Homework 3 | Group 5 | SCMT650

Group 5

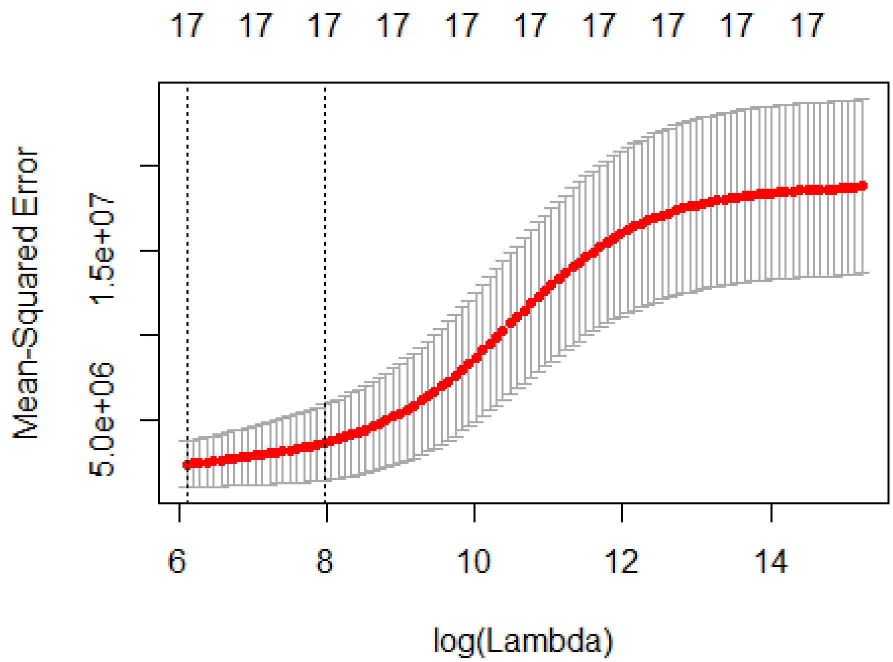
3/28/19

- **Question 1: Predicting with Ridge Regression**
- 1a. Create training and testing subsets

```
rm(list=ls())
setwd("E:/R-HW3")
library(ISLR)

names(College)

##  [1] "Private"     "Apps"        "Accept"      "Enroll"      "Top10perc"
##  [6] "Top25perc"   "F.Undergrad" "P.Undergrad" "Outstate"    "Room.Board"
## [11] "Books"       "Personal"    "PhD"         "Terminal"    "S.F.Ratio"
## [16] "perc.alumni" "Expend"      "Grad.Rate"

set.seed(1)
x=model.matrix(Apps~.,College)[,-1]
y=College$Apps
train=sample(1:nrow(x), nrow(x)/2)
y.test=y[-train]
```

- 1b. Determine the best value of lambda through cross-validation with the default k=10 number of folds and the default values of lambda used by the glmnet() function.
- Answer: Best value of lambda is 450.74 when k=10.

```
library(glmnet)

set.seed(1)
cv.out=cv.glmnet(x[train,],y[train],alpha=0)
plot(cv.out)
```




```
bestlam=cv.out$lambda.min
bestlam

## [1] 450.7435
```

- 1c. What is the test MSE associated with the best value of lambda?
- Answer: The test MSE associated with the value of lambda is 1,038,121.

```
ridge.mod=glmnet(x,y,alpha=0)
ridge.mod=glmnet(x[train,],y[train],alpha=0, thresh=1e-12)
ridge.pred=predict(ridge.mod,s=bestlam,newx=x[-train,])
mean((ridge.pred-y.test)^2)

## [1] 1038121
```

- 1d. Does your best value of lambda depend on the number of folds for cross-validation?
- Answer: From our observations, lambda doesn't change if the number of folds changes. With k=5, k=10, k=15 the lambda was 450.

```
set.seed(1)
cv.out2=cv.glmnet(x[train,],y[train],alpha=0, nfold=5)
lam5=cv.out2$lambda.min
lam5

## [1] 450.7435

set.seed(1)
cv.out3=cv.glmnet(x[train,],y[train],alpha=0, nfold=15)
```

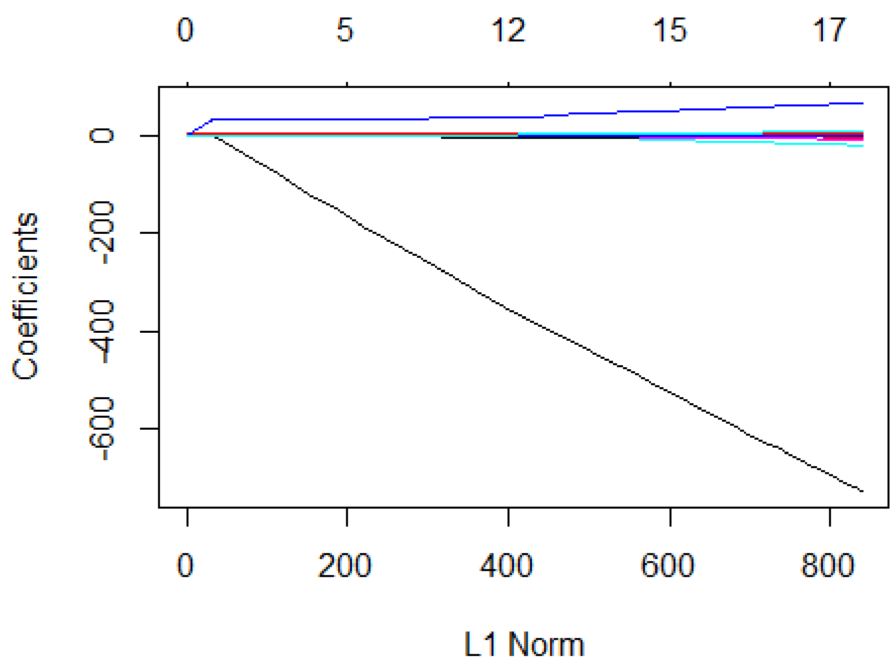
```
lam15=cv.out3$lambda.min
lam15

## [1] 450.7435
```

- 1e. What are the estimated coefficients for your best model?
- Answer: The estimated coefficients is shown below.

```
out=glmnet(x,y,alpha=0)
predict(out,type="coefficients",s=bestlam)[1:18,]

##   (Intercept)    PrivateYes        Accept        Enroll     Top10perc
## -1.575123e+03 -5.312141e+02  9.443508e-01  5.084882e-01  2.395085e+01
##     Top25perc   F.Undergrad   P.Undergrad      Outstate    Room.Board
##  1.676068e+00  8.195201e-02  2.519290e-02 -1.804998e-02  2.008313e-01
##         Books      Personal           PhD      Terminal     S.F.Ratio
##  1.443091e-01 -9.669190e-03 -3.428226e+00 -4.551334e+00  1.260404e+01
##   perc.alumni        Expend     Grad.Rate
## -9.173762e+00  7.444322e-02  1.149378e+01
```

- **Question 2: Predicting with The Lasso**
- 2a. Using the same data as in question 1, determine the best value of lambda using the default values for cross-validations and lambda used by the glmnet() function.
- Answer: The best value of lambda is 24.62.

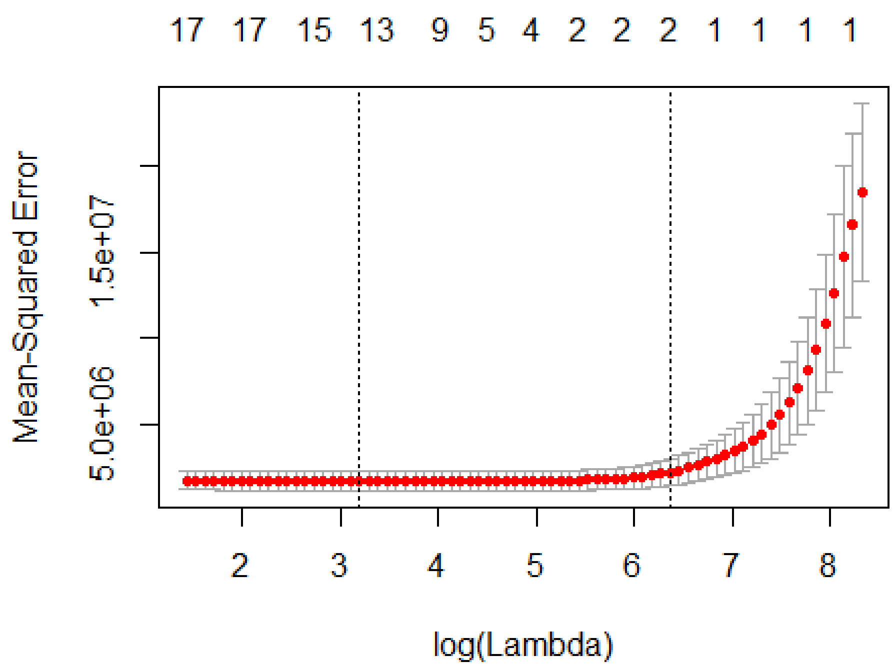
```
lasso.mod=glmnet(x[train,],y[train],alpha=1)
plot(lasso.mod)
```



```r
set.seed(1)
cv.out=cv.glmnet(x[train,],y[train],alpha=1)
bestlam=cv.out$lambda.min
bestlam

## [1] 24.62086
```

- 2b. Plot the results of your cross validation. What information does this provide to you?
- Answer: As Lambda increases (log of lambda increases), the test MSE remains constant through 6, then rapidly increases.

```r
plot(cv.out)
```



- 2c. What is the test MSE associated with the best value of lambda?
- Answer: The test MSE is 1,030,941.

```r
lasso.pred=predict(lasso.mod,s=bestlam,newx=x[-train,])
mean((lasso.pred-y.test)^2)

## [1] 1030941
```

- 2d. Which model is better: ridge or lasso. Explain why?
- Answer: Becuase Lasso has a smaller test error than ridge regression, it is a better method for estimation.

- 2e. Can you combine ridge regression with the lasso (as in elastic net) to arrive at a better predictive model than either ridge regression or the lasso? Evaluate for values of alpha in increments of 0.1 between 0 and 1.
- Answer: As the alpha increases from 0.0 -> 1.0 the test MSE trends generally up then down, but has a small increase fluctuation at alpha = 0.6.

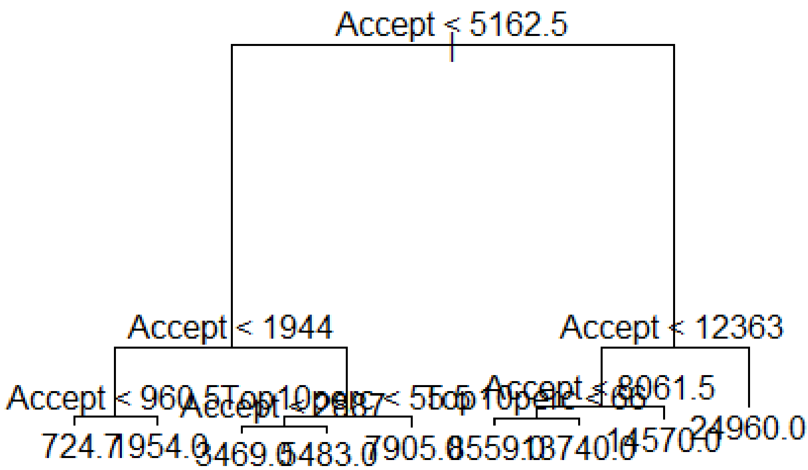
```
set.seed(1)
meanen=rep(0,11)

for (i in 0:11){
  cv.out=cv.glmnet(x[train,],y[train],alpha=((i-1)/10))
  bestlam=cv.out$lambda.min
  bestlam
  en.mod=glmnet(x[train,],y[train],alpha=((i-1)/10))
  en.mod=glmnet(x[train,],y[train],alpha=((i-1)/10), thresh=1e-12)
  en.pred=predict(en.mod,s=bestlam,newx=x[-train,])
  meanen[i]=mean((en.pred-y.test)^2)
}

meanen

##  [1] 1038121 1093089 1090771 1090223 1087914 1077503 1085943 1067693
##  [9] 1045695 1026307 1024606

# Comment: We use (i-1) in the equation because the loop starts at i = 1 even
we set it as (0:10).
```

- **Question 3: Predicting with Decision Trees**
- 3a. Create a regression tree using the training data with Apps as the response variable and all the predictors. Plot your tree.

```
rm(list=ls())
library(tree)

## Warning: package 'tree' was built under R version 3.5.3

set.seed(1)
train = sample(1:nrow(College), nrow(College)/2)
tree.college=tree(Apps~.,College,subset=train)
summary(tree.college)

##
## Regression tree:
## tree(formula = Apps ~ ., data = College, subset = train)
## Variables actually used in tree construction:
## [1] "Accept"    "Top10perc"
## Number of terminal nodes:  9
## Residual mean deviance:  2562000 = 9.71e+08 / 379
## Distribution of residuals:
```

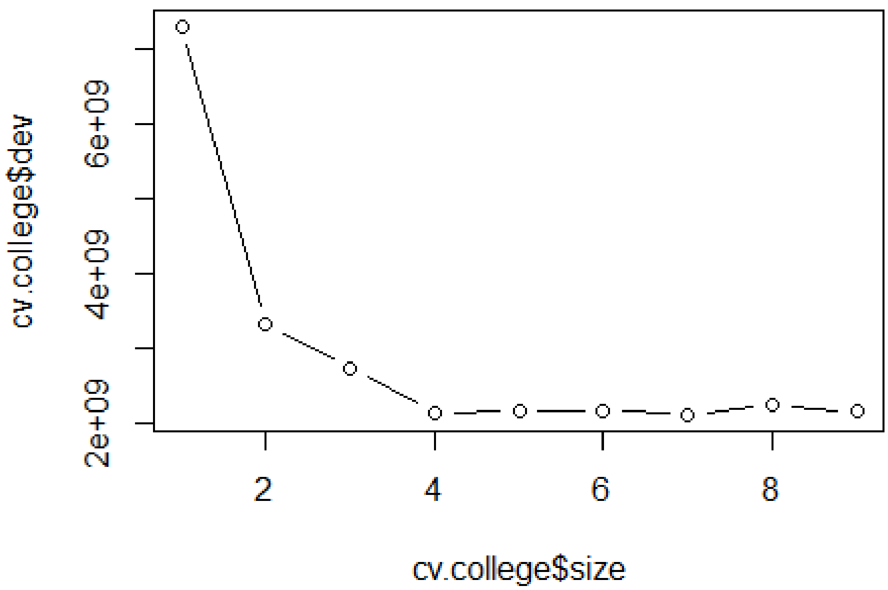
```
##      Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -8371.00  -401.60   -97.38     0.00   274.10 23140.00
```

```r
plot(tree.college)
text(tree.college,pretty=0)
```



- 3b. Prune your tree using cost complexity pruning. What size is the best tree?
- Answer: According to the plot and the deviation value, the best size should be 4. Beyond 4, the deviation value stays stable, with a slight decrease at 7, however the increase in model complexity at that size makes it less desirable choice.

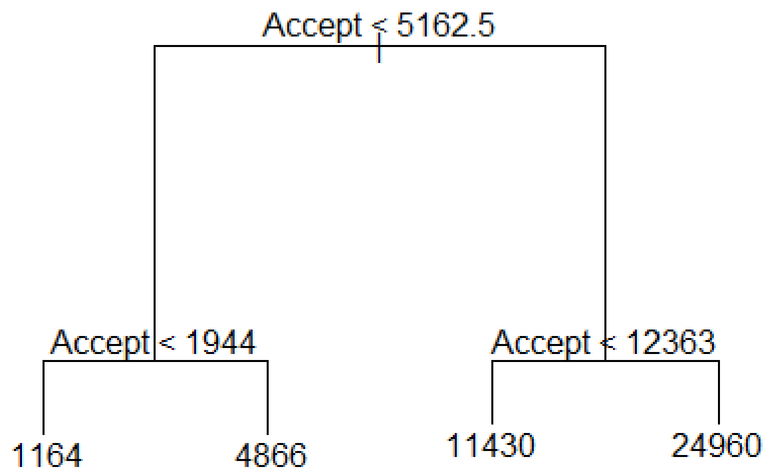
```r
cv.college=cv.tree(tree.college)
plot(cv.college$size,cv.college$dev,type='b')
```




```
#cv.college
```

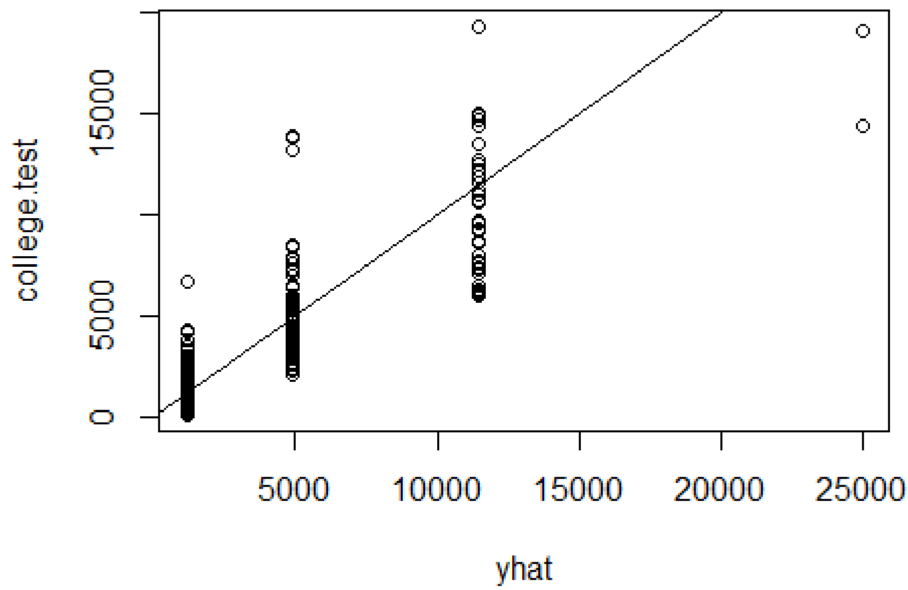
- 3c. Plot the tree corresponding to the size you identified in the prior question.

```
prune.college=prune.tree(tree.college,best=4)
plot(prune.college)
text(prune.college,pretty=0)
```



- 3d. Using the test data, what is the test MSE and/or RMSE? Is this a good predictive model?
- Answer: The test MSE is 3,101,409, and the RMSE is 1761. Since the test MSE is large, this is not a good predicitve model.

```
yhat=predict(prune.college,newdata=College[-train,])
college.test=College[-train,"Apps"]
plot(yhat,college.test)
abline(0,1)
```




```
mean((yhat-college.test)^2)

## [1] 3101409

sqrt(mean((yhat-college.test)^2))

## [1] 1761.082
```

- 3e. How does the accuracy of your best tree compare to ridge regression and the lasso?
- Answer: The test MSE of the regression tree is larger than the ridge regression and the lasso regression. The accuracy of the three models is shown below in increasing order of accuracy.

Best Tree MSE = 3,101,409 — RMSE = 1761

Ridge Regression MSE = 1,038,121 — RMSE = 1018

LASSO Regression MSE = 1,030,941 — RMSE = 1015

A smaller RMSE means better fitting model and high accuracy of predictions. Therefore, the LASSO method produces the most accurate model among the three.