Exercise 5.4

Section 23 - Group 6 Project Groups (Bosan Hsu, Fan Liu, Jimeng Yin, Michael Liu, Richard Wang, Zhuoqian Zhang

Exercise 6.8: Problem 9 (parts a, b, c, & d) Problem 9. In this exercise, we will predict the number of applications received using the other variables in the College data set.

```
library(ISLR)
attach(College)
View(College)
summary(College)
```

```
##  Private        Apps           Accept          Enroll       Top10per
c
##  No :212   Min.   :   81   Min.   :   72   Min.   :  35   Min.   : 1.
00
##  Yes:565   1st Qu.:  776   1st Qu.:  604   1st Qu.: 242   1st Qu.:15.
00
##            Median : 1558   Median : 1110   Median : 434   Median :23.
00
##            Mean   : 3002   Mean   : 2019   Mean   : 780   Mean   :27.
56
##            3rd Qu.: 3624   3rd Qu.: 2424   3rd Qu.: 902   3rd Qu.:35.
00
##            Max.   :48094   Max.   :26330   Max.   :6392   Max.   :96.
00
##    Top25perc      F.Undergrad     P.Undergrad         Outstate
##  Min.   :  9.0   Min.   :  139   Min.   :    1.0   Min.   : 2340
##  1st Qu.: 41.0   1st Qu.:  992   1st Qu.:   95.0   1st Qu.: 7320
##  Median : 54.0   Median : 1707   Median :  353.0   Median : 9990
##  Mean   : 55.8   Mean   : 3700   Mean   :  855.3   Mean   :10441
##  3rd Qu.: 69.0   3rd Qu.: 4005   3rd Qu.:  967.0   3rd Qu.:12925
##  Max.   :100.0   Max.   :31643   Max.   :21836.0   Max.   :21700
##    Room.Board       Books           Personal         PhD
##  Min.   :1780   Min.   :  96.0   Min.   : 250   Min.   :  8.00
##  1st Qu.:3597   1st Qu.: 470.0   1st Qu.: 850   1st Qu.: 62.00
##  Median :4200   Median : 500.0   Median :1200   Median : 75.00
##  Mean   :4358   Mean   : 549.4   Mean   :1341   Mean   : 72.66
##  3rd Qu.:5050   3rd Qu.: 600.0   3rd Qu.:1700   3rd Qu.: 85.00
##  Max.   :8124   Max.   :2340.0   Max.   :6800   Max.   :103.00
##     Terminal       S.F.Ratio      perc.alumni        Expend
##  Min.   : 24.0   Min.   : 2.50   Min.   : 0.00   Min.   : 3186
##  1st Qu.: 71.0   1st Qu.:11.50   1st Qu.:13.00   1st Qu.: 6751
##  Median : 82.0   Median :13.60   Median :21.00   Median : 8377
##  Mean   : 79.7   Mean   :14.09   Mean   :22.74   Mean   : 9660
##  3rd Qu.: 92.0   3rd Qu.:16.50   3rd Qu.:31.00   3rd Qu.:10830
##  Max.   :100.0   Max.   :39.80   Max.   :64.00   Max.   :56233
```

```
##    Grad.Rate
##  Min.   : 10.00
##  1st Qu.: 53.00
##  Median : 65.00
##  Mean   : 65.46
##  3rd Qu.: 78.00
##  Max.   :118.00
```

```
names(College)
```

```
##  [1] "Private"     "Apps"        "Accept"      "Enroll"      "Top10p
erc"
##  [6] "Top25perc"   "F.Undergrad" "P.Undergrad" "Outstate"    "Room.B
oard"
## [11] "Books"       "Personal"    "PhD"         "Terminal"    "S.F.Ra
tio"
## [16] "perc.alumni" "Expend"      "Grad.Rate"
```

```
dim(College)
```

```
## [1] 777  18
```

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

```
set.seed(1)
train <- sample(nrow(College), nrow(College)*0.8, replace = F)
train.data <- College[train, ]
test.data <- College[-train, ]
```

(b) Fit a linear model using least squares on the training set, and report the test error obtained.

```
lm.fit = lm(Apps~., data = train.data)
lm.pred = predict(lm.fit, test.data, type="response")
mean((lm.pred-test.data$Apps)^2)
```

```
## [1] 1567324
```

The test error is 1567324.

(c) Fit a ridge regression model on the training set, with λ chosen by cross-validation. Report the test error obtained.

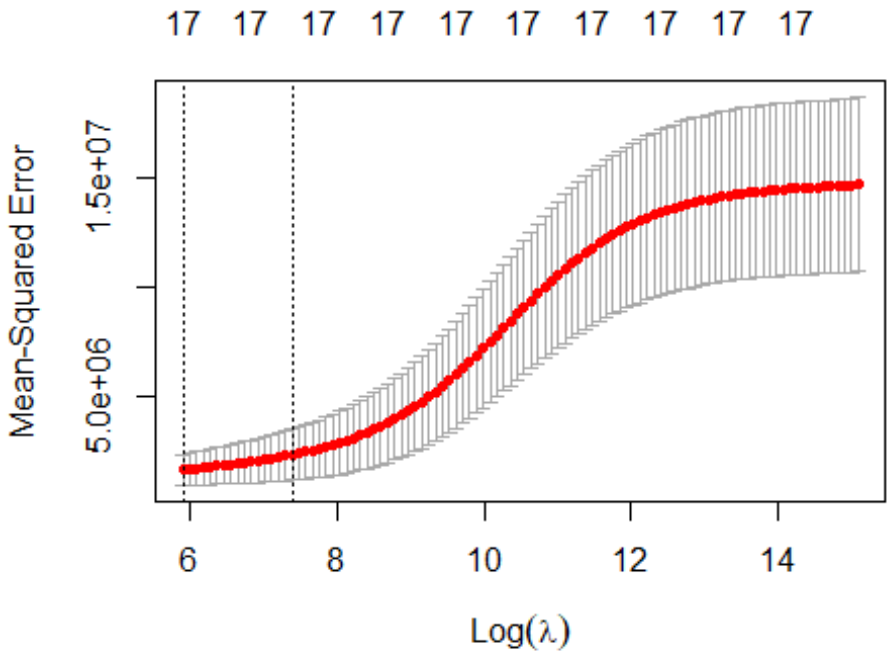
```
library(glmnet)
```

```
## 载入需要的程辑包：Matrix
```

```
## Loaded glmnet 4.1-8
```

```
set.seed(1)
##Create mstrix
x=model.matrix(Apps~.,College)[,-1]
y=College$Apps
##Train model
```

```
grid=10^seq(10,-2,length=100)
ridge.mod=glmnet(x[train,],y[train],alpha=0,lambda=grid,thresh=1e-12)
##Cross Validation
set.seed(1)
cv.out = cv.glmnet(x[train,],y[train], alpha=0)
plot(cv.out)
```




```
bestlam = cv.out$lambda.min
##Make Prediction
ridge.pred = predict(ridge.mod, s= bestlam, x=x[train,],y=y[train],newx
 = x[-train,],exact=T)
mean((ridge.pred-test.data$Apps)^2)

## [1] 1442059
```

(d) Fit a lasso model on the training set, with λ chosen by cross-validation. Report the test error obtained, along with the number of non-zero coefficient estimates.

```
set.seed(1)
lasso.mod=glmnet(x[train,],y[train],alpha=1,lambda=grid)
cv.out=cv.glmnet(x[train,],y[train],alpha=1)
bestlam=cv.out$lambda.min
coef=predict(lasso.mod,type="coefficients",s=bestlam)[-1,]
length(coef[coef!=0])

## [1] 16
```