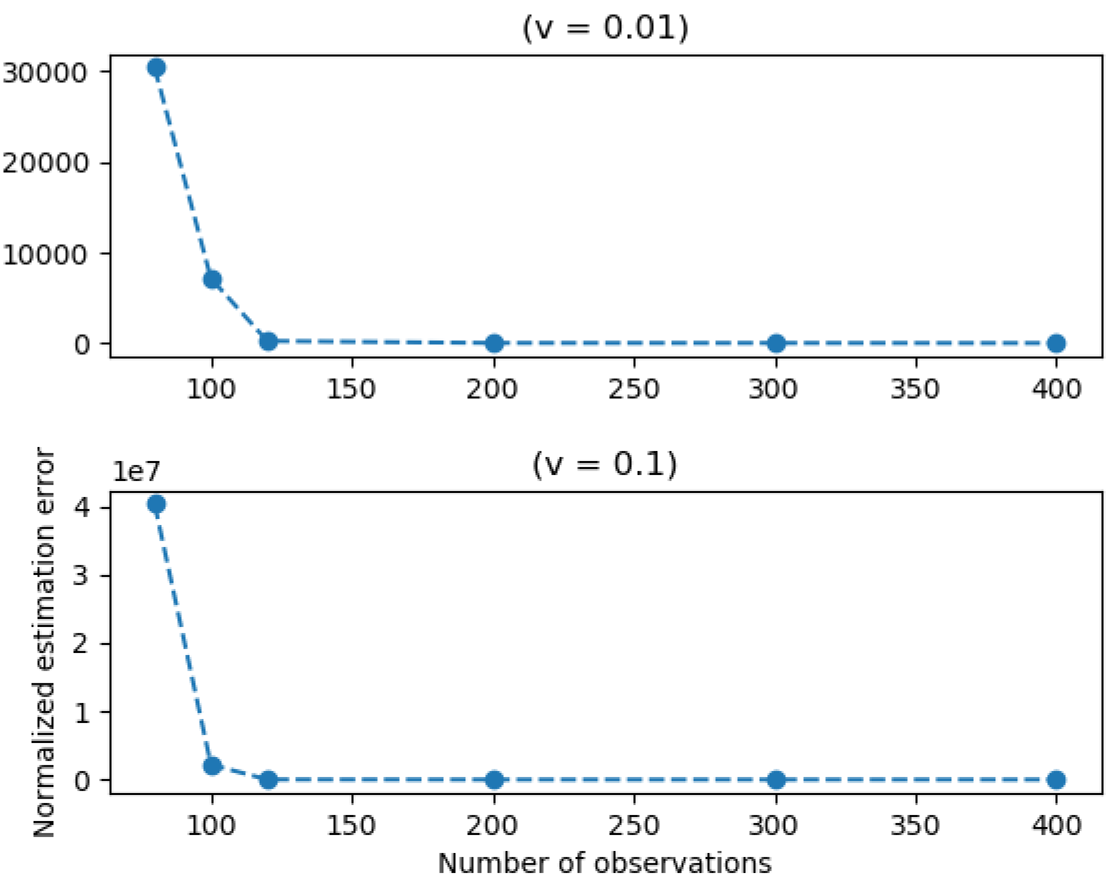
Adam Thelen and An Nguyen

EE 425X

2/7/2020

Homework 1b

1. In question one, the goal is to see how varying m affects the models ability learn theta. To quantify this effect, simulations were run for m = 80, m = 100, m = 120, and m = 400. Plots of the estimation error in theta and normalized Test-MSE are below:



Figures 1 – Normalized estimation Error $\sigma_e^2 = 0.01\|\theta\|_2^2$ and $\sigma_e^2 = 0.1\|\theta\|_2^2$

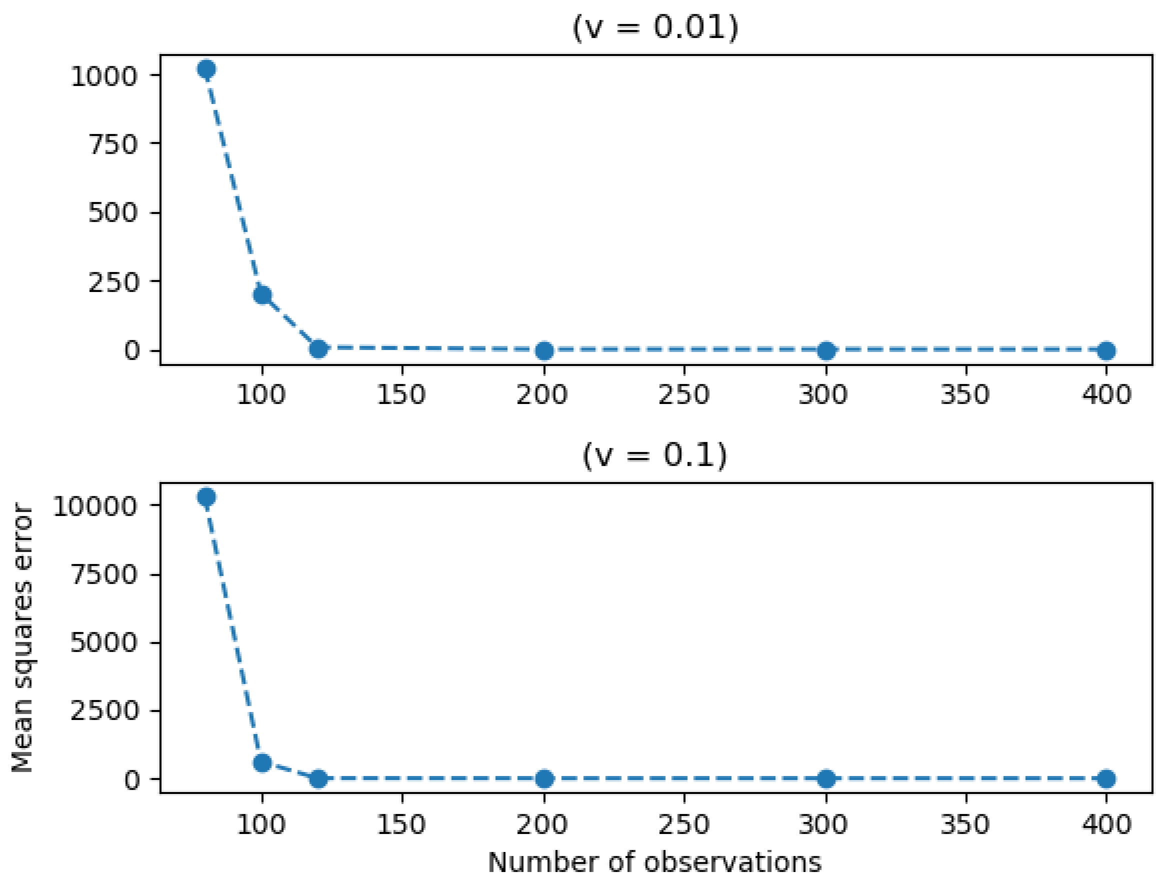


Figure 2 –Test-MSE with $\sigma_e^2 = 0.01\|\theta\|_2^2$ and $\sigma_e^2 = 0.1\|\theta\|_2^2$

Looking at Figures 1 through 4, it is clear the model is unable to learn theta for values of m which are less than n. The solving strategy used for problem one was solving the normal equations. Normal equations only operate given that m is greater than or equal to n. In the case of m = 80, there is no solution, thus the extremely large estimation error and Test-MSE.

For the cases when variance is small, fewer data is needed to estimate the parameters correctly. We can see that by comparing the lines in figure 1 and 2, which dip down much faster for v = 0.01. In figure 1, at 200 observations, we have had a very good estimate of the parameters already, whereas the same cannot be said for figure 2.

2. In this experiment, the goal was to determine the effect of small n values given a fixed m value of m = 80. Simulated data was constructed like problem one. Two sets of data were used, one with $\sigma_e^2 = 0.01\|\theta\|_2^2$ and the other $\sigma_e^2 = 0.1\|\theta\|_2^2$. The results are below:

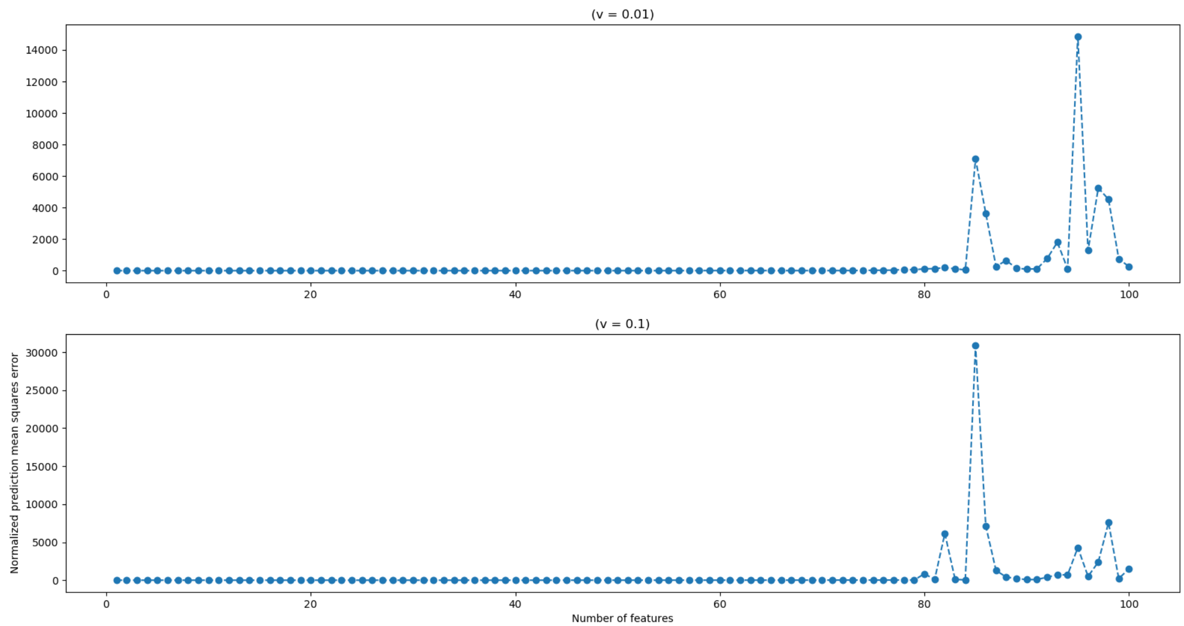


Figure 3: Test-MSE against the number of features included

The np.argmin(test-mse) output indicates that an n_small of 1 provides the smallest normalized test mean squares error (for this set of simulated data).

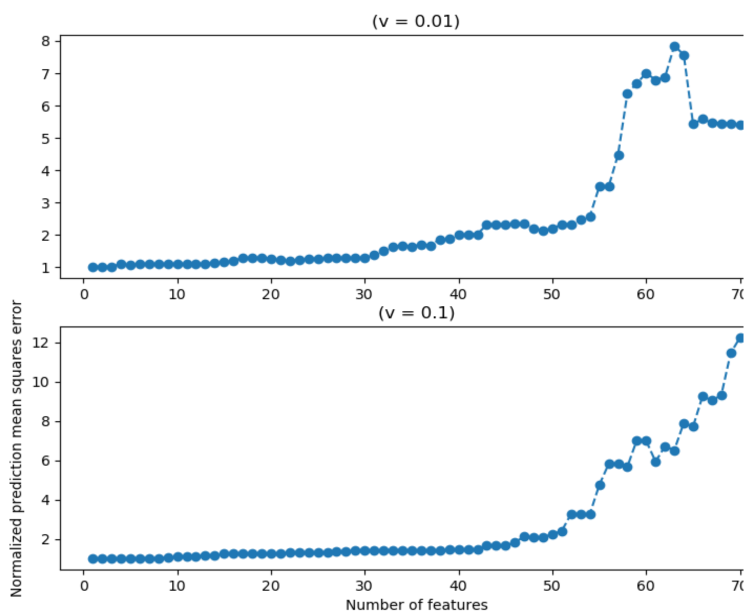


Figure 4: Test-MSE against the number of features included (Zoom in)

3. Based on the variance – bias tradeoff lectures, we were expecting something like a U-shape parabola. But the simulation result showed otherwise. We have tried all three methods of estimating the model parameters (Pseudo inverse, normal equations, gradient descent) and still receive the same result.