Implementation 1

SR Kanna, Ramcharan Sudarsanam, Preston Wipf

I. Problem 1 & 2

[[3.95843212e+01]
[-1.01137046e-01]
[4.58935299e-02]
[-2.73038670e-03]
[3.07201340e+00]
[-1.72254072e+01]
[3.71125235e+00]
[7.15862492e-03]
[-1.59900210e+00]
[3.73623375e-01]
[-1.57564197e-02]
[-1.02417703e+00]
[9.69321451e-03]
[-5.85969273e-01]]

II. Problem 3

SSE Training data: 9561.19128998
SSE Testing data: 1675.23096595

III. Problem 4

SSE Training data: 10598.0572458
SSE Testing data: 1797.625625

The SSE, or loss, increases as when we lose the dummy variable. Because the dummy variable correlates to our b or w_0, which is what translates our line of best fit, removing the variable creates a less accurate prediction.

IV. Problem 5

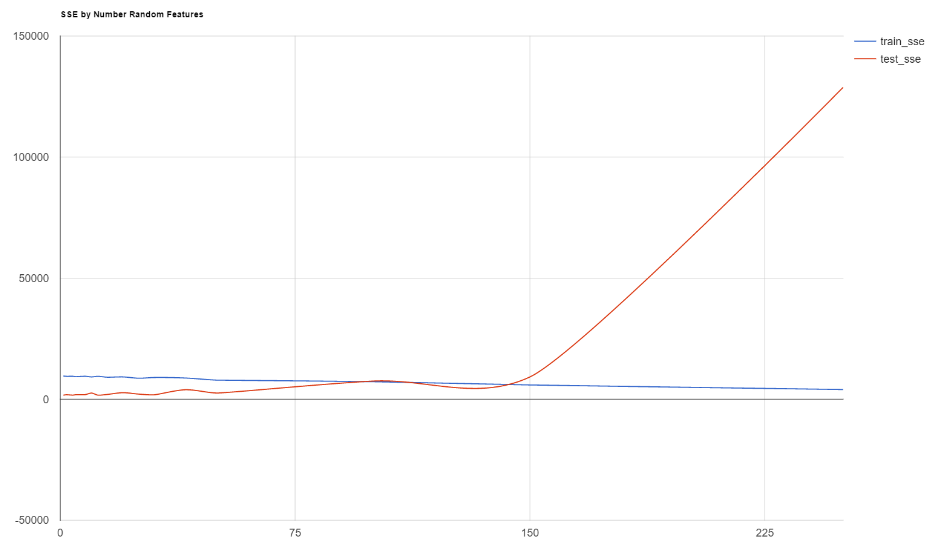

Figure 1: Training SSE decreases linearly; Testing SSE increases exponentially.

Based on the plot above, as more random, uniformly-distributed features are added to the datasets, the SSE of the training predictions decrease linearly, but the SSE of the testing data increases exponentially. Intuitively, this is the result of the model being trained very specifically to match the training dataset, which creates a better prediction particular to the training dataset due to overfitting and results in a far worse prediction of the testing dataset.

V. Problem 6

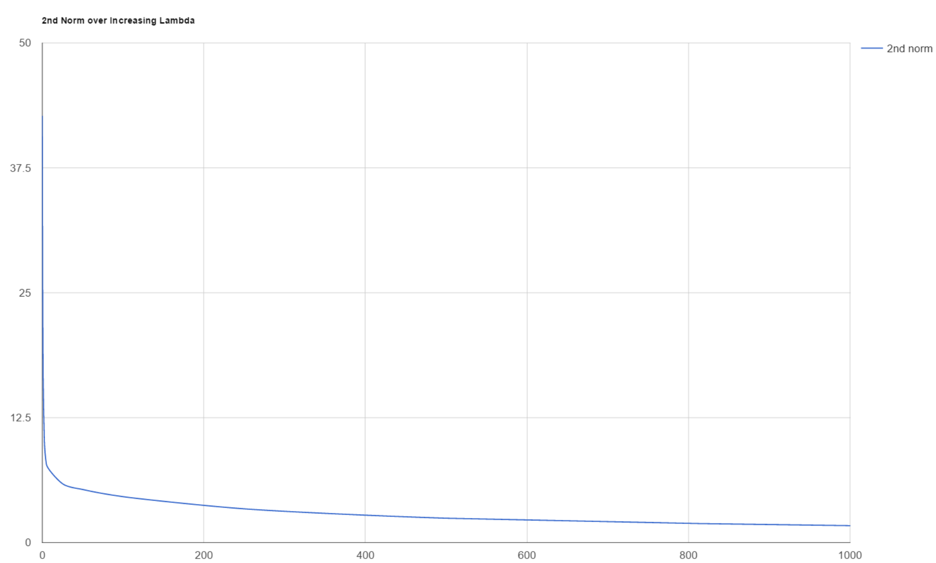

Figure 2: The norm of vector w converges to 0 as the lambda increases.

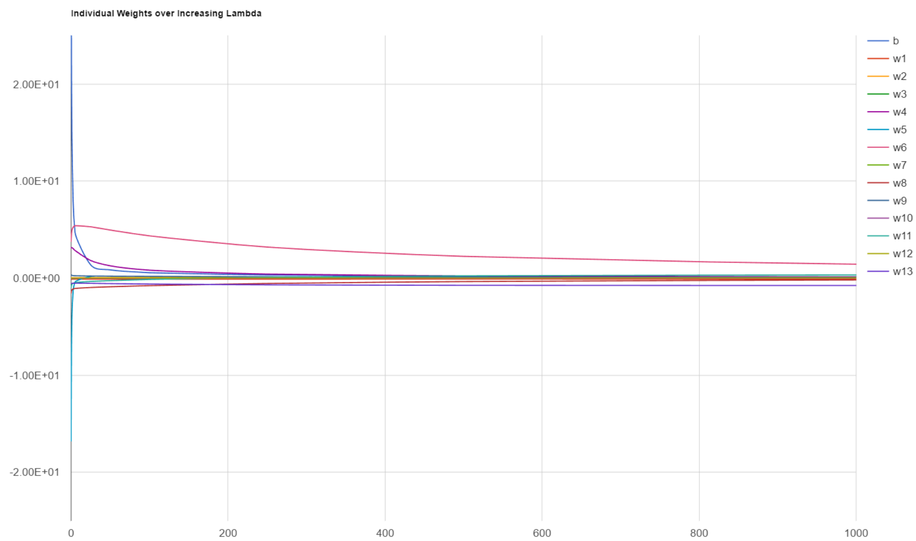

Figure 3: The individual weights converge to 0 as the lambda increases.

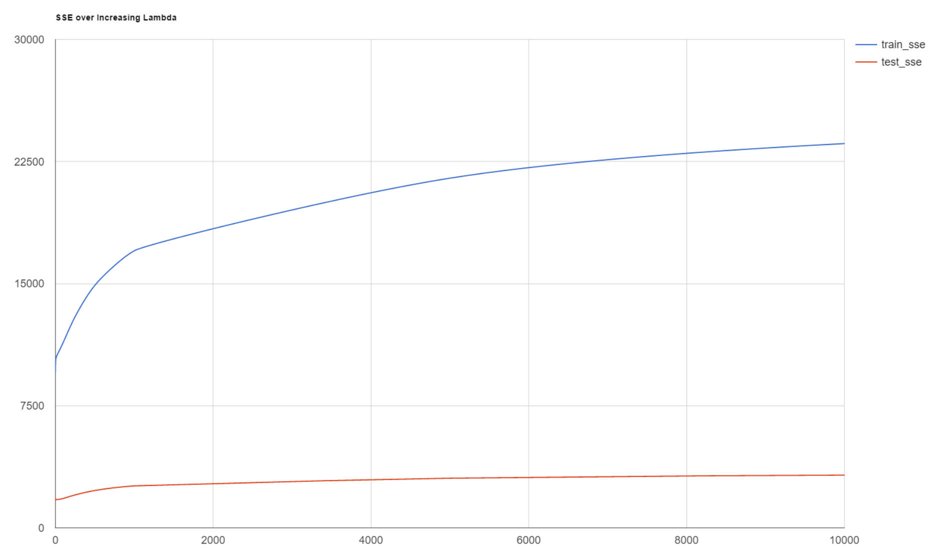

Figure 4: SSE's over the increasing lambda.

VI. Problem 7

Looking at figure 2 and figure 3, we see that as lambda increases the weights of each feature begin to converge to 0. Given a large enough lambda, the weights are essentially 0, meaning each weight holds nearly no predictive power. Obviously if the individual weights converge to 0, then the norm of the vector w also converges to 0 as lambda increases.

VII. Problem 8

Based on this equation, we sum across two terms: the standard SSE and the regularization term. As values of lambda increase, the regularization term dominates the summation. As lambda decreases, the SSE term dominates the summation. The summation as a whole represents the loss of the model; if we are therefore minimizing loss, for large values of lambda we will prioritize small magnitudes of w. On the other hand, for small values of lambda, we are prioritizing the closest fit possible.

Looking at our chart then, it makes sense that as we increase the value of lambda the calculated weights are made forcibly smaller to accommodate our increasing prioritization of regularization. Further, I think the SSE can be explained by looking at the number of instances being summed in the training set relative to the testing set. Based on figure 4 it appears each SSE curve has a similar shape, and the ratio between the two stays relatively constant. What reinforces this theory is the fact that the ratio between the curves at each point is nearly the ratio between the number of training instances to the number of testing instances.