Assignment5

Meng Zhang

4/11/2019

```
carseat = read.csv("carseat.csv")
set.seed(1)
subset<-sample(nrow(carseat),nrow(carseat)*0.8)
carseat.train<-carseat[subset,]
carseat.test<-carseat[-subset,]
```

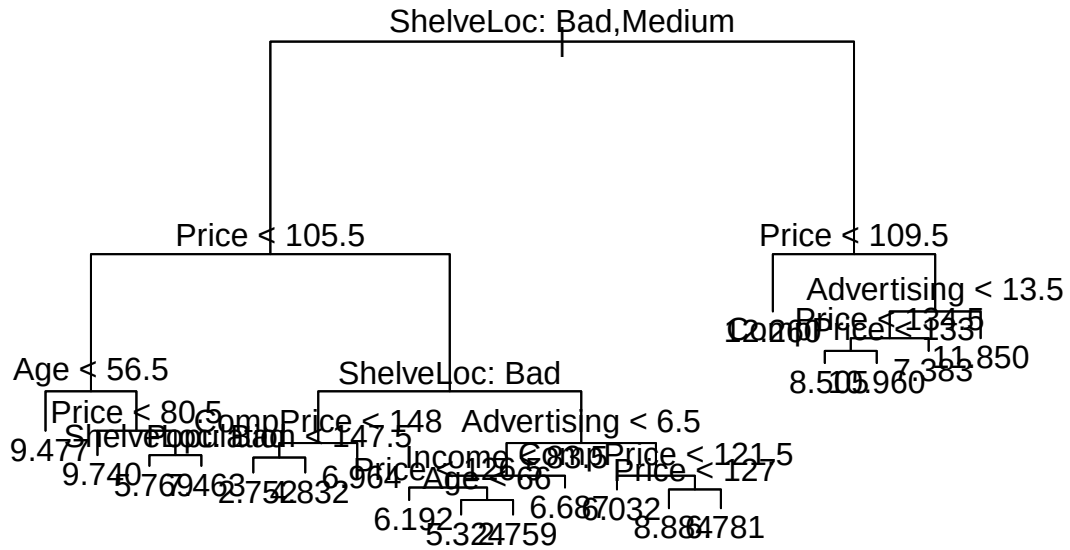
```
install.packages(pkgs='tree')
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/3.5'
## (as 'lib' is unspecified)
```

```
library(tree)
tree.carseat <- tree(Sales ~ ., data = carseat.train)
summary(tree.carseat)
```

```
##
## Regression tree:
## tree(formula = Sales ~ ., data = carseat.train)
## Variables actually used in tree construction:
## [1] "ShelveLoc"   "Price"       "Age"         "CompPrice"   "Population"
## [6] "Advertising" "Income"
## Number of terminal nodes:  19
## Residual mean deviance:  2.452 = 738.2 / 301
## Distribution of residuals:
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -3.80300 -0.97550 -0.06679  0.00000  0.95970  5.30800
```

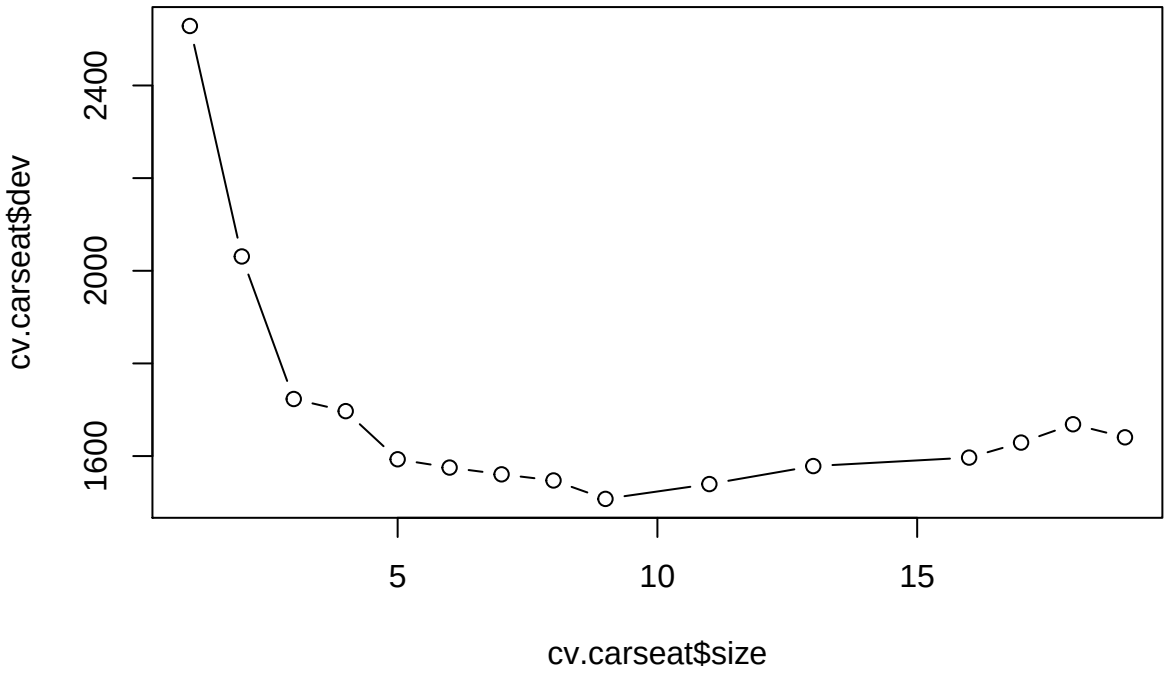
```
plot(tree.carseat)
text(tree.carseat, pretty = 0)
```



```
tree.prediction<-predict(tree.carseat,newdata=carseat.test)
tree.mse<-mean((carseat.test$Sales-tree.prediction)^2)
tree.mse
```

```
## [1] 4.817033
```

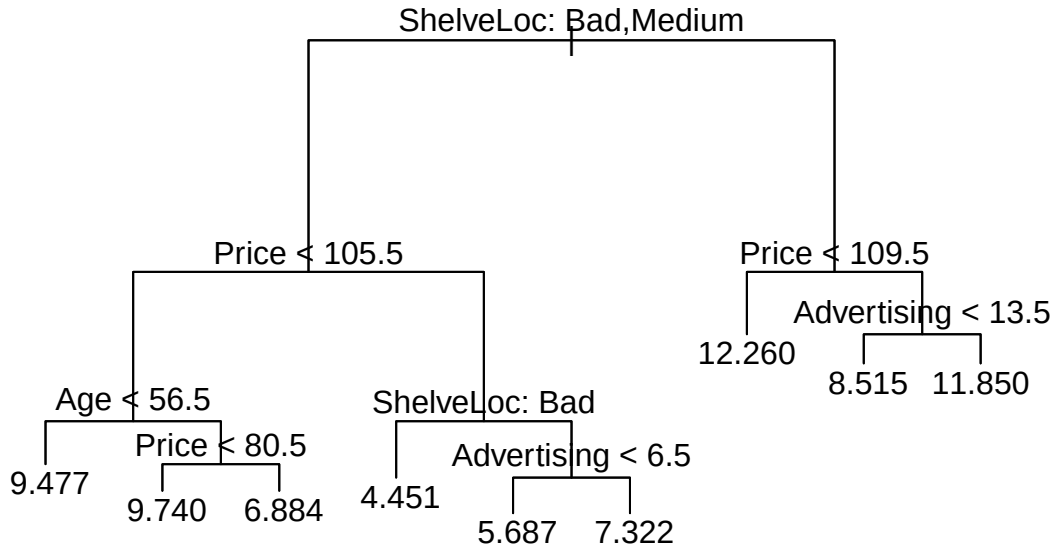
```
cv.carseat <- cv.tree(tree.carseat)
plot(cv.carseat$size, cv.carseat$dev, type = 'b')
```




```
cv.carseat$size[which.min(cv.carseat$dev)]
```

```
## [1] 9
```

```
prune.carseat <- prune.tree(tree.carseat, best = 9)
plot(prune.carseat)
text(prune.carseat, pretty = 0)
```



```
prune.prediction<-predict(prune.carseat,newdata=carseat.test)
prune.mse<-mean((carseat.test$Sales-prune.prediction)^2)
prune.mse
```

```
## [1] 4.831068
```

```
install.packages(pkgs='randomForest')
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/3.5'
## (as 'lib' is unspecified)
```

```
library(randomForest)
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
bag.carseat <- randomForest(Sales ~ ., data = carseat.train, mtry = 10, importance = TRUE)
bag.prediction <- predict(bag.carseat, newdata = carseat.test)
mean((bag.prediction - carseat.test$Sales)^2)
```

```
## [1] 2.230879
```

```
bag.carseat
```

```
##
## Call:
##  randomForest(formula = Sales ~ ., data = carseat.train, mtry = 10,      importance = TRUE)
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 10
##
##           Mean of squared residuals: 2.583164
##                     % Var explained: 67.06
```

```
library(randomForest)
rf.carseat <- randomForest(Sales ~ ., data = carseat.train, mtry = 3, importance = TRUE)
rf.prediction <- predict(rf.carseat, newdata = carseat.test)
mean((rf.prediction - carseat.test$Sales)^2)
```

```
## [1] 2.83626
```

```
rf.carseat
```

```
##
## Call:
##  randomForest(formula = Sales ~ ., data = carseat.train, mtry = 3,      importance = TRUE)
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 3
##
##           Mean of squared residuals: 3.090013
##                     % Var explained: 60.6
```

```
importance(rf.carseat)
```

```
##               %IncMSE IncNodePurity
## X            1.655016     165.29480
## CompPrice   14.997144     209.83480
## Income       7.376841     187.57658
```

```
## Advertising 17.334746     217.95165
## Population  -2.435090     140.98189
## Price       42.115125     564.00297
## ShelveLoc   45.281802     498.83345
## Age         15.354281     237.74921
## Education    1.636557      96.65062
## Urban       -2.532675      20.43989
## US           4.827442      38.06248
```

```
library(randomForest)
MSE_list <- sapply(1:10, function(i){
  randomF <- randomForest(Sales~., data = carseat.train, importance = T, mtry = i, ntree = 500)
  randomF_pred <- predict(randomF, carseat.test)
  mean((randomF_pred-carseat.test$Sales)^2)
})
minNbrTrees <- which.min(MSE_list)
minNbrTrees
```

```
## [1] 7
```

```
min(MSE_list)
```

```
## [1] 2.157084
```