CS 434 Assignment 1 Writeup

Jacob Fenger
Spike Madden

2. Our learn optimal weight vector:

[[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]]

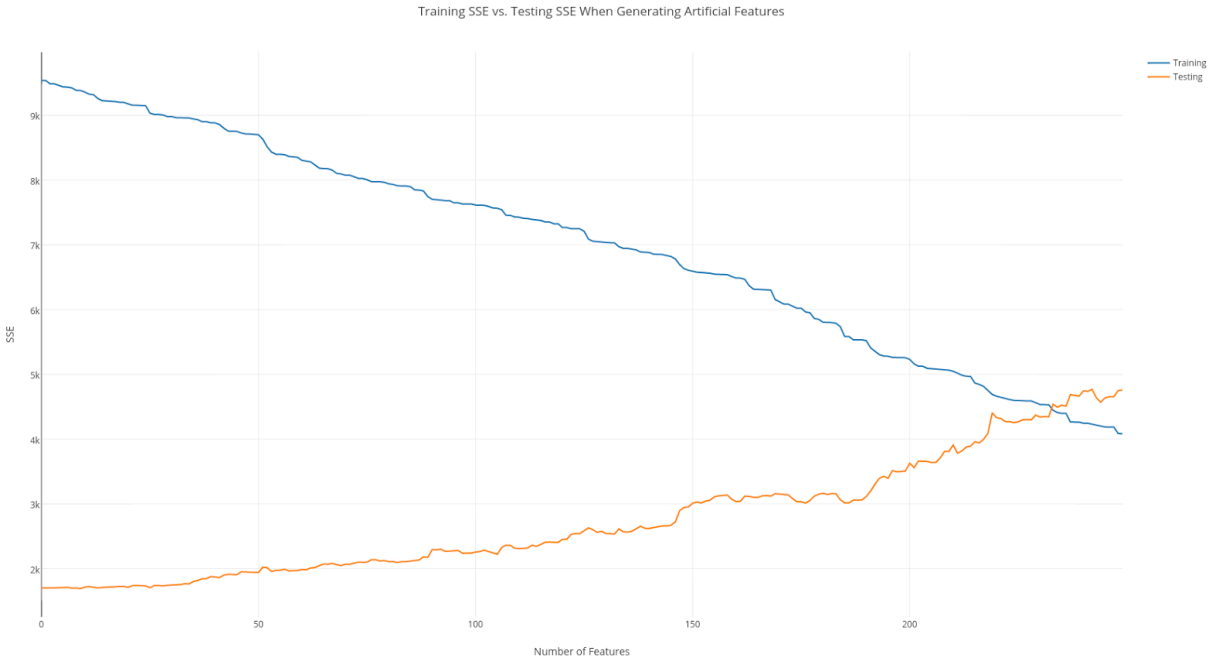
3.

Training data SSE: 9561.19128998
Test data SSE: 1675.28751863

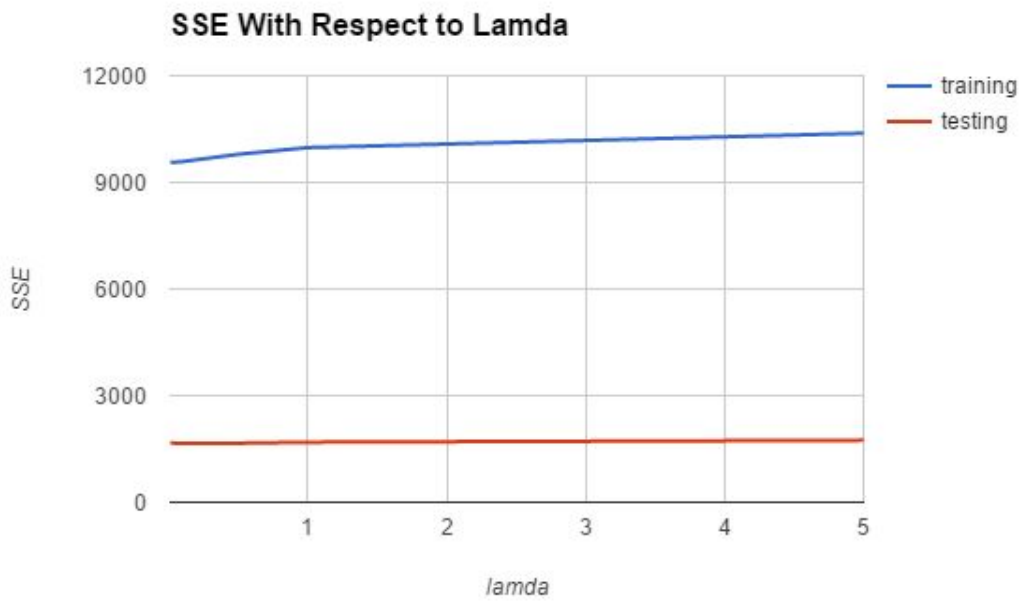
4. With no dummy variable, the SSE goes up for both the training and testing SSEs.

Training data SSE: 10598.2241923
Test data SSE: 1796.28751863

5. Generating more features results in a lower SSE for the training data while the testing data results in a higher SSE. This is because we are computing our optimal weight vector specifically with these additional features, resulting in a much lower SSE. This can be shown by the visual below where we eventually see an intersection of training and testing SSEs as the number of features reaches ~230.



6. As we increase λ, the SSE values increases. This trend is not evident within small values of λ, but with more extreme values show a huge SSE increase. The best value for λ is around 0.1. At this value, we see the lowest SSE for the testing data.



7. As λ gets bigger, the computed weight values decrease. In other words, the weight values converge to 0 as λ increases.

8. As λ increases, the regularization term will increase resulting in a higher SSE. If it decreases, the resulting norm of the weight vector will be dependent on the data term and not the regularization term.