CS434 Assignment

Wenbo Hou, Zhi Jiang

June 14, 2017

1 Instruction

Please put our source file `Assignment_1.m` with data files in the same directory. This is a Matlab script, so it must be executed on Matlab. Type "Run" button, it will output all of results. Which include computed results of problem 2, 3, 4 and 6 on the command window as well as figures of problem 5, 6, 7 and 8.

2 Problem 1-3

The first part of this project requires us to calculate the weight vector,**W**, based on the training data. After that, we should use **W** to compute the sum of square error (**SSE**) for both the training data and the testing data to evaluate **W**. In this part, I will import the **dummy variable** that locates at the **at column of the new matrix**, as the requirement file states. The results are:

$$\mathbf{W} = \begin{bmatrix} -0.1011 \\ 0.0459 \\ -0.0027 \\ 3.0720 \\ -17.2254 \\ 3.7113 \\ 0.0072 \\ -1.5990 \\ 0.3736 \\ -0.0158 \\ -1.0242 \\ 0.0097 \\ -0.5860 \\ 39.5843 \end{bmatrix}$$

SSE of training data with above Weight Vector $= 9.561191 \times 10^3$
SSE of testing data with above Weight Vector $= 1.675231 \times 10^3$

3 Second Section: Problem 4

The problem 4 requires us to calculate **W** and **SSE** without **dummy variable**. The results are:

$$\mathbf{W} = \begin{bmatrix} -0.0979 \\ 0.0490 \\ -0.0254 \\ 3.4509 \\ -0.3555 \\ 5.8165 \\ -0.0033 \\ -1.0205 \\ 0.2266 \\ -0.0122 \\ -0.3880 \\ 0.0170 \\ -0.4850 \end{bmatrix}$$

SSE of training data with above Weight Vector $= 1.059806 \times 10^4$
SSE of testing data with above Weight Vector $= 1.797626 \times 10^3$
Comparing results generated in this section and the last section, I noticed that importing dummy variable makes SSE smaller. The hypothesis I made was that the dummy variable reduce the model's dependency of the given data set.

4 Problem 5

The problem 5 requires us to explore effects of additional random features to **W** and **SSE**. To get a comprehensive result, I decide to continuously add random features to the training data and calculate **W** and **SSE**. As the assignment description states, I will provide plots showing the relationship between the number of additional random features and SSE of the training data and the relationship between the number of additional random features and SSE of the testing data.

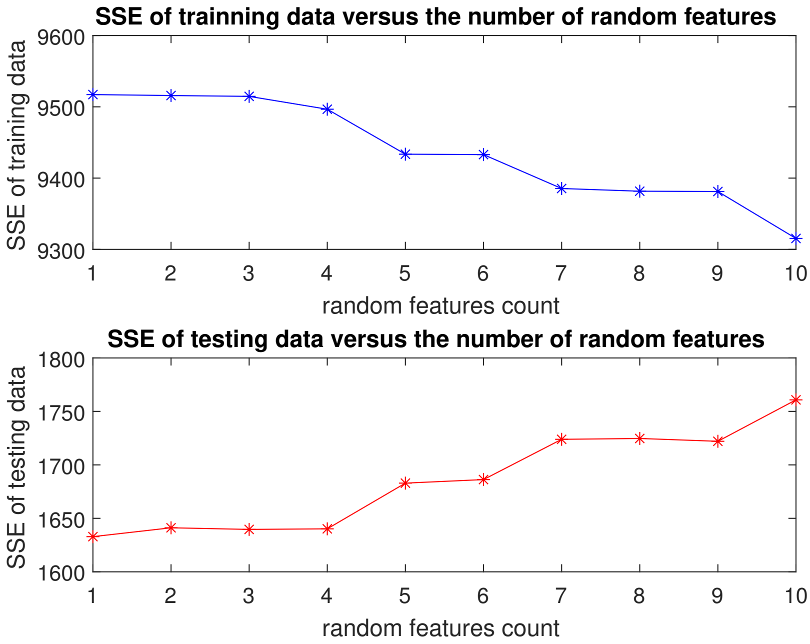

Figure 1: SSE versus the number of additional random features

By observing the above two plots, we can say that more additional random features lead to **smaller SSE of the training data** and **greater SSE of the testing data**. Specifically, the large amount of features in the training data force the model to fit the training data in a deep scale. Correspondingly, the model generated cannot fit the testing data well, which shows as a greater SSE. There exists a over-fitting because this model has a lot of features so it is excessively complex. So this is why we need to do regularization in next problems

5 Problem 6

In this section, we are required to import a new method to calculate **W** that is:

$$w = (X^T X + \lambda I)^{-1} X^T Y \tag{1}$$

To show the effect of λ to the model clearly, I create an array of λ whose size is 1000 and range is 0.0001-1000. I will provide two plots to show relationships between λ and SSE.The most optimal λ value should make **SSE of the testing data** and **SSE of the training data** as small as possible.
The most optimal λ is 2.001200×10^{-1}
The smallest SSE of testing data is 1.649620×10^3
The corresponding SSE of training data is 9.634307×10^3

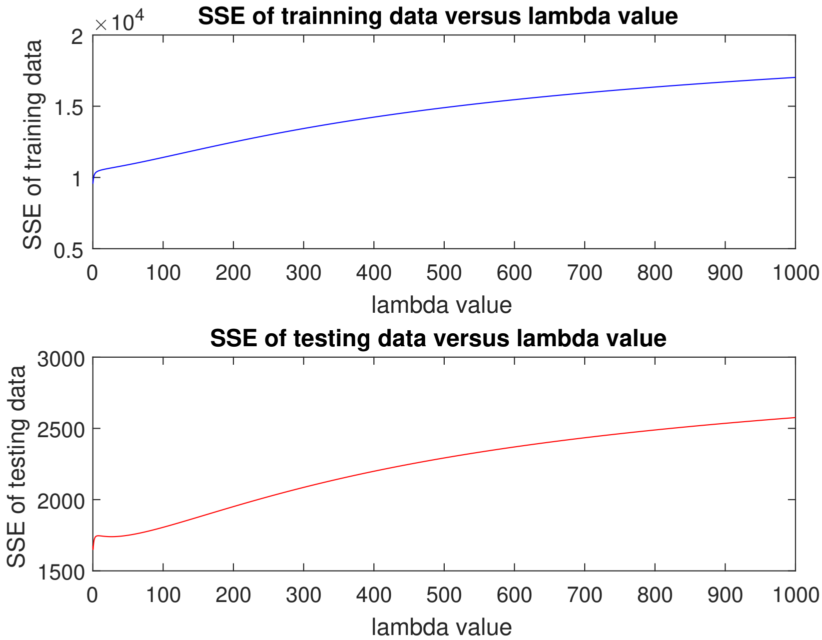

Figure 2: SSE versus λ value

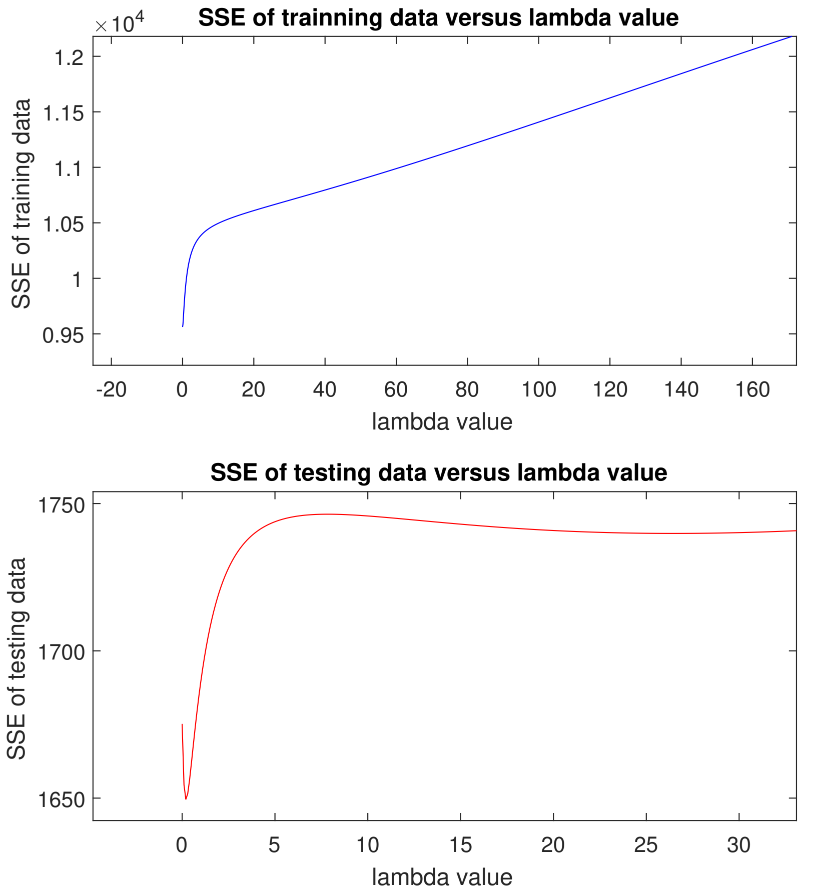

Figure 3: SSE versus λ value (zoomed in)

6 Problem 7

Problem 7 requires us to analyze how λ affects the Weight Vector. The model reflects that contributions of certain features to the model get insignificant when corresponding weight values approach to 0. The following figure shows that **some elements in W approaches to 0 when λ becomes greater**. In other words, the model becomes simpler when λ gets greater.

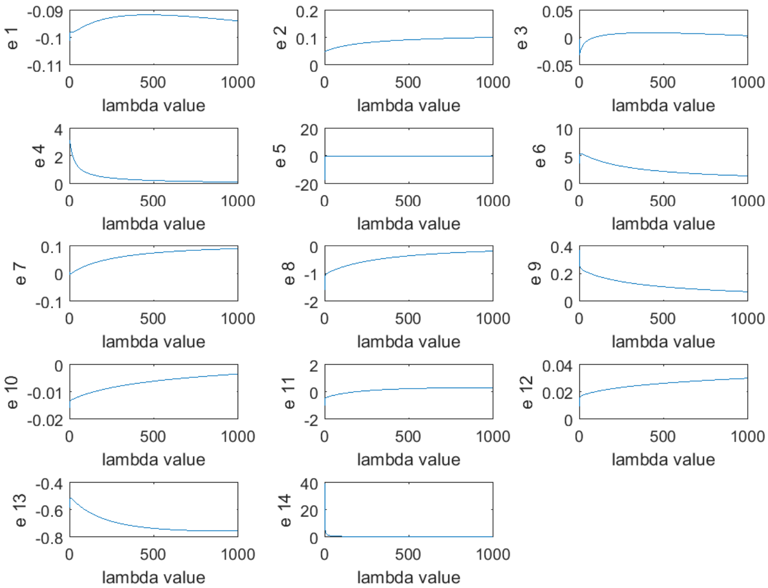

Figure 4: Elements of Weight Vector changes versus λ value

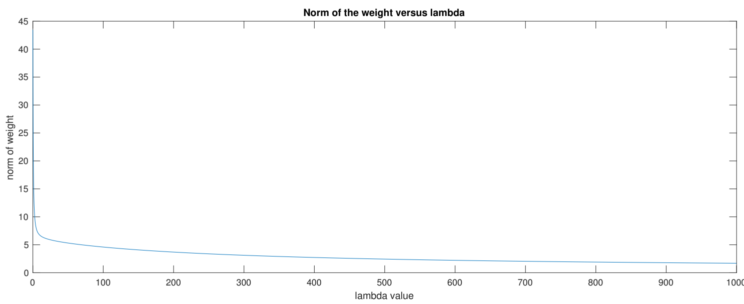

Figure 5: Norm of Weight Vector changes versus λ value

7 Problem 8

The equation provided in problem 8 aims to minimize SSE of the given data set by balancing the model complexity and the the level of fit of the data. Usually, the bigger the λ is the simpler the model is, and the smaller the λ is, the model more fit the given data set. Go back to the plot in last section, we can notify that a larger λ value leads to a Weight Vector containing more zero element. This fact implies that more features are being abundant and the model become simpler when λ grows. By controlling λ, programmers can find a most optimal model (Weight Vector) for the given problem.