Homework 3 Linear Regression: Experiment

CSE 847: Machine Learning

Nan Cao

Question1:

$$f(\omega) = \frac{1}{2}||X\omega - y||_2^2 + \frac{\lambda}{2}||\omega||_2^2$$

$$f'(\omega) = X^T X - X^T y + \lambda\omega$$

$$Let f'(\omega) = 0$$

$$X^T X\omega - X^T y + \lambda\omega = 0$$

$$\hat{\omega}_{Ridge} = \frac{X^T y}{\lambda I + X^T X}$$

The following function is used to compute the $\hat{\omega}_{Ridge}$ with the given x and y.

```
function [ w ] = RidgeReg( x, y, Lamda )
        X2 = x' * x;
        [n1, n2] = size(X2);
        a = Lamda * (eye(n1, n2))+X2;
        if det(a) == 0
                display('singular matrix, unable to implement the ridge regression')
                ;
        else
                w = a^(-1)* (x' * y);
end
```

Question2: The following function is used to compute the MSE with the given $\hat{\omega}_{Ridge}$ x and y.

```
function [ MSE ] = MSE( w, x, y )
        a=x * w - y;
        N = length(y);
        MSE = (1/N) * ((norm(a,2))^2);
end
```

The follow codes are used to compute the MSEof different λ and plot them in the same figure.

```
LamdaError = zeros(3, 7);
LamdaError(1,1:7) = [1e-5, 1e-4, 1e-3, 1e
    -2, 1e-1, 1, 10];
for i=1 : 1 : 7
        w1=RidgeReg(x_train, y_train,
            LamdaError(1, i));
        LamdaError(2, i)=MSE(w1, x_train,
            y_train);
        LamdaError(3, i)=MSE(w1, x_test,
            y_test);
end
LogScale = log (LamdaError(1,1:7));
LamdaError
plot(LogScale, LamdaError(2,1:7),'k.',
    LogScale, LamdaError(3,1:7),'kp');
```

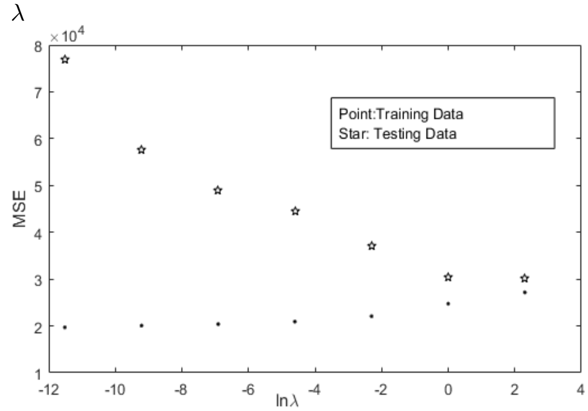
Following are the output results of both training and testing dataset

```
LamdaError =

1.0e+04 *

0.0000 0.0000 0.0000 0.0000 0.0000 0.0001 0.0010
1.9695 2.0060 2.0385 2.0919 2.2076 2.4731 2.7165
7.691  5.7532 4.9061 4.4405 3.6995 3.0518 3.0057
```

The plot of MSE of training and testing data under different λ



Question3: To perform 5-fold cross validation

```
Lamda(1:7) = [1e-5, 1e-4, 1e-3, 1e-2, 1e
    -1, 1, 10]
for i=1 : 1 : 7
        for k=1 : 1 : 5
                KFC_testX=x_train(2+48*(k
                    -1):1+48*k,1:64)
```

ACM ISBN N/A.
DOI: N/A

```
            KFC_testY=y_train(2+48*(k
                -1):1+48*k,1)
            KFC_trainX(1:194,1:64)=
                x_train([1:1+48*(k-1)
                ,2+48*k:242],1:64)
            KFC_trainY=y_train
                ([1:1+48*(k-1),2+48*k
                :242])
            KFC_w=RidgeReg(KFC_trainX,
                 KFC_trainY, Lamda(i))
            KFC_MSE(k)=MSE(KFC_w,
                KFC_testX, KFC_testY)
        end
        Lam_MSE(i)=mean(KFC_MSE)
end
Lam_MSE
```

```
Lam_MSE =

1.0e+05 *

1.2870 0.9394  0.6376 0.4533 0.3369 0.2889 0.2801
```

It shows that the best λ for estimated from the training data is 10