to the left, improve=0.2479108, (0 missing)  
## RAD < 7 to the left, improve=0.1692147, (0 missing)  
## INDUS < 6.065 to the left, improve=0.1652488, (0 missing)  
## DIS < 6.56935 to the right, improve=0.1646723, (0 missing)  
## Surrogate splits:  
## CRIM < 0.046275 to the left, agree=0.857, adj=0.4, (0 split)  
## RAD < 3.5 to the left, agree=0.857, adj=0.4, (0 split)  
## AGE < 79.65 to the right, agree=0.810, adj=0.2, (0 split)  
## PTRATIO < 17.15 to the left, agree=0.810, adj=0.2, (0 split)  
##   
## Node number 162: 11 observations, complexity param=0.0002843224  
## mean=20.53636, MSE=1.489587   
## left son=324 (6 obs) right son=325 (5 obs)  
## Primary splits:  
## DIS < 4.6206 to the right, improve=0.4568150, (0 missing)  
## LSTAT < 10.54 to the left, improve=0.3547936, (0 missing)  
## AGE < 45.25 to the right, improve=0.2846225, (0 missing)  
## NOX < 0.502 to the right, improve=0.1989643, (0 missing)  
## PTRATIO < 18.85 to the right, improve=0.1989643, (0 missing)  
## Surrogate splits:  
## INDUS < 6.935 to the left, agree=0.909, adj=0.8, (0 split)  
## LSTAT < 10.54 to the left, agree=0.909, adj=0.8, (0 split)  
## RAD < 4.5 to the right, agree=0.818, adj=0.6, (0 split)  
## TAX < 223.5 to the right, agree=0.818, adj=0.6, (0 split)  
## PTRATIO < 19.65 to the right, agree=0.818, adj=0.6, (0 split)  
##   
## Node number 163: 3 observations  
## mean=24.2, MSE=2.48   
##   
## Node number 164: 23 observations, complexity param=0.0006670202  
## mean=20.80435, MSE=3.513459   
## left son=328 (4 obs) right son=329 (19 obs)  
## Primary splits:  
## RM < 6.4135 to the right, improve=0.2173021, (0 missing)  
## NOX < 0.5945 to the left, improve=0.2044513, (0 missing)  
## INDUS < 18.84 to the left, improve=0.2041132, (0 missing)  
## CRIM < 0.40083 to the left, improve=0.1396524, (0 missing)  
## LSTAT < 10.425 to the left, improve=0.1339062, (0 missing)  
##   
## Node number 165: 5 observations  
## mean=23.86, MSE=0.6904   
##   
## Node number 166: 3 observations  
## mean=20.36667, MSE=2.722222   
##   
## Node number 167: 40 observations, complexity param=0.0007468992  
## mean=23.475, MSE=2.541375   
## left son=334 (5 obs) right son=335 (35 obs)  
## Primary splits:  
## RAD < 2.5 to the left, improve=0.19342880, (0 missing)  
## TAX < 237 to the right, improve=0.13767720, (0 missing)  
## AGE < 69.2 to the right, improve=0.10845510, (0 missing)  
## CRIM < 0.048715 to the left, improve=0.10506450, (0 missing)  
## ZN < 41.25 to the right, improve=0.09837194, (0 missing)  
## Surrogate splits:  
## INDUS < 3.73 to the left, agree=0.950, adj=0.6, (0 split)  
## CRIM < 0.0242 to the left, agree=0.925, adj=0.4, (0 split)  
## NOX < 0.3935 to the left, agree=0.925, adj=0.4, (0 split)  
## ZN < 53.75 to the right, agree=0.900, adj=0.2, (0 split)  
## LSTAT < 11.28 to the right, agree=0.900, adj=0.2, (0 split)  
##   
## Node number 172: 22 observations, complexity param=0.001524593  
## mean=27.37273, MSE=6.538347   
## left son=344 (16 obs) right son=345 (6 obs)  
## Primary splits:  
## DIS < 3.2948 to the right, improve=0.2845182, (0 missing)  
## TAX < 278 to the right, improve=0.1896673, (0 missing)  
## LSTAT < 5.005 to the left, improve=0.1425160, (0 missing)  
## NOX < 0.524 to the right, improve=0.1175262, (0 missing)  
## ZN < 9 to the right, improve=0.1143909, (0 missing)  
## Surrogate splits:  
## CRIM < 0.56764 to the left, agree=0.818, adj=0.333, (0 split)  
## RM < 6.564 to the right, agree=0.818, adj=0.333, (0 split)  
## AGE < 68.4 to the left, agree=0.818, adj=0.333, (0 split)  
## TAX < 227.5 to the right, agree=0.773, adj=0.167, (0 split)  
##   
## Node number 173: 3 observations  
## mean=31.23333, MSE=0.3755556   
##   
## Node number 320: 11 observations, complexity param=0.0005214363  
## mean=17.69091, MSE=3.268099   
## left son=640 (3 obs) right son=641 (8 obs)  
## Primary splits:  
## PTRATIO < 18.65 to the left, improve=0.3818573, (0 missing)  
## INDUS < 9.02 to the right, improve=0.2605840, (0 missing)  
## NOX < 0.541 to the right, improve=0.2605840, (0 missing)  
## DIS < 3.15795 to the left, improve=0.1387821, (0 missing)  
## CRIM < 0.964645 to the right, improve=0.1387821, (0 missing)  
## Surrogate splits:  
## RM < 5.187 to the left, agree=0.909, adj=0.667, (0 split)  
## INDUS < 9.02 to the right, agree=0.818, adj=0.333, (0 split)  
## NOX < 0.541 to the right, agree=0.818, adj=0.333, (0 split)  
##   
## Node number 321: 5 observations  
## mean=20.94, MSE=2.0184   
##   
## Node number 322: 5 observations  
## mean=18.88, MSE=0.5096   
##   
## Node number 323: 16 observations, complexity param=0.0003206623  
## mean=20.68125, MSE=2.150273   
## left son=646 (5 obs) right son=647 (11 obs)  
## Primary splits:  
## INDUS < 6.015 to the left, improve=0.1872810, (0 missing)  
## AGE < 68.15 to the right, improve=0.1812330, (0 missing)  
## RM < 5.934 to the left, improve=0.1721305, (0 missing)  
## DIS < 3.5159 to the left, improve=0.1641655, (0 missing)  
## CRIM < 0.4432 to the left, improve=0.1375007, (0 missing)  
## Surrogate splits:  
## CRIM < 0.108955 to the left, agree=0.938, adj=0.8, (0 split)  
## NOX < 0.5015 to the left, agree=0.938, adj=0.8, (0 split)  
## ZN < 18.75 to the right, agree=0.875, adj=0.6, (0 split)  
## DIS < 6.715 to the right, agree=0.875, adj=0.6, (0 split)  
## TAX < 294 to the left, agree=0.875, adj=0.6, (0 split)  
##   
## Node number 324: 6 observations  
## mean=19.78333, MSE=0.7447222   
##   
## Node number 325: 5 observations  
## mean=21.44, MSE=0.8864   
##   
## Node number 328: 4 observations  
## mean=18.9, MSE=8.225   
##   
## Node number 329: 19 observations, complexity param=0.0004522232  
## mean=21.20526, MSE=1.597341   
## left son=658 (16 obs) right son=659 (3 obs)  
## Primary splits:  
## RM < 6.3155 to the left, improve=0.3922739, (0 missing)  
## INDUS < 18.84 to the left, improve=0.3484566, (0 missing)  
## AGE < 93.15 to the left, improve=0.3484566, (0 missing)  
## LSTAT < 11.725 to the right, improve=0.3444098, (0 missing)  
## NOX < 0.526 to the left, improve=0.2861453, (0 missing)  
## Surrogate splits:  
## INDUS < 18.84 to the left, agree=0.947, adj=0.667, (0 split)  
## AGE < 93.15 to the left, agree=0.947, adj=0.667, (0 split)  
## PTRATIO < 15.8 to the right, agree=0.895, adj=0.333, (0 split)  
##   
## Node number 334: 5 observations  
## mean=21.62, MSE=0.7216   
##   
## Node number 335: 35 observations, complexity param=0.0004601024  
## mean=23.74, MSE=2.239543   
## left son=670 (29 obs) right son=671 (6 obs)  
## Primary splits:  
## TAX < 245 to the right, improve=0.15453070, (0 missing)  
## INDUS < 4.27 to the right, improve=0.13967880, (0 missing)  
## LSTAT < 10.925 to the left, improve=0.11535280, (0 missing)  
## RM < 6.1245 to the left, improve=0.08752197, (0 missing)  
## ZN < 23.5 to the right, improve=0.08146211, (0 missing)  
## Surrogate splits:  
## RAD < 3.5 to the right, agree=0.914, adj=0.500, (0 split)  
## RM < 6.499 to the left, agree=0.886, adj=0.333, (0 split)  
##   
## Node number 344: 16 observations, complexity param=0.001380497  
## mean=26.5375, MSE=5.484844   
## left son=688 (7 obs) right son=689 (9 obs)  
## Primary splits:  
## TAX < 291 to the right, improve=0.41413230, (0 missing)  
## RM < 6.6385 to the left, improve=0.27858780, (0 missing)  
## DIS < 6.7684 to the right, improve=0.14205890, (0 missing)  
## LSTAT < 5.005 to the left, improve=0.08934659, (0 missing)  
## NOX < 0.5225 to the right, improve=0.07906476, (0 missing)  
## Surrogate splits:  
## CRIM < 0.288615 to the right, agree=0.75, adj=0.429, (0 split)  
## NOX < 0.4895 to the right, agree=0.75, adj=0.429, (0 split)  
## RM < 6.611 to the left, agree=0.75, adj=0.429, (0 split)  
## AGE < 61.65 to the right, agree=0.75, adj=0.429, (0 split)  
## RAD < 4.5 to the right, agree=0.75, adj=0.429, (0 split)  
##   
## Node number 345: 6 observations  
## mean=29.6, MSE=2.526667   
##   
## Node number 640: 3 observations  
## mean=15.86667, MSE=0.1622222   
##   
## Node number 641: 8 observations  
## mean=18.375, MSE=2.716875   
##   
## Node number 646: 5 observations  
## mean=19.74, MSE=0.8704   
##   
## Node number 647: 11 observations, complexity param=0.0003206623  
## mean=21.10909, MSE=2.146281   
## left son=1294 (8 obs) right son=1295 (3 obs)  
## Primary splits:  
## NOX < 0.535 to the right, improve=0.4422170, (0 missing)  
## DIS < 3.4119 to the left, improve=0.3129009, (0 missing)  
## AGE < 72.4 to the right, improve=0.2843682, (0 missing)  
## INDUS < 9.02 to the right, improve=0.2680644, (0 missing)  
## LSTAT < 11.84 to the right, improve=0.2415351, (0 missing)  
## Surrogate splits:  
## INDUS < 7.17 to the right, agree=0.909, adj=0.667, (0 split)  
## AGE < 41.1 to the right, agree=0.818, adj=0.333, (0 split)  
##   
## Node number 658: 16 observations, complexity param=0.0003173233  
## mean=20.8625, MSE=1.082344   
## left son=1316 (13 obs) right son=1317 (3 obs)  
## Primary splits:  
## CRIM < 4.62009 to the left, improve=0.4823970, (0 missing)  
## AGE < 84.3 to the right, improve=0.2947404, (0 missing)  
## NOX < 0.526 to the left, improve=0.1870940, (0 missing)  
## LSTAT < 11.725 to the right, improve=0.1706760, (0 missing)  
## DIS < 2.7187 to the left, improve=0.1627387, (0 missing)  
##   
## Node number 659: 3 observations  
## mean=23.03333, MSE=0.3755556   
##   
## Node number 670: 29 observations, complexity param=0.0004066649  
## mean=23.47241, MSE=1.039239   
## left son=1340 (9 obs) right son=1341 (20 obs)  
## Primary splits:  
## AGE < 52.65 to the right, improve=0.35523110, (0 missing)  
## DIS < 3.63345 to the left, improve=0.19395930, (0 missing)  
## PTRATIO < 18.55 to the left, improve=0.10617940, (0 missing)  
## TAX < 306 to the left, improve=0.08934249, (0 missing)  
## CRIM < 0.11892 to the left, improve=0.08418051, (0 missing)  
## Surrogate splits:  
## DIS < 3.6589 to the left, agree=0.862, adj=0.556, (0 split)  
## NOX < 0.5 to the right, agree=0.828, adj=0.444, (0 split)  
## LSTAT < 9.99 to the right, agree=0.793, adj=0.333, (0 split)  
## CRIM < 0.31696 to the right, agree=0.759, adj=0.222, (0 split)  
## RM < 6.1245 to the left, agree=0.724, adj=0.111, (0 split)  
##   
## Node number 671: 6 observations  
## mean=25.03333, MSE=6.022222   
##   
## Node number 688: 7 observations  
## mean=24.82857, MSE=3.516327   
##   
## Node number 689: 9 observations  
## mean=27.86667, MSE=2.977778   
##   
## Node number 1294: 8 observations  
## mean=20.5125, MSE=1.163594   
##   
## Node number 1295: 3 observations  
## mean=22.7, MSE=1.286667   
##   
## Node number 1316: 13 observations, complexity param=9.268411e-05  
## mean=20.51538, MSE=0.5782249   
## left son=2632 (7 obs) right son=2633 (6 obs)  
## Primary splits:  
## AGE < 73.5 to the right, improve=0.32460330, (0 missing)  
## NOX < 0.5395 to the left, improve=0.14518810, (0 missing)  
## CRIM < 0.13559 to the right, improve=0.09170029, (0 missing)  
## DIS < 2.7224 to the left, improve=0.08197145, (0 missing)  
## PTRATIO < 19.45 to the right, improve=0.08159611, (0 missing)  
## Surrogate splits:  
## DIS < 2.7224 to the left, agree=0.846, adj=0.667, (0 split)  
## CRIM < 0.0943 to the right, agree=0.769, adj=0.500, (0 split)  
## LSTAT < 11.275 to the right, agree=0.769, adj=0.500, (0 split)  
## NOX < 0.49 to the right, agree=0.692, adj=0.333, (0 split)  
## RM < 6.1385 to the left, agree=0.692, adj=0.333, (0 split)  
##   
## Node number 1317: 3 observations  
## mean=22.36667, MSE=0.4822222   
##   
## Node number 1340: 9 observations  
## mean=22.56667, MSE=0.9088889   
##   
## Node number 1341: 20 observations, complexity param=0.0001134724  
## mean=23.88, MSE=0.5626   
## left son=2682 (17 obs) right son=2683 (3 obs)  
## Primary splits:  
## DIS < 4.56715 to the right, improve=0.2654901, (0 missing)  
## NOX < 0.4195 to the left, improve=0.2426828, (0 missing)  
## TAX < 279 to the right, improve=0.2146396, (0 missing)  
## LSTAT < 8.87 to the left, improve=0.1727693, (0 missing)  
## INDUS < 5.59 to the left, improve=0.1600382, (0 missing)  
## Surrogate splits:  
## TAX < 279 to the right, agree=0.95, adj=0.667, (0 split)  
## NOX < 0.4585 to the left, agree=0.90, adj=0.333, (0 split)  
##   
## Node number 2632: 7 observations  
## mean=20.11429, MSE=0.4269388   
##   
## Node number 2633: 6 observations  
## mean=20.98333, MSE=0.3480556   
##   
## Node number 2682: 17 observations, complexity param=8.508791e-05  
## mean=23.71765, MSE=0.4720415   
## left son=5364 (12 obs) right son=5365 (5 obs)  
## Primary splits:  
## RM < 6.405 to the left, improve=0.2791428, (0 missing)  
## NOX < 0.4195 to the left, improve=0.2013830, (0 missing)  
## TAX < 317.5 to the left, improve=0.1363451, (0 missing)  
## LSTAT < 8.09 to the left, improve=0.1218466, (0 missing)  
## ZN < 27.5 to the right, improve=0.1136986, (0 missing)  
##   
## Node number 2683: 3 observations  
## mean=24.8, MSE=0.08   
##   
## Node number 5364: 12 observations, complexity param=5.324232e-05  
## mean=23.48333, MSE=0.4397222   
## left son=10728 (4 obs) right son=10729 (8 obs)  
## Primary splits:  
## NOX < 0.4195 to the left, improve=0.2656349, (0 missing)  
## LSTAT < 8.09 to the left, improve=0.2293114, (0 missing)  
## AGE < 40.25 to the right, improve=0.1535060, (0 missing)  
## PTRATIO < 19.15 to the right, improve=0.1460044, (0 missing)  
## INDUS < 5.03 to the right, improve=0.1243390, (0 missing)  
## Surrogate splits:  
## ZN < 41.25 to the right, agree=0.833, adj=0.50, (0 split)  
## AGE < 18.1 to the left, agree=0.833, adj=0.50, (0 split)  
## TAX < 291 to the right, agree=0.833, adj=0.50, (0 split)  
## INDUS < 9.095 to the right, agree=0.750, adj=0.25, (0 split)  
## DIS < 5.344 to the left, agree=0.750, adj=0.25, (0 split)  
##   
## Node number 5365: 5 observations  
## mean=24.28, MSE=0.1016   
##   
## Node number 10728: 4 observations  
## mean=23, MSE=0.235   
##   
## Node number 10729: 8 observations  
## mean=23.725, MSE=0.366875

printcp(reg\_fit\_train)

##   
## Regression tree:  
## rpart(formula = MEDV ~ CRIM + ZN + INDUS + CHAS + NOX + RM +   
## AGE + DIS + RAD + TAX + PTRATIO + LSTAT, data = Training\_Set,   
## method = "anova", minsplit = 10, cp = -1)  
##   
## Variables actually used in tree construction:  
## [1] AGE CRIM DIS INDUS LSTAT NOX PTRATIO RAD RM   
## [10] TAX   
##   
## Root node error: 26326/304 = 86.599  
##   
## n= 304   
##   
## CP nsplit rel error xerror xstd  
## 1 4.6941e-01 0 1.000000 1.01169 0.108379  
## 2 1.4625e-01 1 0.530591 0.70843 0.085512  
## 3 1.0036e-01 2 0.384341 0.56280 0.070724  
## 4 6.9125e-02 3 0.283984 0.49364 0.064996  
## 5 4.1088e-02 4 0.214859 0.40147 0.060445  
## 6 2.1351e-02 5 0.173771 0.35272 0.057262  
## 7 2.1211e-02 6 0.152419 0.34789 0.058647  
## 8 1.1997e-02 7 0.131208 0.31412 0.057685  
## 9 6.7388e-03 8 0.119211 0.26690 0.048353  
## 10 6.5233e-03 9 0.112472 0.24629 0.047210  
## 11 4.3927e-03 10 0.105949 0.25580 0.049233  
## 12 3.6138e-03 11 0.101556 0.25572 0.049223  
## 13 3.2684e-03 13 0.094328 0.26367 0.049719  
## 14 3.2209e-03 14 0.091060 0.26208 0.049684  
## 15 3.1825e-03 15 0.087839 0.26907 0.050066  
## 16 2.4730e-03 17 0.081474 0.26064 0.049659  
## 17 2.3453e-03 18 0.079001 0.26476 0.049910  
## 18 2.0310e-03 19 0.076656 0.26506 0.050334  
## 19 1.8332e-03 20 0.074625 0.26178 0.050307  
## 20 1.7633e-03 21 0.072792 0.26350 0.050515  
## 21 1.5246e-03 22 0.071028 0.26210 0.050455  
## 22 1.4567e-03 24 0.067979 0.26197 0.050446  
## 23 1.3805e-03 25 0.066522 0.26154 0.050452  
## 24 1.3348e-03 26 0.065142 0.26124 0.050442  
## 25 1.2408e-03 27 0.063807 0.25315 0.046154  
## 26 1.1833e-03 28 0.062566 0.25289 0.046138  
## 27 1.1494e-03 30 0.060200 0.25269 0.046074  
## 28 1.1015e-03 31 0.059050 0.25334 0.046097  
## 29 1.0242e-03 33 0.056847 0.25444 0.046117  
## 30 9.8052e-04 34 0.055823 0.25625 0.046139  
## 31 8.3578e-04 35 0.054842 0.25685 0.046146  
## 32 7.4690e-04 36 0.054007 0.25509 0.045712  
## 33 7.2918e-04 37 0.053260 0.25566 0.045707  
## 34 7.1473e-04 39 0.051801 0.25566 0.045707  
## 35 6.6702e-04 40 0.051087 0.25600 0.045702  
## 36 6.6029e-04 41 0.050420 0.25577 0.045680  
## 37 5.2144e-04 42 0.049759 0.25559 0.045635  
## 38 4.6949e-04 43 0.049238 0.25556 0.045625  
## 39 4.6010e-04 44 0.048768 0.25564 0.045623  
## 40 4.5222e-04 45 0.048308 0.25550 0.045625  
## 41 4.0666e-04 46 0.047856 0.25390 0.045594  
## 42 3.2066e-04 47 0.047450 0.25487 0.045599  
## 43 3.1732e-04 49 0.046808 0.25494 0.045595  
## 44 2.8432e-04 50 0.046491 0.25446 0.045600  
## 45 1.1347e-04 51 0.046207 0.25211 0.045573  
## 46 9.2684e-05 52 0.046093 0.25226 0.045574  
## 47 8.5088e-05 53 0.046000 0.25226 0.045574  
## 48 5.3242e-05 54 0.045915 0.25241 0.045572  
## 49 -1.0000e+00 55 0.045862 0.25241 0.045572

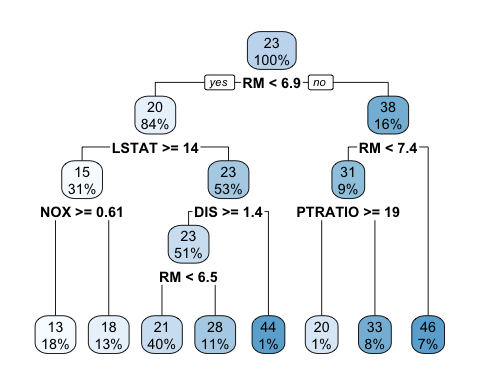
plotcp(reg\_fit\_train)

## Warning in sqrt(cp0 \* c(Inf, cp0[-length(cp0)])): NaNs produced



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

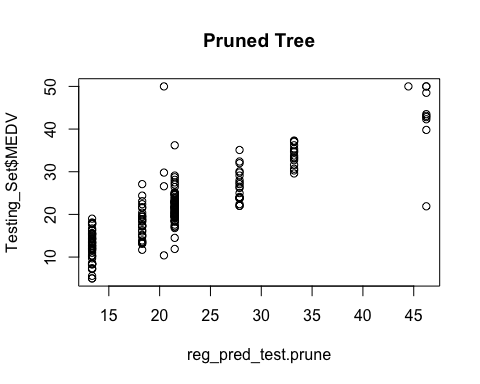
bestcp <-fit\_train$cptable[which.min(fit\_train$cptable[,"xerror"]),"CP"]  
reg\_pruned.tree <- prune(reg\_fit\_train, cp = bestcp)  
rpart.plot(reg\_pruned.tree)

 ###

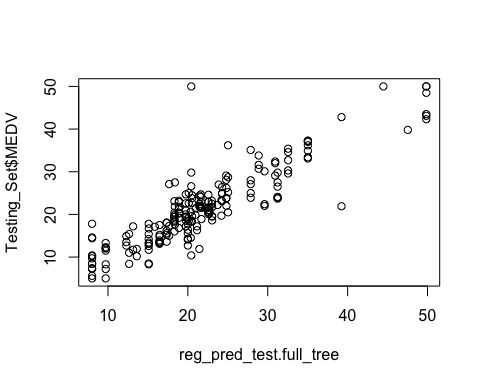
## B5.1 Predict on the Testing set with the pruned tree. Plot the predicted values vs the response values in the test set.

## B5.2 Predict on the Testing set with the Entire tree fitted to the training set. Plot the predicted values vs the response values in the test set.

### B5.1  
# Alternate specification   
reg\_pred\_test.prune = predict(reg\_pruned.tree, Testing\_Set)  
plot(reg\_pred\_test.prune,Testing\_Set$MEDV, main="Pruned Tree" )



### B5.2  
reg\_pred\_test.full\_tree=predict(reg\_fit\_train, Testing\_Set, main="Entire Tree on Trainig Set")  
plot(reg\_pred\_test.full\_tree,Testing\_Set$MEDV )

 ###

### B6 calculate the MSE for both the prediction and compare their MSE. Comment on your finding.

##  
mean((reg\_pred\_test.prune -Testing\_Set$MEDV )^2)

## [1] 20.98047

mean((reg\_pred\_test.full\_tree-Testing\_Set$MEDV )^2)

## [1] 18.55217