Individual Assignment 6

Zixiao Wu 515491

2022-10-27

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 6.8: Problem 8 (parts e & f)

```
set.seed(114514)
X = c(rnorm(100))
N = c(rnorm(100))
β0=114
β1=514
β2=1919
β3=810
Y = c(β0 + β1*X + β2*X^2 + β3*X^3 + N)
df=data.frame(X,X^2,X^3,X^4,X^5,X^6,X^7,X^8,X^9,X^10,Y)
```

#(e)

```
set.seed(114514)
library(glmnet)
```

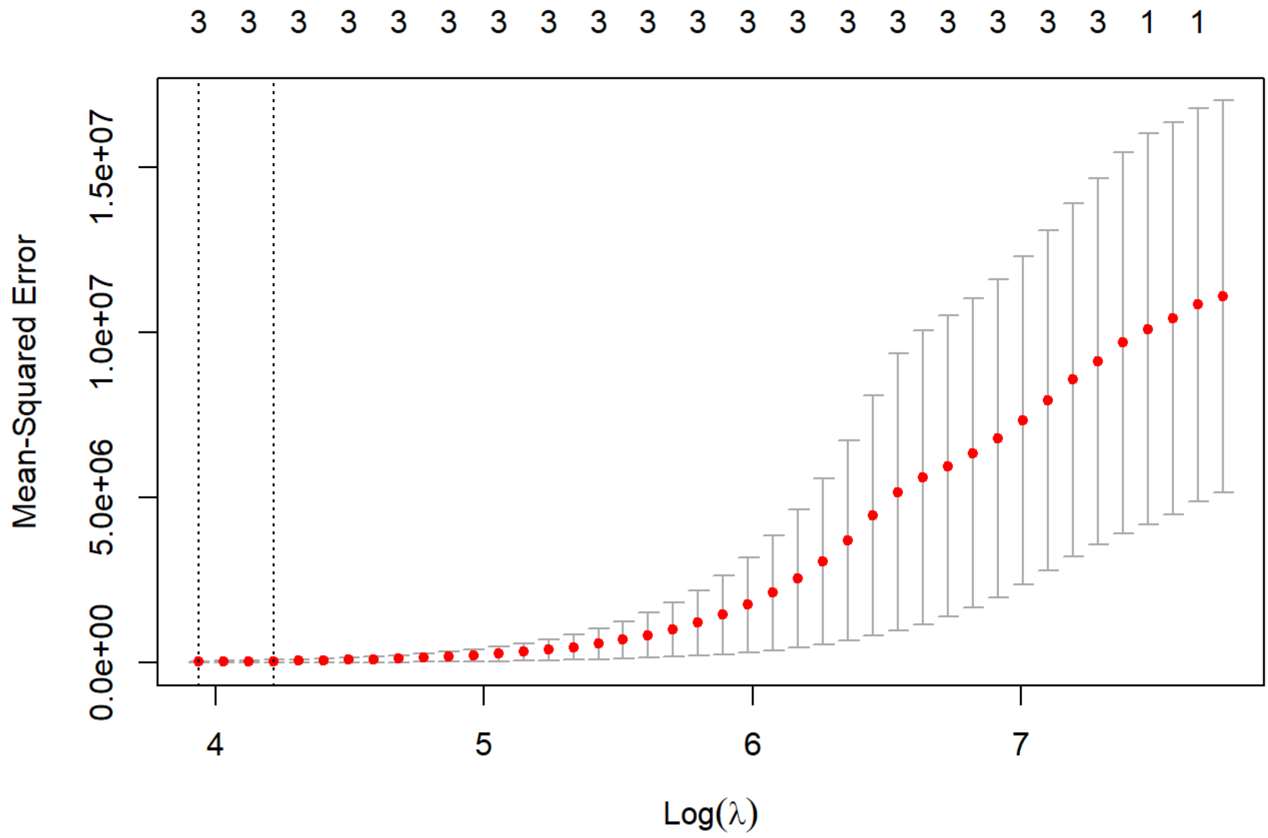
```
## 载入需要的程辑包：Matrix
```

```
## Loaded glmnet 4.1-4
```

```
x = model.matrix(Y~.,df)[,-1]
y = df$Y

train = sample(1:nrow(x), nrow(x)/2)
test = (-train)
y.test = y[test]

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




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

```
## [1] 51.30221
```

```
out = glmnet(x,y,alpha=1)
lasso.coef = predict(out, type="coefficients",s=bestlam)
lasso.coef
```

```
## 11 x 1 sparse Matrix of class "dgCMatrix"
##                      s1
## (Intercept)   180.3299
## X             516.2386
## X.2          1850.6751
## X.3           782.3113
## X.4             .
## X.5             .
## X.6             .
## X.7             .
## X.8             .
## X.9             .
## X.10            .
```

```
We can find that Lasso method creates a model with β0, β1, β2 and β3, and other coef are al
l 0. That means the model is clearly provides an accurate estimation of the real model.
```

#(f)

```
set.seed(114514)
β7 = 100
Y2 = c(β0 + β7*X^7 + N)
df2 = data.frame(X,X^2,X^3,X^4,X^5,X^6,X^7,X^8,X^9,X^10,Y2)
```

```
#best subset
library(leaps)
regfit.full2 = regsubsets(Y2~.,data=df2,nvmax=10)
reg.summary = summary(regfit.full2)
reg.summary
```

```
## Subset selection object
## Call: regsubsets.formula(Y2 ~ ., data = df2, nvmax = 10)
## 10 Variables  (and intercept)
##      Forced in Forced out
## X        FALSE      FALSE
## X.2      FALSE      FALSE
## X.3      FALSE      FALSE
## X.4      FALSE      FALSE
## X.5      FALSE      FALSE
## X.6      FALSE      FALSE
## X.7      FALSE      FALSE
## X.8      FALSE      FALSE
## X.9      FALSE      FALSE
## X.10     FALSE      FALSE
## 1 subsets of each size up to 10
## Selection Algorithm: exhaustive
##           X   X.2 X.3 X.4 X.5 X.6 X.7 X.8 X.9 X.10
## 1  ( 1 )  " " " " " " " " " " " " "*" " " " " " "
## 2  ( 1 )  "*" " " " " " " " " " " "*" " " " " " "
## 3  ( 1 )  " " " " "*" " " " " " " "*" " " " " " "
## 4  ( 1 )  " " " " "*" " " " " " " "*" " " "*" " "
## 5  ( 1 )  "*" "*" " " " " "*" " " "*" " " " " "*"
## 6  ( 1 )  "*" "*" "*" "*" "*" " " "*" " " " " " "
## 7  ( 1 )  "*" "*" "*" "*" "*" " " "*" " " "*" " "
## 8  ( 1 )  "*" "*" "*" "*" "*" "*" "*" "*" " " " "
## 9  ( 1 )  "*" "*" "*" "*" "*" "*" "*" "*" " " "*"
## 10  ( 1 ) "*" "*" "*" "*" "*" "*" "*" "*" "*" "*"
```

```
reg.summary$cp
```

```
##  [1] -5.0955806 -3.3273523 -1.9718588 -0.4736197  1.3837341  3.1393441
##  [7]  5.1329252  7.1141030  9.0834519 11.0000000
```

```
reg.summary$bic
```

```
##  [1] -1870.200 -1865.850 -1861.958 -1857.912 -1853.466 -1849.135 -1844.537
##  [8] -1839.953 -1835.382 -1830.871
```

```
reg.summary$adjr2
```

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

According to cp and bic, they are lowest at one variable model. So the model with with the x^7 term is the best choice.

```
#Lasso
x2 = model.matrix(Y2~., df2)[,-1]
y2 = df2$Y2

train2 = sample(1:nrow(x2), nrow(x2)/2)

cv.out = cv.glmnet(x[train,], y[train], alpha=1)
bestlam2 = cv.out$lambda.min

out2 = glmnet(x2, y2, alpha=1)
lasso.coef2 = predict(out2, type="coefficients", s=bestlam2)
lasso.coef2
```

```
## 11 x 1 sparse Matrix of class "dgCMatrix"
##                   s1
## (Intercept) 86.23706
## X            .
## X.2          .
## X.3          .
## X.4          .
## X.5          .
## X.6          .
## X.7         97.08370
## X.8          .
## X.9          .
## X.10         .
```

We can see Lasso model also results in a model with one variable X7. Also, the model is a good estimation of the real model.