

# MAST90083 Assignment 1

Haonan Zhong

## Question 1.1

```
library("MASS")
library("ISLR")
suppressMessages(library("glmnet"))
data(Hitters)
# Remove rows with NA in the Salary column
Hitters <- Hitters[!is.na(Hitters$Salary),]
```

## Question 1.2

```
# Construct design matrix and response variable
x <- model.matrix(~.-1, data = subset(Hitters, select = -(Salary)))
y <- Hitters$Salary
lambda <- 10^seq(10, -2, length = 100)

# Estimate ridge coefficients for 100 lambda values
coef_estimate <- glmnet(x, y, alpha = 0, lambda = lambda)
# Observe the coefficients for the largest lambda
coef(coef_estimate)[,1]
```

```
##      (Intercept)      AtBat      Hits      HmRun      Runs
## 5.359257e+02 5.443467e-08 1.974589e-07 7.956523e-07 3.339178e-07
##      RBI      Walks      Years      CAtBat      CHits
## 3.527222e-07 4.151323e-07 1.697711e-06 4.673743e-09 1.720071e-08
##      CHmRun      CRuns      CRBI      CWalks      LeagueA
## 1.297171e-07 3.450846e-08 3.561348e-08 3.767877e-08 5.800263e-07
##      LeagueN      DivisionW      PutOuts      Assists      Errors
## -5.800262e-07 -7.807263e-06 2.180288e-08 3.561198e-09 -1.660460e-08
##      NewLeagueN
## -1.152288e-07
```

```
#Observe the coefficients for the smallest lambda
coef(coef_estimate)[,100]
```

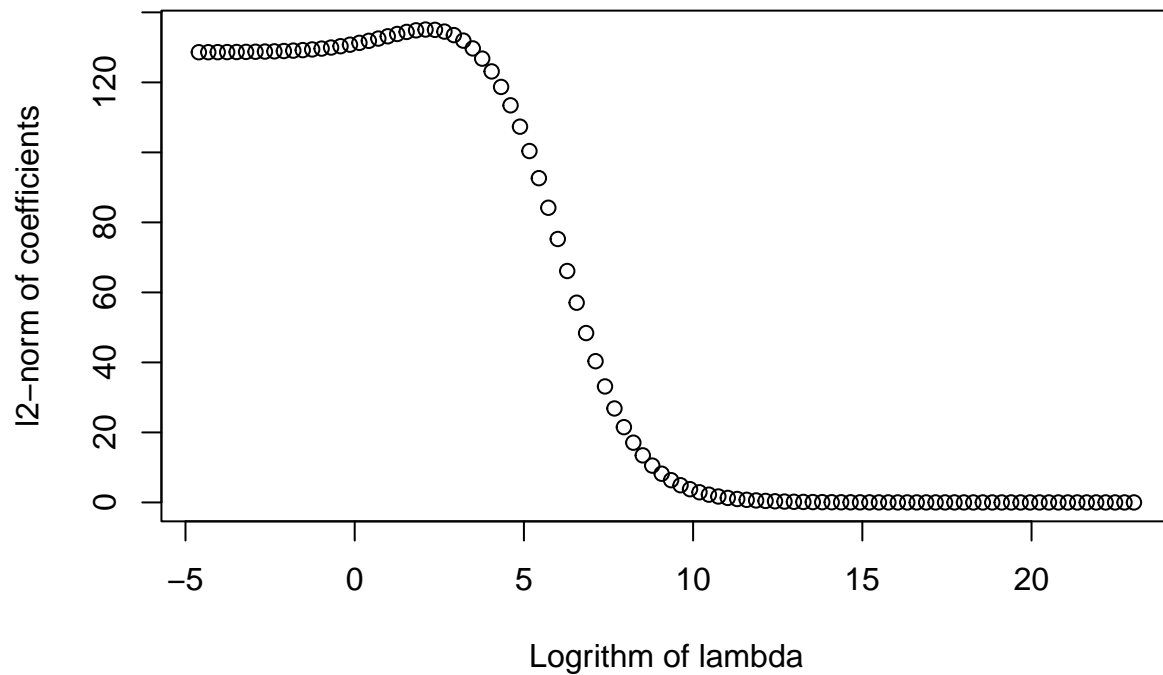
```
##      (Intercept)      AtBat      Hits      HmRun      Runs
## 195.91660826 -1.97384734 7.37766330 3.93654704 -2.19869563
##      RBI      Walks      Years      CAtBat      CHits
## -0.91621489 6.20035808 -3.71378518 -0.17510678 0.21135663
##      CHmRun      CRuns      CRBI      CWalks      LeagueA
```

```
##      0.05633262      1.36604022      0.70963872     -0.79581596    -31.80455226
##      LeagueN      DivisionW      PutOuts      Assists      Errors
##      31.60419354    -117.08236876      0.28202517      0.37318700     -3.42404313
##      NewLeagueN
##     -25.99406318
```

As we can see from the output above, coefficients for the largest lambda is much more closer to 0. Which is quite reasonable, as the effect of shrinkage penalty grows as  $\lambda$  increases, and the ridge regression coefficients will get closer to 0.

### Question 1.3

```
l2norm <- rep(0, length(lambda))
for (i in 1:100) {
  l2norm[i] <- norm(coef_estimate$beta[,i], type="2")
}
plot(log(lambda), l2norm, xlab = "Logrithm of lambda", ylab = "l2-norm of coefficients")
```



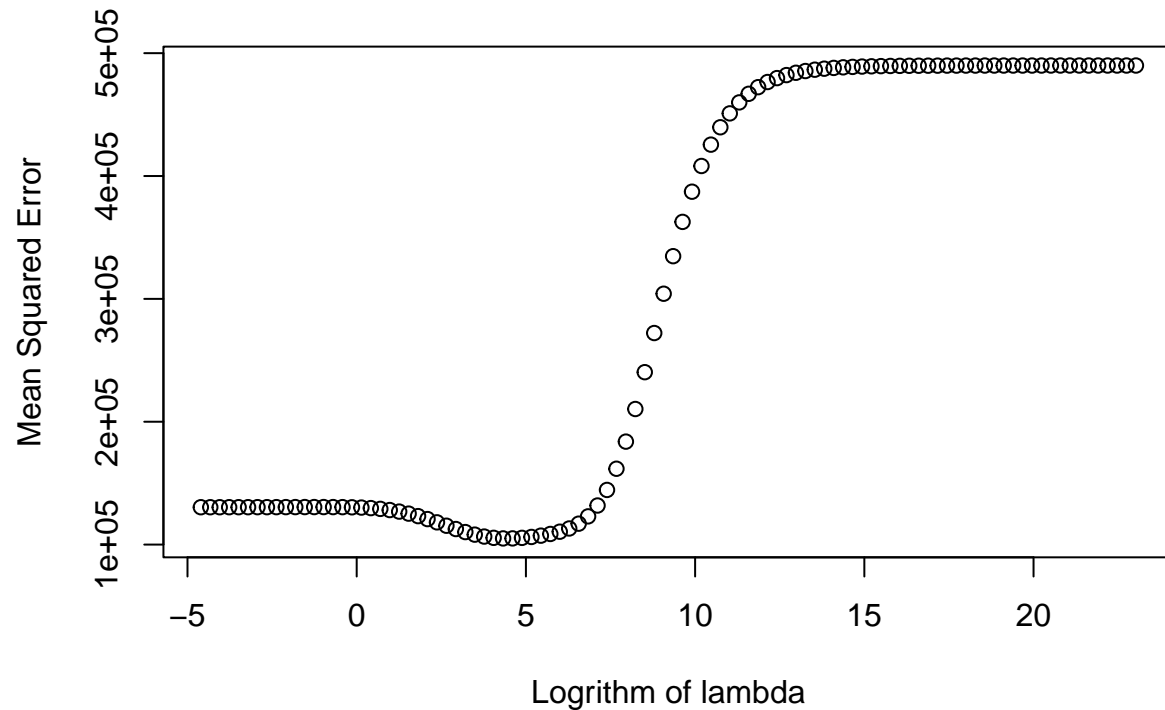
```
# Computing the MSE for each lambda
mse <- rep(0, length(lambda))
for (i in 1:length(lambda)) {
  prediction <- rep(0, length(y))
  for (j in 1:length(y)) {
```

```

    prediction[j] <- t(x[j,]) %*% matrix(coef_estimate$beta[, i])
  }
  mse[i] <- mean((y - prediction)^2)
}

plot(log(lambda), mse, xlab = "Logrithm of lambda", ylab = "Mean Squared Error")

```



We cannot really say anything about the optimal value of  $\lambda$  with  $l_2$ -norm, since it only tells us the size of the coefficients. On the other hand, mean squared error can tell us how accurate the coefficients are in terms of estimating the response, *Salary*.