Individual Assignment 7

Zixiao Wu

2022-11-03

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com (http://rmarkdown.rstudio.com).

Exercise 8.4: Problem 8 (parts a, b, & c) (a)

```r
library(ISLR)
library(tree)
```

```
## Warning: 程辑包'tree'是用R版本4.2.2 来建造的
```

```r
set.seed(100)

num = sample(1:nrow(Carseats), nrow(Carseats)/2)
train = data.frame(Carseats[num,])
test = data.frame(Carseats[-num,])
```

b.

```r
tree.car = tree(Sales ~., data = train)
summary(tree.car)
```

```
##
## Regression tree:
## tree(formula = Sales ~ ., data = train)
## Variables actually used in tree construction:
## [1] "ShelveLoc"  "Price"       "Age"         "CompPrice"   "Advertising"
## [6] "Education"
## Number of terminal nodes:  17
## Residual mean deviance:  1.844 = 337.5 / 183
## Distribution of residuals:
##      Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -3.81700 -0.89580 -0.01857  0.00000  0.92000  2.91800
```

```r
plot(tree.car)
text(tree.car,pretty=0)
```
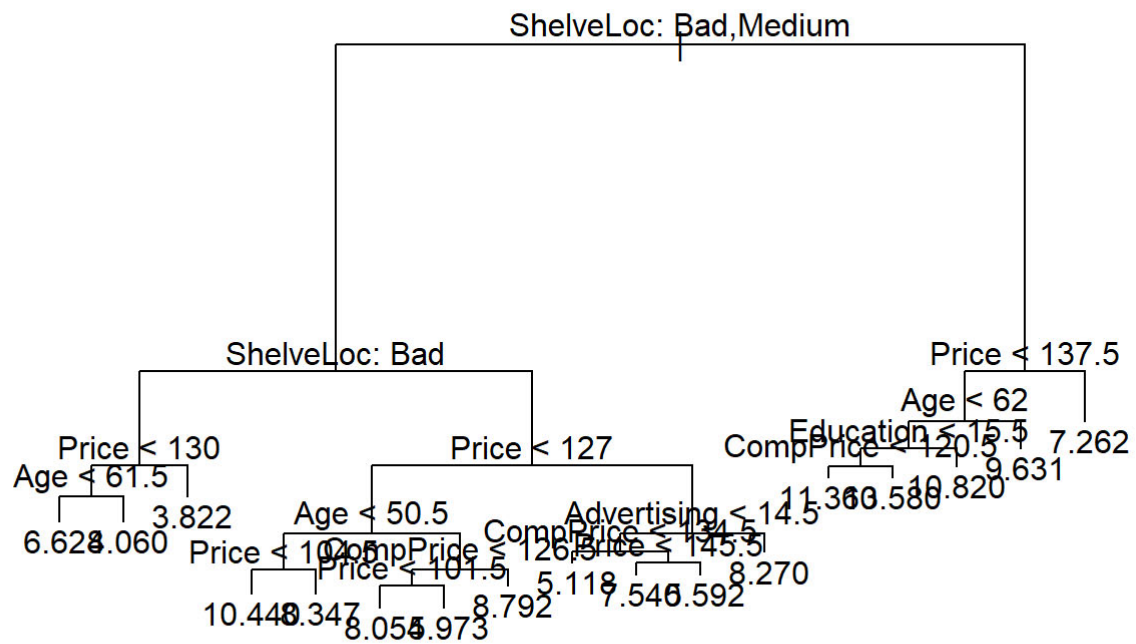
```
tree.pred = predict(tree.car, test)
test.mse = mean((tree.pred-test$Sales)^2)
test.mse
```

```
## [1] 5.395751
```

We can know that the test MSE = 5.3957

   c.

```
set.seed(100)
cv.car = cv.tree(tree.car)
cv.car
```

```
## $size
##  [1] 17 16 14 13 12 10  9  8  7  6  5  4  3  2  1
##
## $dev
##  [1]  911.0873  870.1659  870.1659  836.4578  844.3387  855.8490  862.0321
##  [8]  859.3948  877.2921  895.8563  905.0300  972.7206  981.5440 1070.2611
## [15] 1535.3160
##
## $k
##  [1]       -Inf  16.29507  16.51993  25.79705  27.26126  27.92693  35.37058
##  [8]  37.18435  40.02062  44.83473  52.50480  72.69219  98.74993 137.17898
## [15] 473.72357
##
## $method
## [1] "deviance"
##
## attr(,"class")
## [1] "prune"          "tree.sequence"
```

```
plot(cv.car$size, cv.car$dev, type = 'b')
```
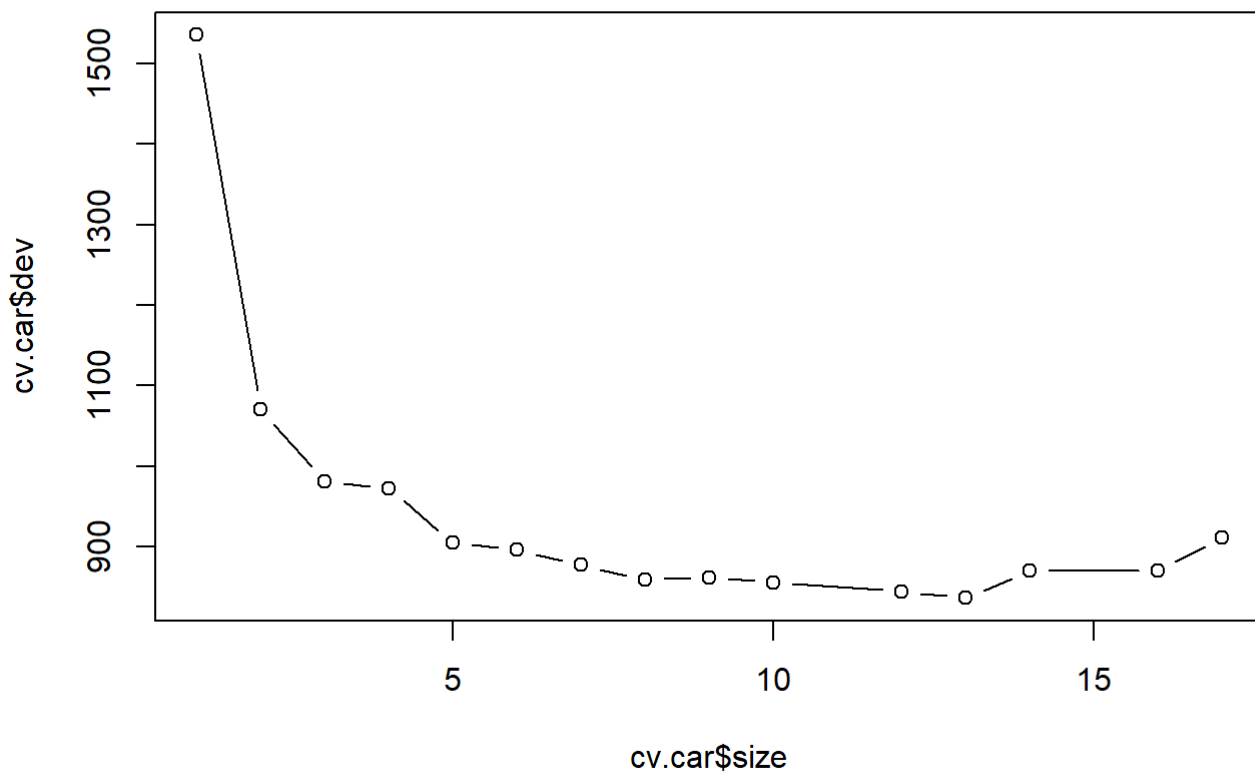

We can see that the best size is 13.

```
prune.car = prune.tree(tree.car, best = 13)
pred.prune = predict(prune.car, newdata = test)
pune_test.mse = mean((pred.prune-test$Sales)^2)
pune_test.mse
```

```
## [1] 5.39241
```

So, pruning the tree improve the test MSE from 5.3957 to 5.3924.