Assignment 1

Brandon Lee, Alex Nguyen, Michael Lee

April 15, 2017

To set up the code, simply run *sh setup.sh*.

Part 1 and 2

Optimal weight vector w of training data with dummy column:

```
[  3.9584e+01  -1.0114e-01   4.5894e-02  -2.7304e-03   3.0720e+00
  -1.7225e+01   3.7113e+00   7.1586e-03  -1.5990e+00   3.7362e-01
  -1.5756e-02  -1.0242e+00   9.6932e-03  -5.8597e-01 ]
```

Part 3

SSE values with dummy variable

```
Training SSE:
9561.19128998
Test SSE:
1675.23096595
```

Part 4

SSE values without dummy variable

```
Training SSE:
10598.0572458
Test SSE:
1797.625625
```

How does the dummy variable impact training and testing SSEs?

Introducing the dummy variable decreases our SSE. Since the dummy variable is also correlated to our line of best fit (through w_0 and b), introducing the dummy variable improves our predictions.

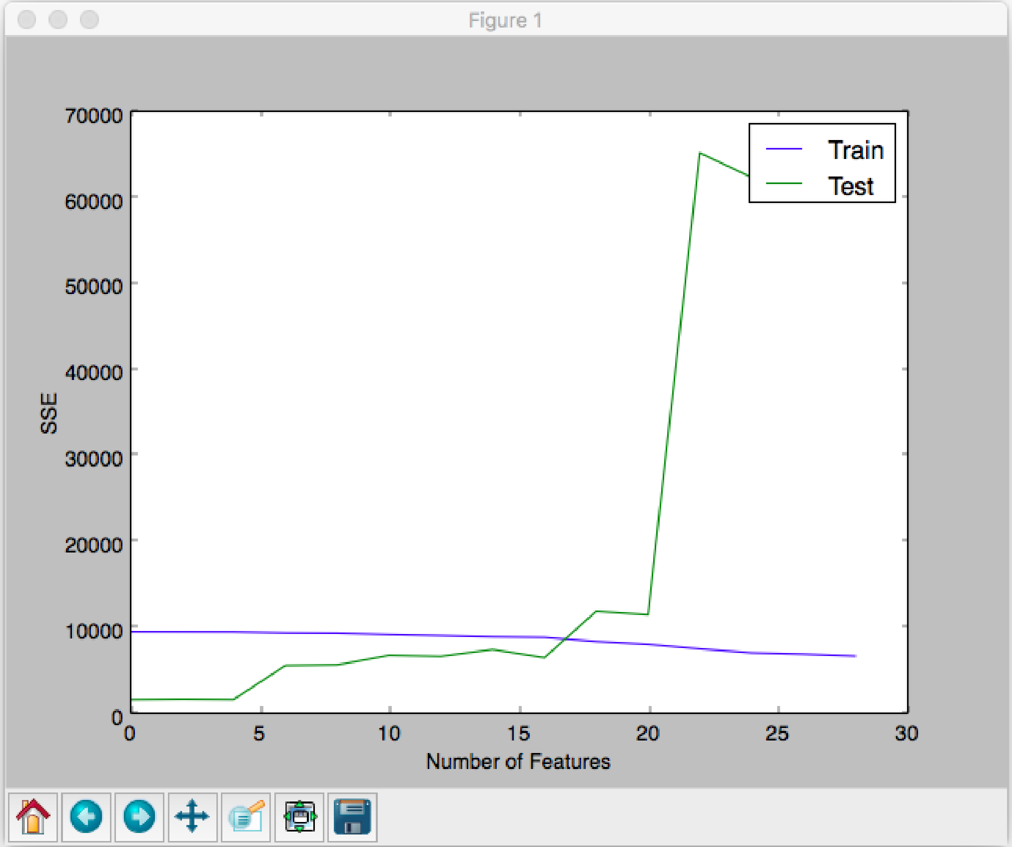

Figure 1: SSE Relation to Increasing Features

Part 5

No. of random features	Training SSE	Test SSE
0	9561.19128998	1797.625625
2	9560.09873557	1673.91252833
4	9545.30893303	1671.71935438
6	9524.89152143	1664.41585319

What trends do you observe? Do more features lead to better performance?

As observed from Figure 1, the more random features are uniformly distributed into the dataset, the greater the SSE value.

Part 6

λ	Training SSE	Test SSE
0.01	9525.23515759	1661.36416373
0.05	9532.30410068	1652.30507951
0.1	9549.9012481	1646.18662914
0.5	9754.35298078	1661.84766007
1	9938.27928143	1694.39789917
5	10328.9892371	1757.6160561

What behavior do you observe? What do you think is the best lambda value for this problem?

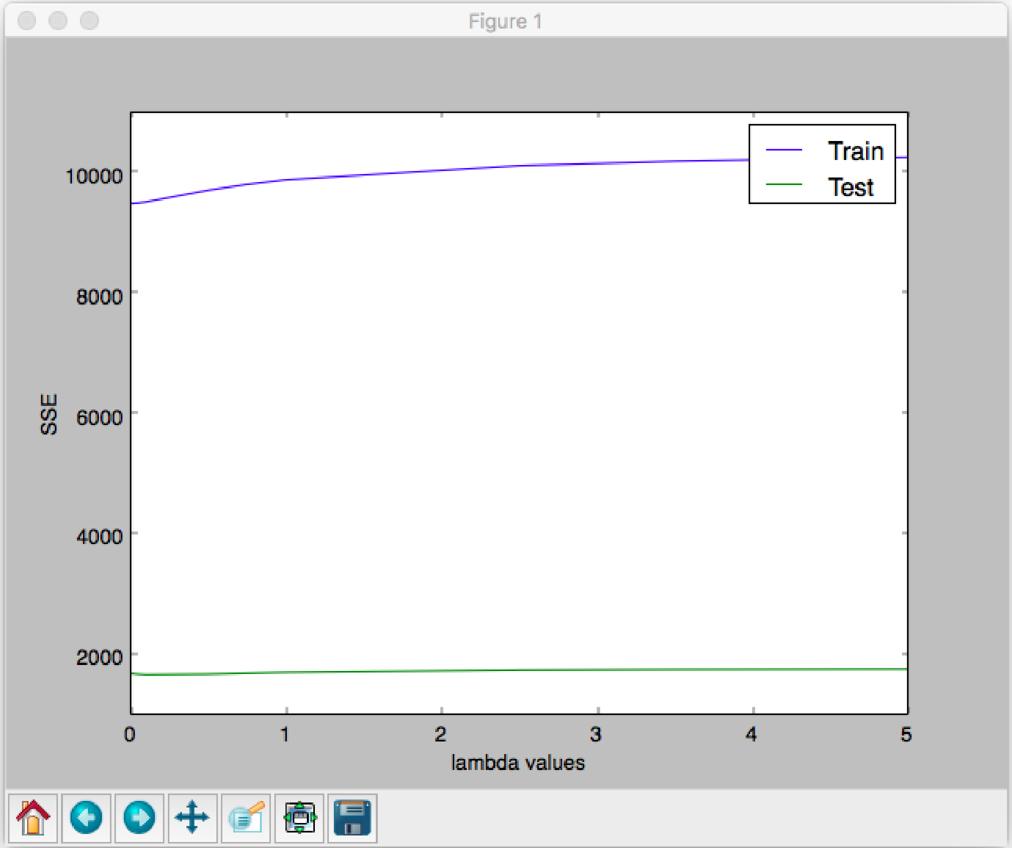

Figure 2: SSE Values Over Lambda

We noticed that the SSEs of the training data increase rapidly then start increase less and less as lambda increases. For the testing data, increasing lambda does not have that great of an effect. A good lambda for our data would be below 1 because there is a large jump in the SSE from going from 1 lambda to 5 lambda.

Part 7

Compare the different w's that you got in [6]. As the lambda value gets bigger, what impact do you observe it has on the weight values?

As the lambda values increase, the w values slowly starts decreasing until the until we hit a certain threshold. In our data the w values were continuously decreasing along the range of [0, 1]. When lambda $= 2.5$ we had a spike in our w values, but then as it increase on the range [2.5, 5] the w values again started to approach 0.

Part 8

Can you use this objective to explain the behavior that you observe in [7]?

In analyzing the given equation, we observe that we sum the SSE value and the regularization term. When the lambda values gradually increase, the regularization begins to predominantly control the summation. When lambda gradually decreases, the SSE value maintains predominant control over the summation. Since the summation represents the loss of the model, we focus on maintaining small w values for large lambda values. Inversely, we focus on maintaining closest fit for small lambda values.

From the data gathered from our experiment displayed in the tables, it is clear that when the lambda value is increased, the calculated weights are decreased to match the increase in focus on regularization. Additionally, the SSE values is demonstrated by observing the total instances being summed in the training data versus the testing data.

From Figure 2, both SSE trends appear to share a similar curve pattern. Subsequently the ratio between training and testing maintains at a constant rate. This is observed in our data as the ratio between the data points is almost equal to the ratio between the total training instances versus the total testing instances.