Individual Assignment 7

Charlie Ling

10/29/2021

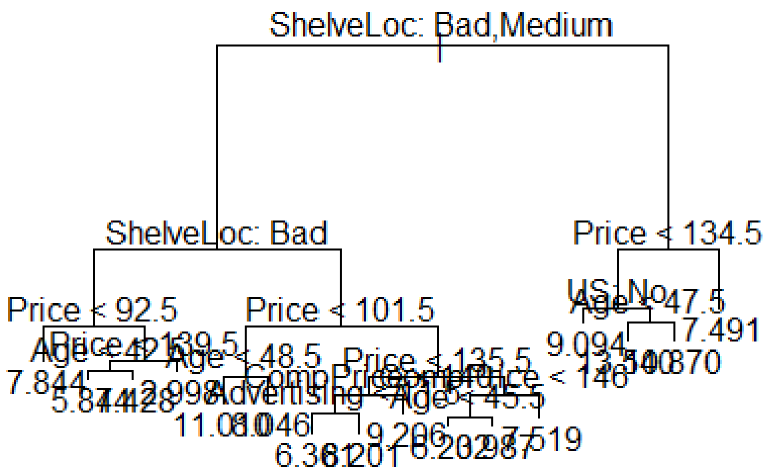
Problem #8: In the lab, a classification tree was applied to the Carseats data set after converting Sales into a qualitative response variable. Now we will seek to predict Sales using regression trees and related approaches, treating the response as a quantitative variable.

(a) Split the data set into a training set and a test set.

```
library(ISLR)
Carseats=na.omit(Carseats)
fix(Carseats)
set.seed(3)
train=sample(nrow(Carseats),nrow(Carseats)/2)
test=-train
```

(b) Fit a regression tree to the training set. Plot the tree, and interpret the results. What test MSE do you obtain?

```
library(tree)
tree.Carseats = tree(Sales~.,Carseats,subset=train)
plot(tree.Carseats)
text(tree.Carseats,pretty=0)
```




```
summary(tree.Carseats)

##
## Regression tree:
## tree(formula = Sales ~ ., data = Carseats, subset = train)
## Variables actually used in tree construction:
## [1] "ShelveLoc"   "Price"       "Age"         "CompPrice"   "Advertising"
## [6] "US"
## Number of terminal nodes:  16
## Residual mean deviance:  2.134 = 392.6 / 184
## Distribution of residuals:
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -4.37400 -0.90790 -0.05181  0.00000  0.92840  3.82600

tree.Carseats

## node), split, n, deviance, yval
##       * denotes terminal node
##
##  1) root 200 1507.000  7.338
##    2) ShelveLoc: Bad,Medium 161  861.900  6.653
##      4) ShelveLoc: Bad 50  232.300  5.236
##        8) Price < 92.5 8   44.980  7.844 *
##        9) Price > 92.5 42  122.600  4.740
##         18) Price < 139.5 37   89.680  4.975
##           36) Age < 42.5 14   24.690  5.874 *
##           37) Age > 42.5 23   46.790  4.428 *
```

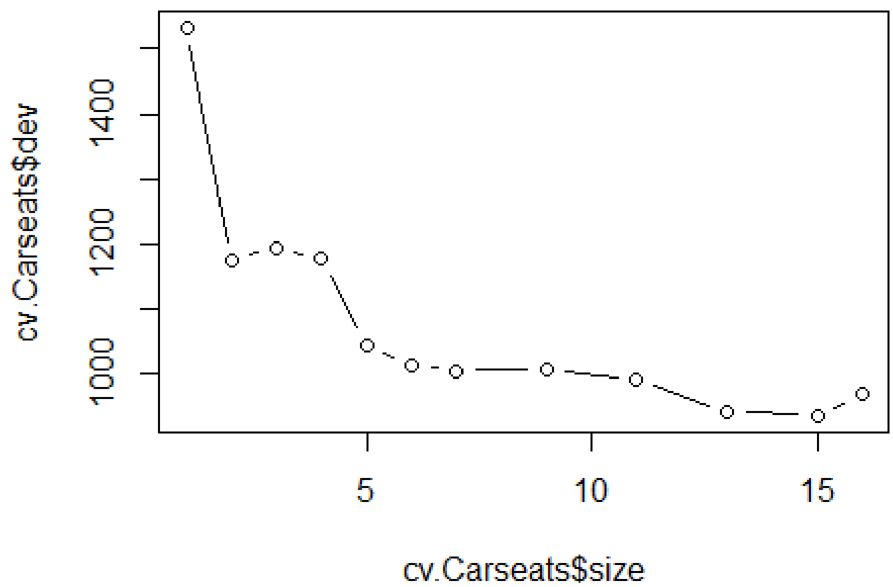
```
##         19) Price > 139.5 5   15.670  2.998 *
##       5) ShelveLoc: Medium 111  484.100  7.291
##        10) Price < 101.5 27  107.700  8.924
##          20) Age < 48.5 8   17.760 11.010 *
##          21) Age > 48.5 19   40.480  8.046 *
##        11) Price > 101.5 84  281.300  6.767
##          22) Price < 135.5 61  163.500  7.170
##            44) CompPrice < 140 54  128.600  6.906
##              88) Advertising < 11.5 38   62.230  6.361 *
##              89) Advertising > 11.5 16   28.230  8.201 *
##            45) CompPrice > 140 7    2.180  9.206 *
##          23) Price > 135.5 23   81.560  5.697
##            46) CompPrice < 146 15   26.820  4.725
##              92) Age < 45.5 5    2.980  6.202 *
##              93) Age > 45.5 10    7.489  3.987 *
##            47) CompPrice > 146 8   14.020  7.519 *
##     3) ShelveLoc: Good 39  257.000 10.170
##       6) Price < 134.5 28  120.100 11.220
##        12) US: No 5    5.763  9.094 *
##        13) US: Yes 23   86.850 11.680
##          26) Age < 47.5 7   20.020 13.540 *
##          27) Age > 47.5 16   32.230 10.870 *
##       7) Price > 134.5 11   27.070  7.491 *

tree.pred=predict(tree.Carseats,Carseats[test,])
mean((tree.pred-Carseats$Sales[test])^2)

## [1] 4.784151

#test MSE is 4.784151
```

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

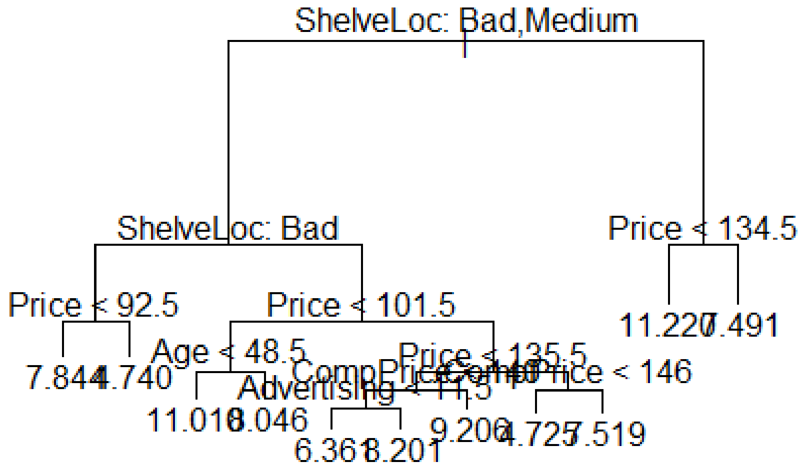
```
set.seed(3)
cv.Carseats=cv.tree(tree.Carseats)
names(cv.Carseats)

## [1] "size"   "dev"    "k"      "method"

cv.Carseats

## $size
##  [1] 16 15 13 11  9  7  6  5  4  3  2  1
##
## $dev
##  [1]  968.6673  936.3228  942.1916  992.6679 1008.3147 1005.6006 1015.2121
##  [8] 1045.9225 1177.7032 1195.3073 1174.9886 1532.9066
##
## $k
##  [1]      -Inf  16.35408  17.70200  31.04855  35.43549  38.47700  49.42328
##  [8]  64.75540  95.06820 109.78062 145.59777 387.80108
```

```
##
## $method
## [1] "deviance"
##
## attr(,"class")
## [1] "prune"         "tree.sequence"
```

```
plot(cv.Carseats$size,cv.Carseats$dev,type="b")
```




```
# optimal tree size is 10

prune.Carseats=prune.tree(tree.Carseats,best=10)
plot(prune.Carseats)
text(prune.Carseats,pretty=0)
```




```
tree.pred=predict(tree.Carseats,Carseats[test,])
mean((tree.pred-Carseats$Sales[test])^2)

## [1] 4.784151

#test MSE is 4.784151, which is the same as unpruned trees, so this pruning
the tree does not improve the test MSE
```