# Homework 3 Report

Peide Li

March 13, 2016

In this experiment, I used Matlab to solve regression problems with $l_2$-norm regularization. The original problem can be described in this way. Given training data set, we need to find a linear model which can minimize the error of prediction. In order to avoid over-fitting problems, we add regularization term to penalize the error if the norm of parameters is very large. Assume $X$ is the designed matrix, $y$ is the target vector in our training set, $\lambda$ is the regularization term. We need to solve the optimization problem:

$$E(w) = \underset{w}{min} \frac{1}{2} \left\| Xw - y \right\|_2^2 + \frac{\lambda}{2} \left\| w \right\|_2^2$$

By setting the derivative $\frac{\partial E(w)}{\partial w} = 0$, we can get the least square estimate of the parameter:

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

Then I use this formula to code the ridge regression solver in Matlab. The function is shown below. It needs three inputs: designed matrix, target vector, and regularization parameter. It will give out the $w$ which can minimize the loss function $E(w)$.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% Ridge Regression Function
%%% Input : X(predictors), Y(response), lambda(regularizaiotn parameter)
%%% Output: w (renewed parameters)

function [w] = RidgeRegress(X, Y, lambda)
[a, b] = size(X' * X);

%calculate the invers part
den = X' * X + eye(a,b) * lambda;

if det(den) == 0  %Prevent that the projection matrix is not inversible
    disp('The matrix is not inversible');
    w = 0;
else
    w = den^(-1) * X' * Y;
end

end
```

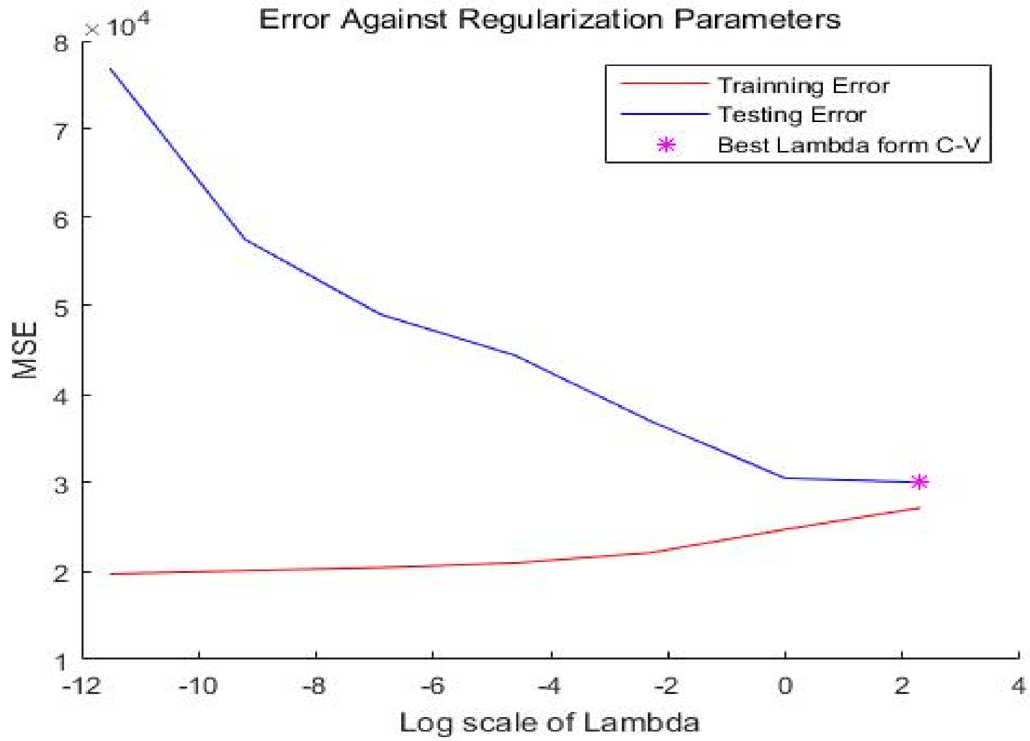


Figure 1: MSE against $log\lambda$

Next, I use the ridge regression solver to get the least square estimate of the parameters on the Diabetes dataset. The regularization parameter vary from $1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1, 10$. And I use mean squared error to measure the error. The Figure 1 shows the MSE against the value of the log $\lambda$.

Finally, I performed 5-fold cross validation on the training data to find the best $\lambda$ for the model. I divided the training set almost evenly into five groups, NO.1-50, 51-100, 101-150, 151-200, 201-242. So there are five different combinations of training set and cross validation set. For each value of $\lambda$, I use these sets to train the models and computed their cross validation errors(just the same as the MSE). I found the $\lambda = 10$ has the minimum cross validation error $1.4044 \times 10^5$. So I set $\lambda = 10$ and use the whole training set to train the model again. The best $\lambda$ obtained from cross validation procedure is also pointed out in the Figure 1.

The PDF report and the original Matlab code can be found at my github site: https://github.com/Kira233767/HomeWork-3.git