Tut3_ST5227

Name:Zhu Xu
User ID:E0337988
Matriculation ID:A0191344H

13/Mar/2019

Question 1:

$$\min_x f(x) \quad s.t. \quad h_i(x), g_i(x) \leq 0$$

$$\Rightarrow \quad L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{i=1}^{n} \mu_i g_i(x)$$

For $g_i(\beta) = \beta^2 - t$

$$L = argmin\{\sum_{i=1}^{n}(y_i - \beta x_i)^2 + \mu(\beta^2 - t)\}$$

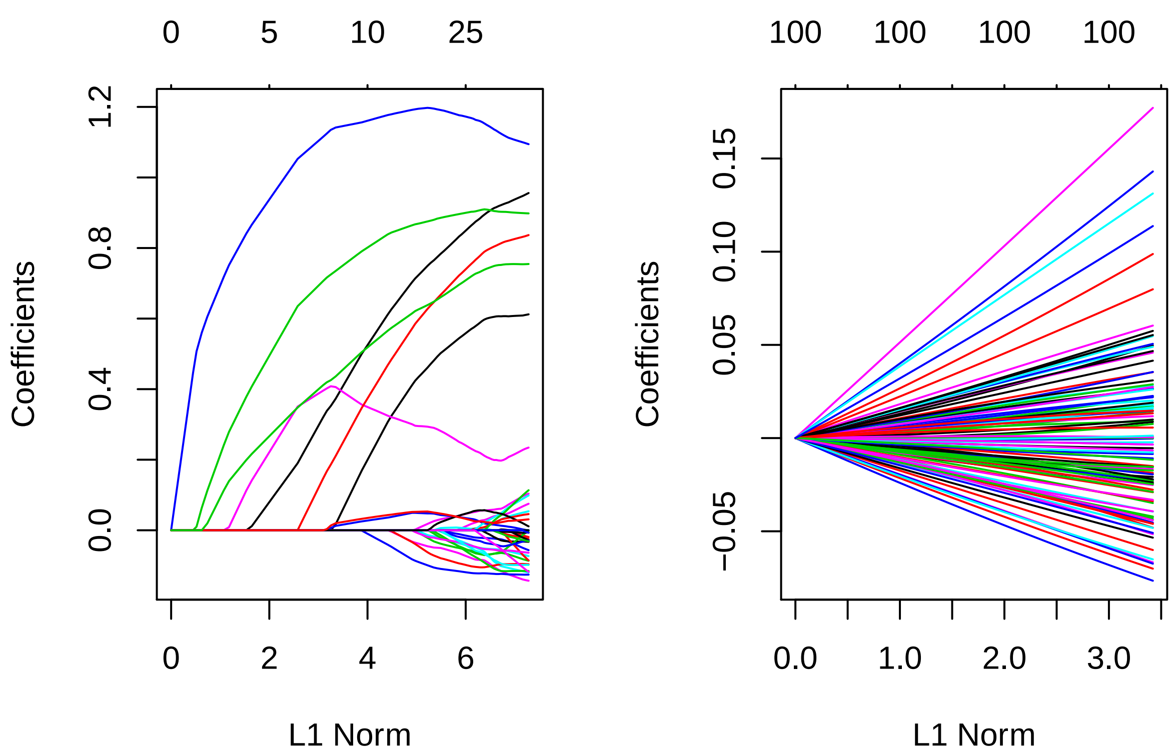
$$\mu\beta^2 - \mu t = \lambda\beta^2 \quad \Rightarrow \quad t = \frac{\lambda-\mu}{\mu}\beta^2$$

Question 2:

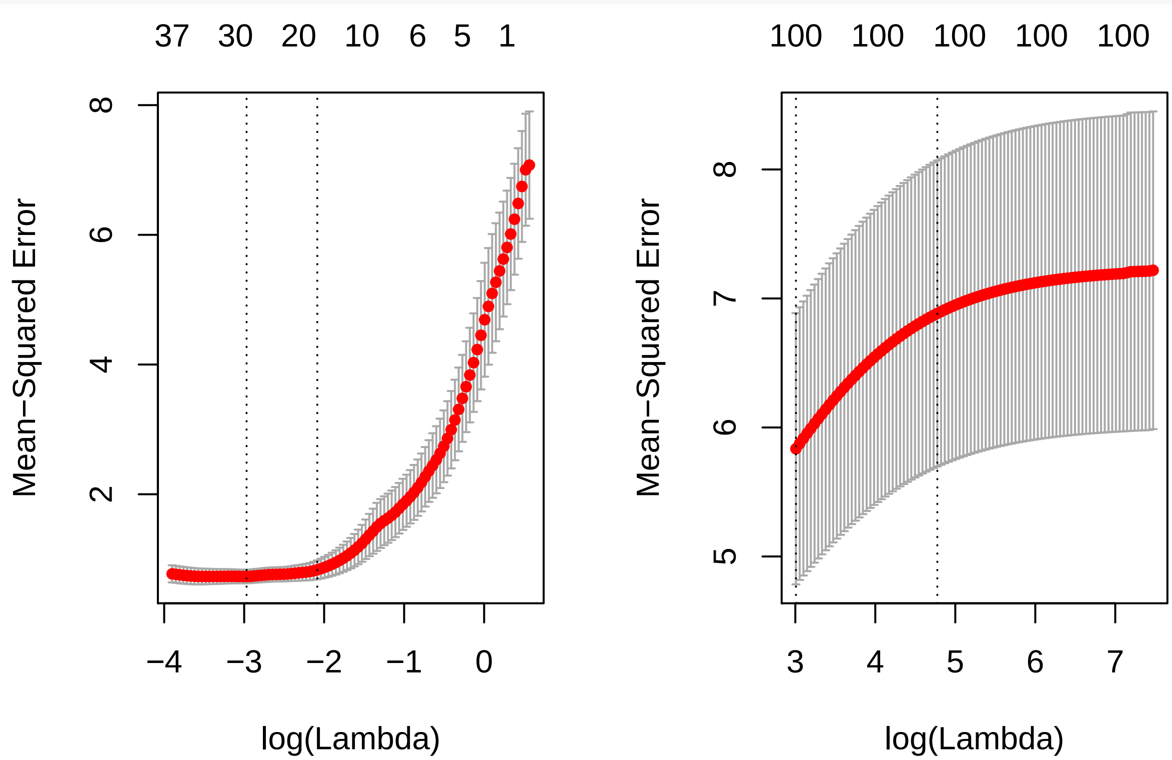
```r
library(glmnet)
```

```
## Loading required package: Matrix

## Loading required package: foreach

## Loaded glmnet 2.0-16
```

```r
mydata = read.csv(
  "/Users/xuzhu/Desktop/Notes/Sem2/ST5227-Applied_Data_Mining/Tut/T3/dataT03a1.csv")

X = data.matrix(mydata[,1:100])
y = data.matrix(mydata[,101])

reg_lasso = glmnet(X, y, alpha=1, standardize=TRUE)
reg_ridge = glmnet(X, y, alpha=0, standardize=TRUE)

par(mfrow = c(1, 2))

plot(reg_lasso)
plot(reg_ridge)
```




```r
cv_lasso = cv.glmnet(X, y, alpha=1, nfolds=5, standardize=TRUE)
cv_ridge = cv.glmnet(X, y, alpha=0, nfolds=5, standardize=TRUE)

plot(cv_lasso)
plot(cv_ridge)
```



```r
BestLambda_lasso = lambda=cv_lasso$lambda.min
BestLambda_ridge = lambda=cv_ridge$lambda.min


reg_lasso = glmnet(X, y, alpha=1, intercept=TRUE, lambda=BestLambda_lasso)
reg_ridge = glmnet(X, y, alpha=0, intercept=TRUE, lambda=BestLambda_ridge)

beta_lasso = reg_lasso$beta
beta_ridge = reg_ridge$beta

mydata_pred = read.csv(
  "/Users/xuzhu/Desktop/Notes/Sem2/ST5227-Applied_Data_Mining/Tut/T3/dataT03a2.csv")
X_pred = data.matrix(mydata_pred[,1:100])
y_pred = data.matrix(mydata_pred[,101])

mypred_lasso = predict(reg_lasso, X_pred)
mypred_ridge = predict(reg_ridge, X_pred)

error_lasso = mean((mypred_lasso-y_pred)^2)
error_ridge = mean((mypred_ridge-y_pred)^2)

error_lasso
```

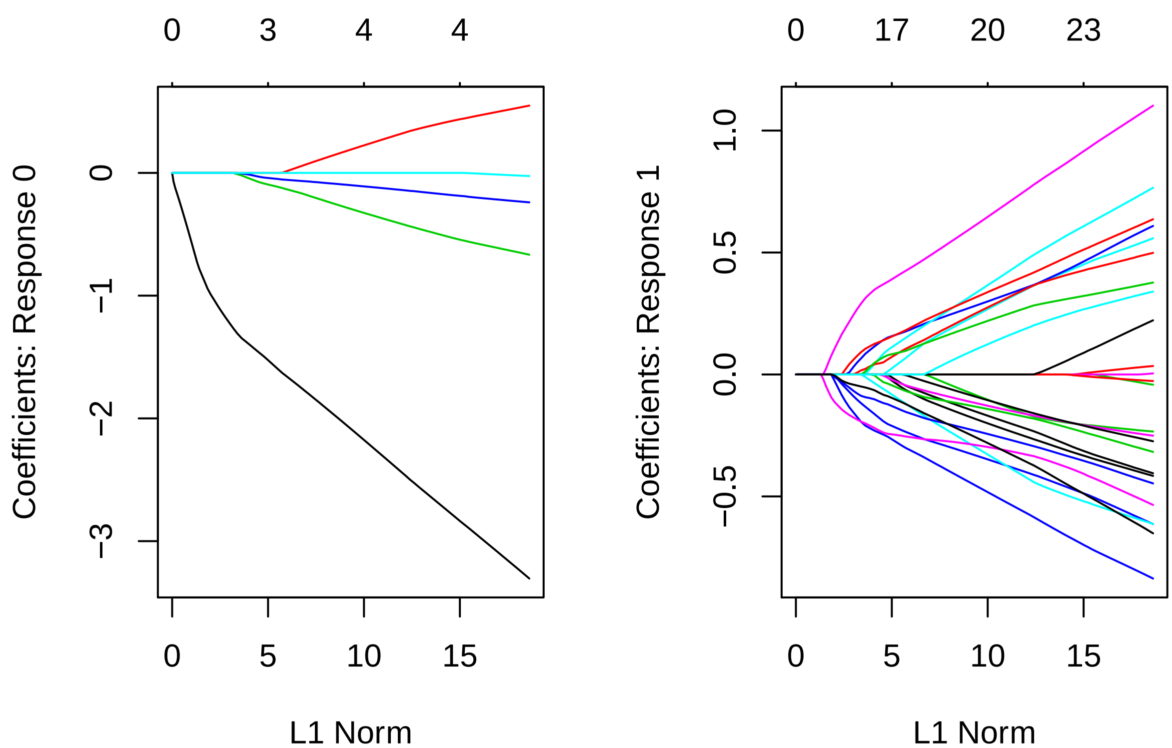
```
## [1] 0.1761772
```

```r
error_ridge
```

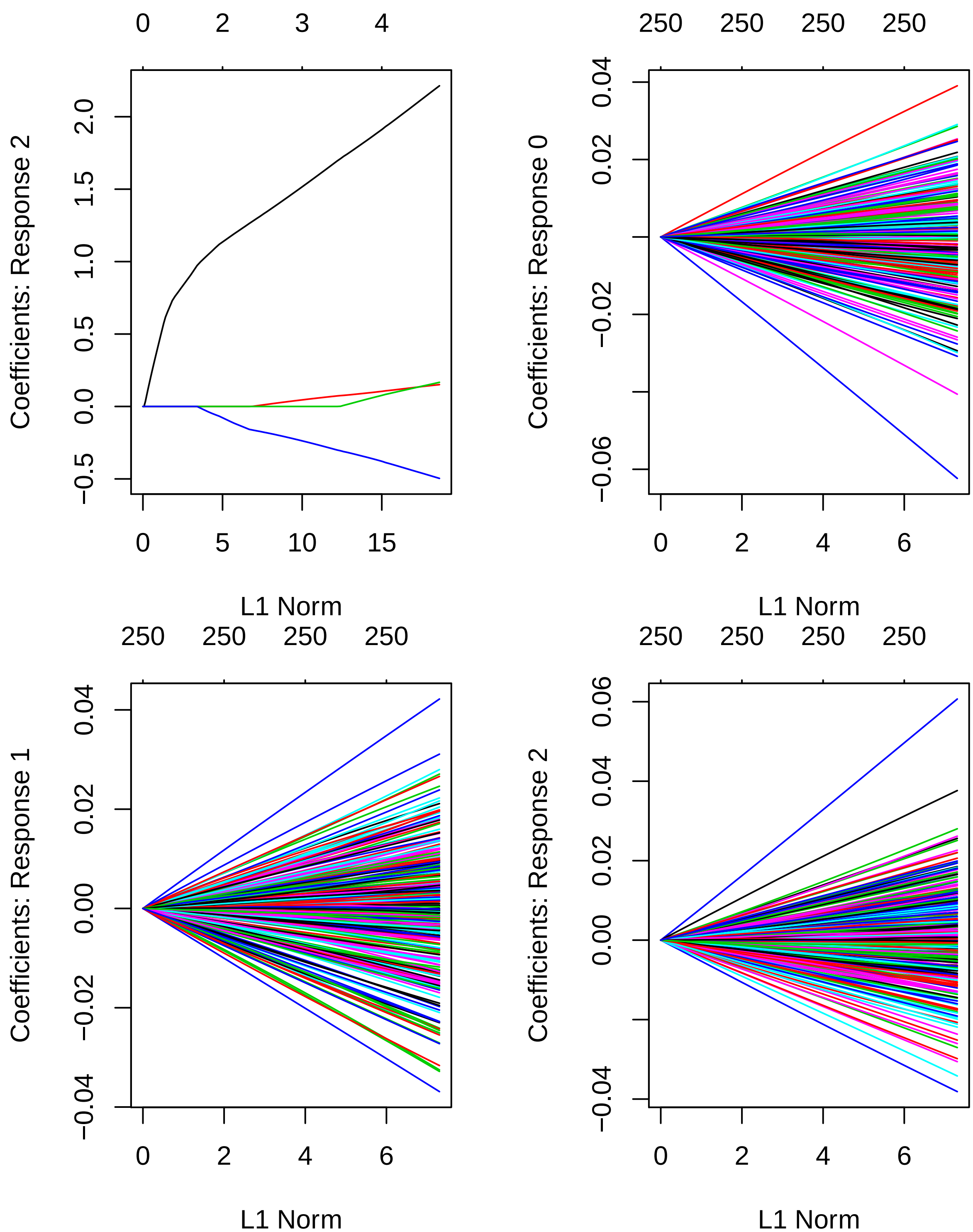
```
## [1] 0.9864321
```

Question 3:

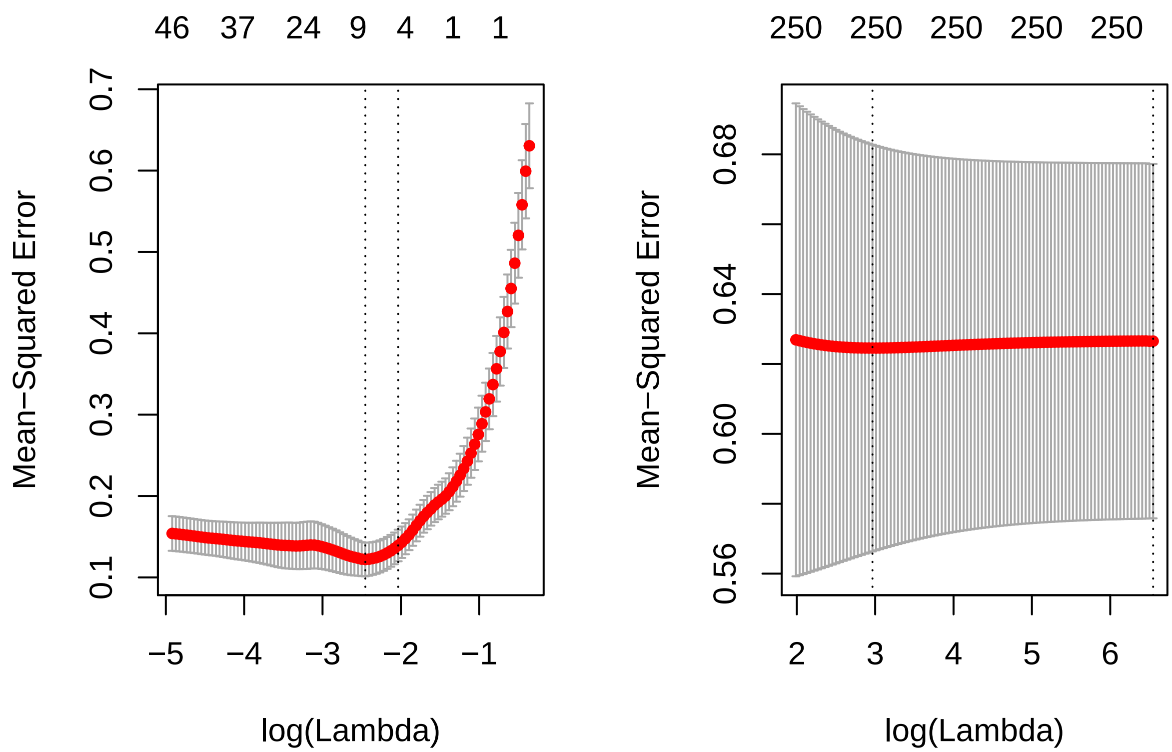
```
library(glmnet)

mydata = read.csv(
  "/Users/xuzhu/Desktop/Notes/Sem2/ST5227-Applied_Data_Mining/Tut/T3/dataT03b1.csv")

X = data.matrix(mydata[,1:250])
y = data.matrix(mydata[,251])

reg_lasso = glmnet(X, y, family="multinomial", alpha=1)
reg_ridge = glmnet(X, y, family="multinomial", alpha=0)

par(mfrow = c(1, 2))

plot(reg_lasso)
```




```
plot(reg_ridge)
```




```
cv_lasso = cv.glmnet(X, y, alpha=1, nfolds=5, standardize=TRUE)
cv_ridge = cv.glmnet(X, y, alpha=0, nfolds=5, standardize=TRUE)

plot(cv_lasso)
plot(cv_ridge)
```




```
BestLambda_lasso = lambda=cv_lasso$lambda.min

reg_lasso = glmnet(X, y, family="multinomial", alpha=1, intercept=TRUE, lambda=BestLambda_lasso)

beta_lasso = reg_lasso$beta

mydata_pred = read.csv(
  "/Users/xuzhu/Desktop/Notes/Sem2/ST5227-Applied_Data_Mining/Tut/T3/dataT03b2.csv")
X_pred = data.matrix(mydata_pred[,1:250])

mypred_lasso = predict(reg_lasso, X_pred, type="class")
mypred_lasso = as.numeric(mypred_lasso)
error_lasso = mean((mypred_lasso-y_pred)^2)
```

```
## Warning in mypred_lasso - y_pred: longer object length is not a multiple of
## shorter object length
```

```
error_lasso
```

```
## [1] 4.695274
```

Question 4:

Let X_j be the feature that we duplicate and let X_{-j} denote all other features except X_j. Let β_j denote the coefficient of X_j in the original Lasso problem, and let β_{-j} denote all the other coefficients. Then the original Lasso Problem be written as:

$$\min_{\beta} ||Y - X_{-j}\beta_{-j} - X_j\beta_j||_2^2 \quad s.t. \quad ||\tilde{\beta_{-j}}||_1 + |\tilde{\beta}_j| + |\beta_j^*| \leq t$$

Let X_j^* denote the duplicated feature and let $\tilde{\beta}_j$ and β_j^* denote the coefficients of the original feature X_j and the duplicated feature X_j^* in the new Lasso problem. Let $\tilde{\beta}_{-j}$ denote the coefficients of other feature vectors in the new Lasso problem. Then the updated Lasso problem can be written as:

$$\min_{\beta} ||Y - X_{-j}\tilde{\beta}_{-j} - X_j\tilde{\beta}_j||_2^2 \quad s.t. \quad ||\tilde{\beta_{-j}}||_1 + |\tilde{\beta}_j| + |\beta_j^*| \leq t$$

Now for a particular solution to the updated Lasso problem, our coefficients are:$\tilde{\beta}_{-j}$, $\tilde{\beta}_j$ and β_j^*. Now if we choose $\beta_{-j} = \tilde{\beta}_{-j}$, and $\beta_j = \tilde{\beta}_j + \beta_j^*$, then this set of β_{-j} and β_j is also a solution to the original Lasso problem.

Using Triangle Inequality($|a + b| \leq |a| + |b|$) and we get:

$$||\tilde{\beta}_{-j}||_1 + |\tilde{\beta}_j| + |\beta_j^*| \leq t \quad \Rightarrow \quad ||\tilde{\beta}_{-j}||_1 + |\tilde{\beta}_j + \beta_j^*| \leq t$$

For the given value of t, the optimal coefficient of X_j for the original Lasso problem is $\beta_j = a$. Therefore, this new coefficient $\tilde{\beta}_j + \beta_j^*$ also has to equal a. Further, we also know the absolute value of each indibidual coefficient can never exceed t.

Thus

$$\tilde{\beta}_j + \beta_j^* = a \quad s.t. \quad ||\tilde{\beta}_j|| \leq t, \quad |\beta_j^*| \leq t$$

Question 5:

Let X be the $n \times p$ feature matrix, and y the n-vector of labels.

Ridge regression solves

$$\min_{\beta} ||X\beta - y||^2 + \lambda ||\beta||^2$$

and has the general closed form solution

$$\beta = (X^T X + \lambda I)^{-1} X^T y$$

In this question, q=1 initially, and we can write

$$\beta = \frac{X^T y}{X^T X + \lambda}$$

With the duplication dimension, the feature matrix ix $[XX]_{n \times 2p}$.

Then

$$\min_{\beta_1, \beta_2} ||X\beta_1 + X\beta_2 - y||^2 + \lambda ||\beta_1||^2 + \lambda ||\beta_2||$$

Take the partial derivatives and set them to zero

$$2X^T(X\beta_1 + X\beta_2 - y) + 2\lambda\beta_1 = 0$$

$$2X^T(X\beta_1 + X\beta_2 - y) + 2\lambda\beta_2 = 0$$

Thus, in terms of the old β, we have

$$\beta_1 = \beta_2 = \frac{X^T y}{2X^T X + \lambda} = \frac{X^T X + \lambda}{2X^T X + \lambda}\beta$$