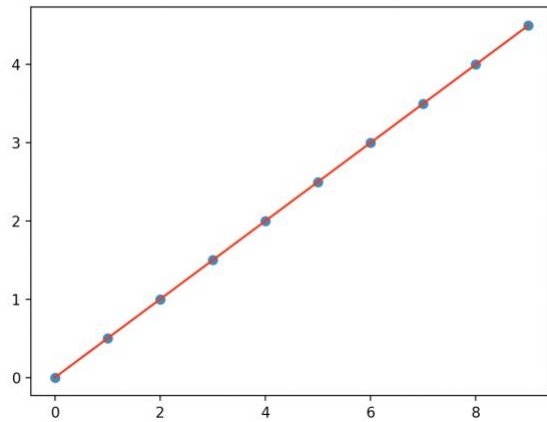
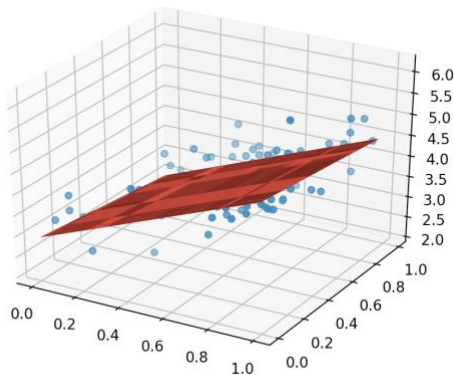


Zachary Panzarino

1a. For the 1D with no noise, the loss is $8.45868431574874e-32$. The plot is shown below:



For the 2D noisy, the loss is 0.10759283475200875. The plot is shown below:



1b. When using the closed form solution and one of the features is duplicated, the function errors because it is no longer invertible.

1c. This error does not occur when one of the training points is duplicated because it is still invertible.

1d. Gradient descent will work in either of the above cases. This happens because it does not require an inversion.

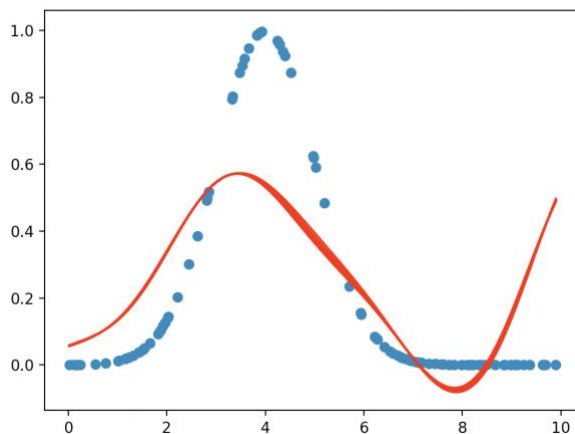
2a. Iteration: 1 , Loss: 3.5625 , Theta: [[0], [0]]
 Iteration: 2 , Loss: 2835.890625 , Theta: [[2.25], [14.25]]
 Iteration: 3 , Loss: 2258035.0078125 , Theta: [[-61.875], [-387.75]]
 Iteration: 4 , Loss: 1797926500.0751953 , Theta: [[1747.125], [10955.8125]]
 Iteration: 5 , Loss: 1431572003454.5208 , Theta: [[-49298.90625], [-309132.65625]]
 Iteration: 6 , Loss: 1139867731516952.2 , Theta: [[1391099.203125], [8723007.375]]
 Iteration: 7 , Loss: 9.076025810914652e+17 , Theta: [[-3.92535309e+07], [-2.46142635e+08]]
 Iteration: 8 , Loss: 7.226649394729697e+20 , Theta: [[1.10764186e+09], [6.94556337e+09]]
 Iteration: 9 , Loss: 5.754111167416799e+23 , Theta: [[-3.12550351e+10], [-1.95987381e+11]]
 Iteration: 10 , Loss: 4.581624694721907e+26 , Theta: [[8.81943214e+11], [5.53030063e+12]]

2b. The model parameters are not the same as the closed form solution as the theta value becomes quite large along with the loss. The closed form solution was able to give values that were more accurate.

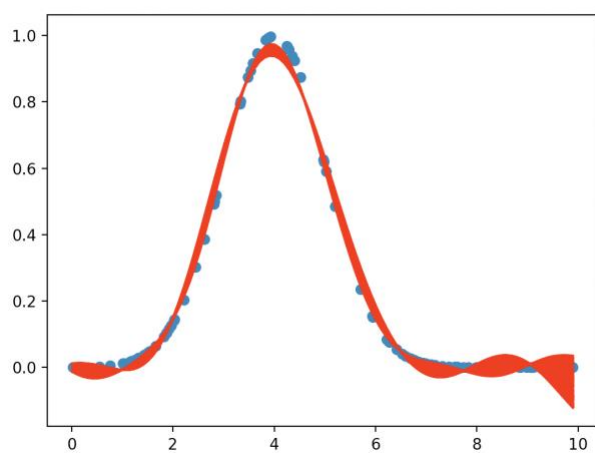
2c. A set of an alpha value of .01 with a number of iterations of 100 and an alpha value of .02 with a number of iterations of 100 provides the same answers. This occurs because they both have a low learning rate, so they do not pass over the minimum, and a high number of iterations ensures that they are both close to the minimum. An alpha value of .01 with a number of iterations of 100 and an alpha value of 0.5 with a number of iterations of 100 provides noticeably different answers. This occurs because the learning rate for the second is so high that it passes over the minimum, while the first has a low learning rate and correctly approaches the minimum.

3a. For all of the models, the small value of K was 10, the medium value of K was 100, and the large value of K was 1000.

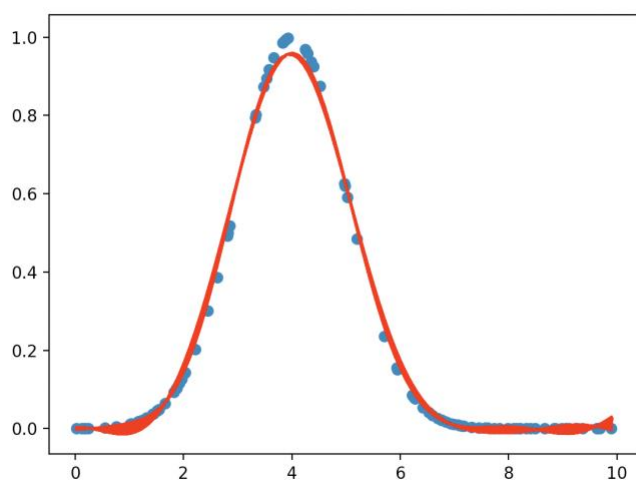
1D-exp-samp.txt small



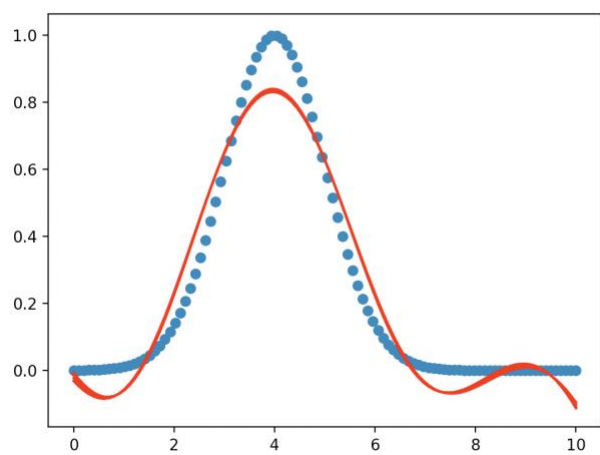
1D-exp-samp.txt medium



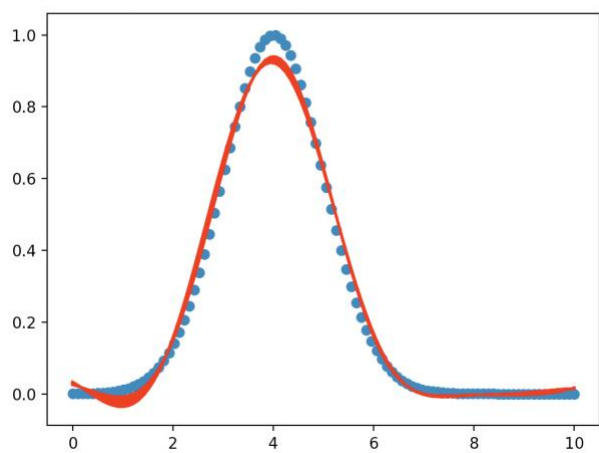
1D-exp-samp.txt large



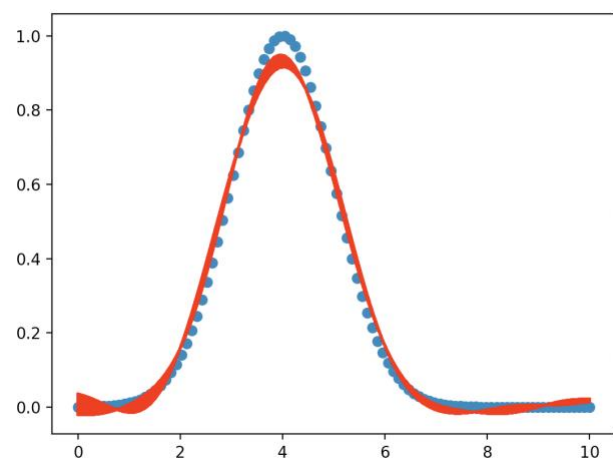
1D-exp-uni.txt small



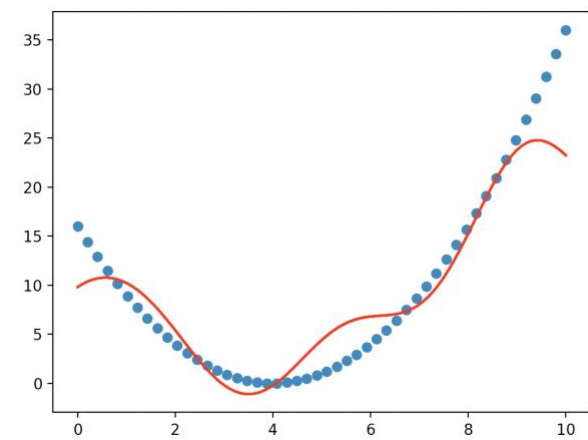
1D-exp-uni.txt medium



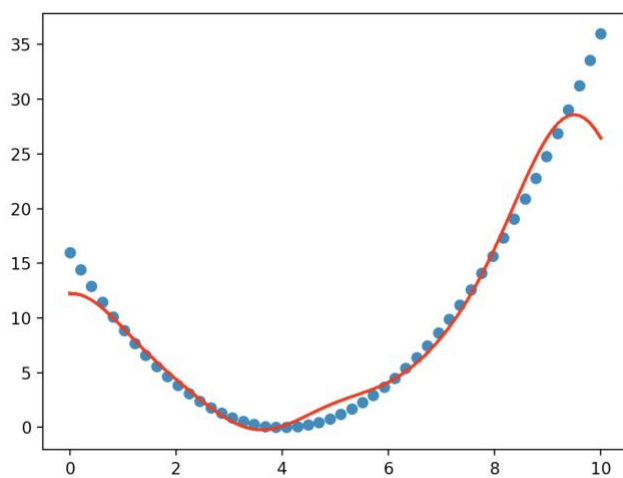
1D-exp-uni.txt large



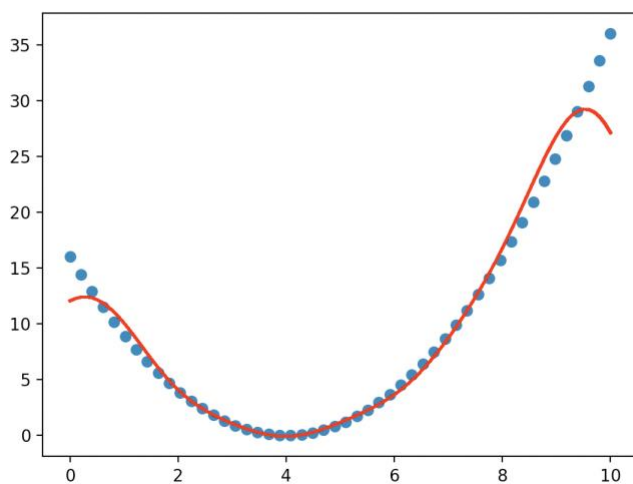
1D-quad-uni.txt small



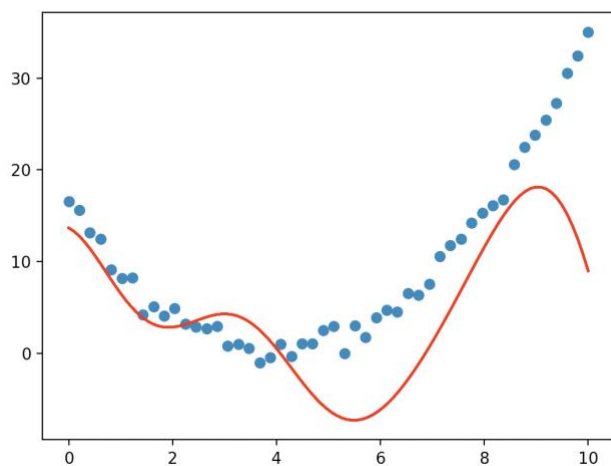
1D-quad-uni.txt medium



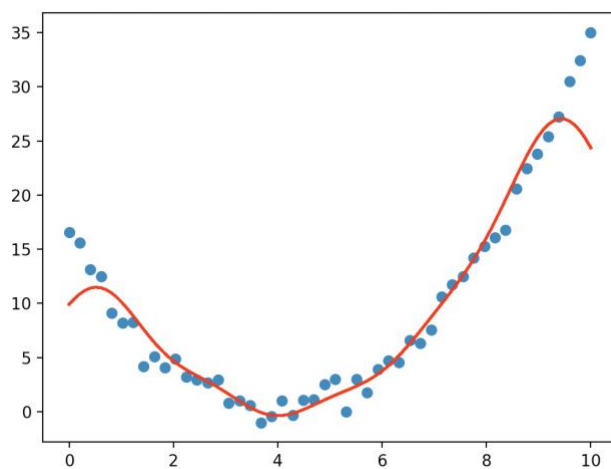
1D-quad-uni.txt large



1D-quad-uni-noise.txt small



1D-quad-uni-noise.txt medium



1D-quad-uni-noise.txt large

