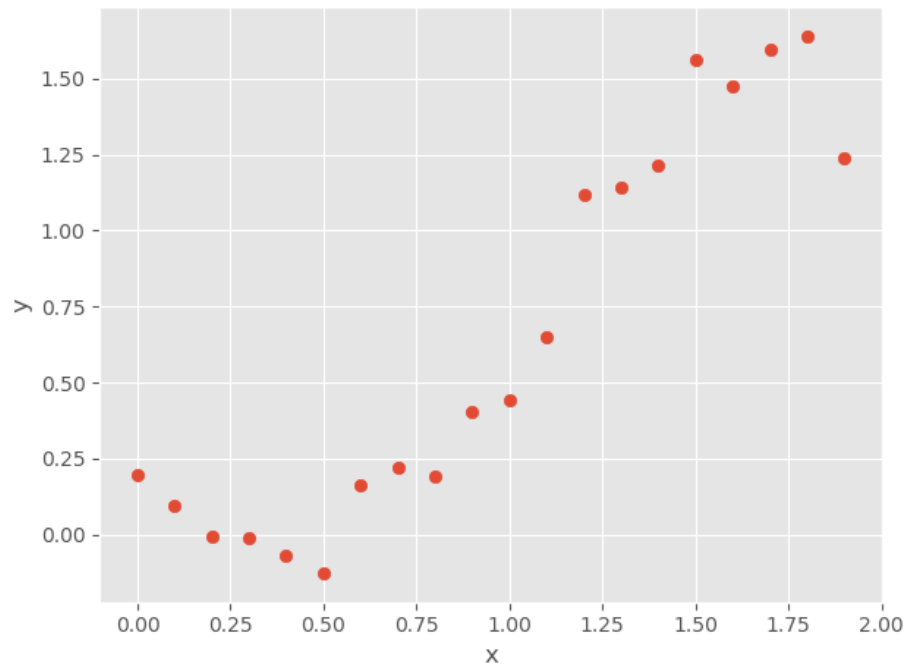


Machine Learning Project 1

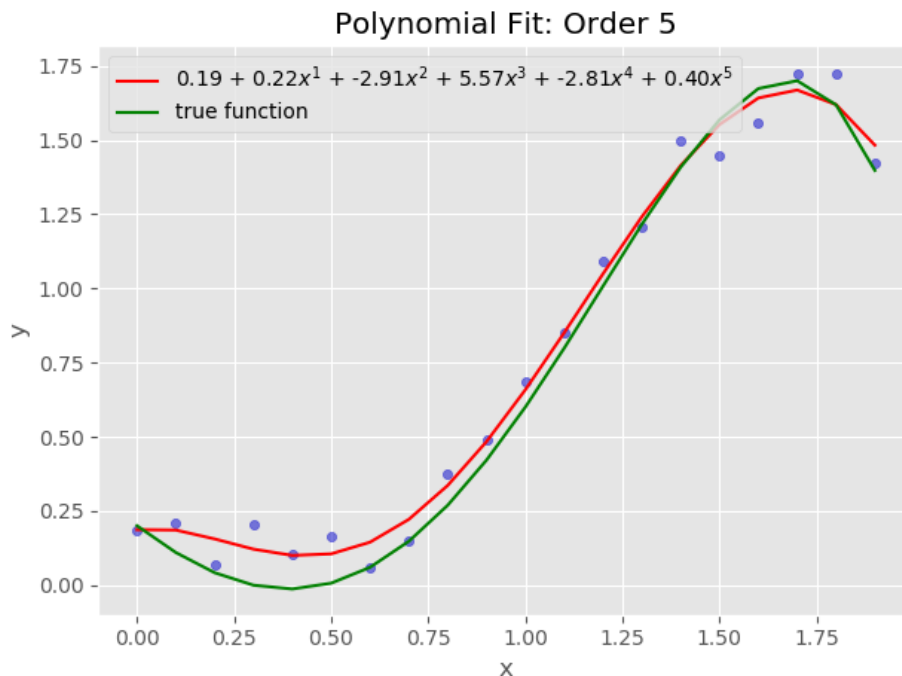
Σιατήρας Βαγγέλης

Problem 1

1.1)

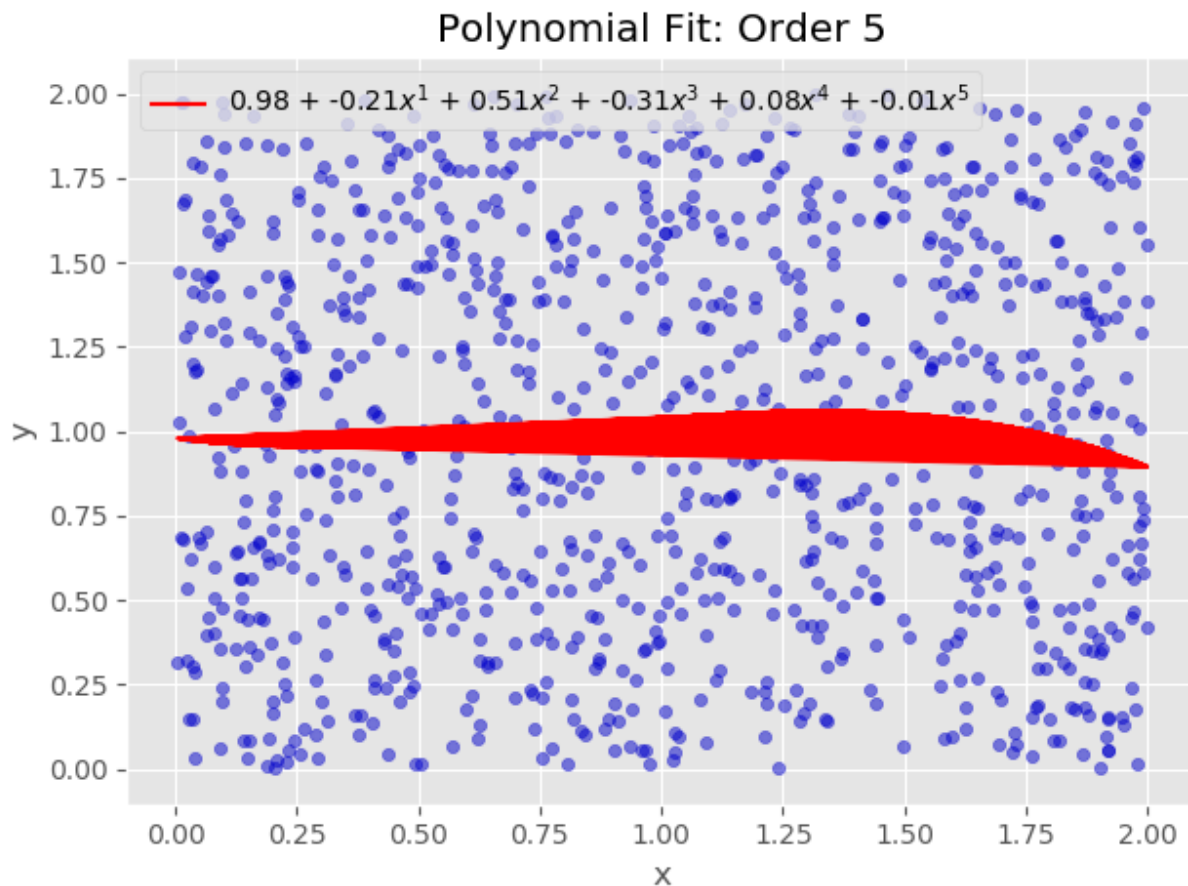


In the first plot we see the points that are generated by applying in our true function 20 equidistant points in $[0,2]$. We can see that they are following and underlying normal-gaussian distribution due to the values of the coefficients and the added gaussian i.i.d samples.



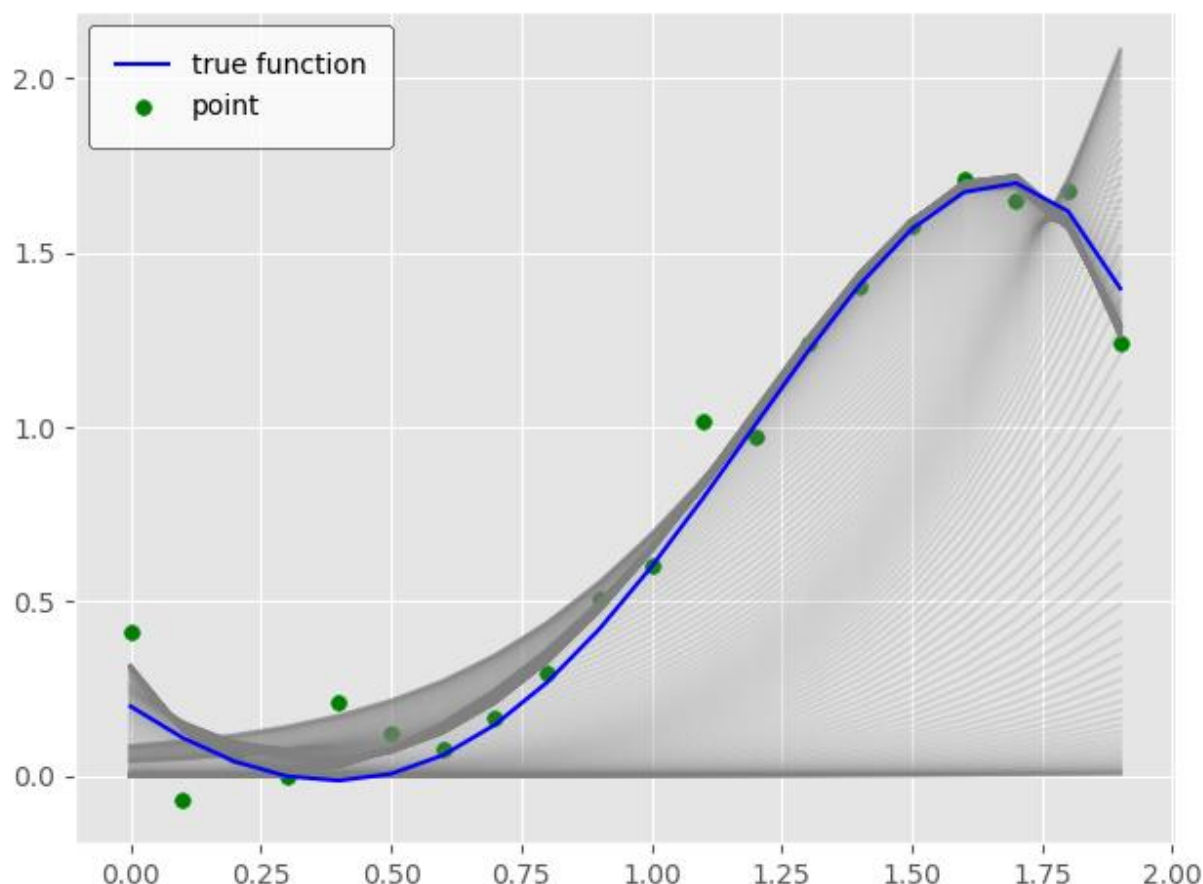
We fit our generated points to a least squares polynomial regressor of order 5. We see that the predicted line is very close to the true line and underfits slightly the data. From that we can assume that the Mean Square Error is very small because the distance of the points from the curve is very small. Actually, the MSE of the predicted model mentioned in the legend of the plot is proportional to 0.006. A small MSE means that our regressor is very efficient.

Now if we introduce 1000 points following a random distribution (a continuous uniform distribution) as they are limited to $[0,2]$ and fit them to a 5th degree polynomial least squares regressor, we see that the model is seriously underfitting the data. As they are not scattered uniformly, the variance between the predicted line is much bigger, therefore the MSE will be much bigger. Actually, the MSE for the below plotted line is proportional to 0.32.



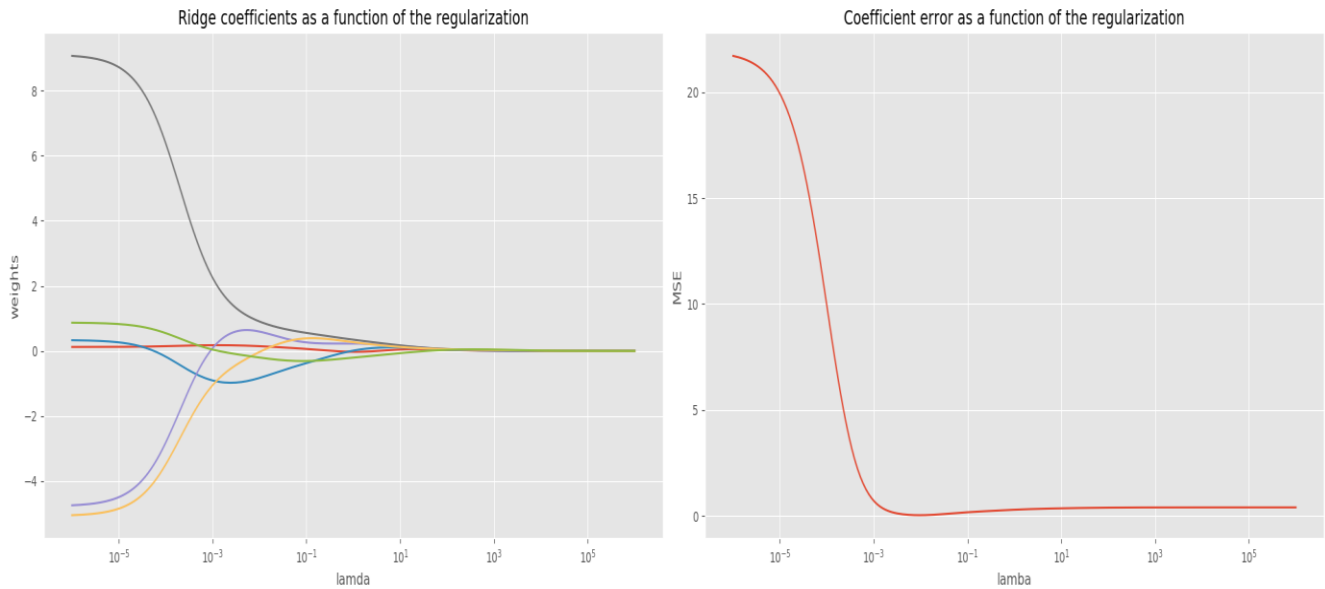
It is very clear that our regressor is doing very bad because we have points scattered non uniformly and because we apply LS with a low order polynomial relatively to our training set.

1.3) In the term of ridge regression we introduce the amount of bias by changing the importance parameter (lamda) and we are getting the following results.



As per the legend of the plot we have our true model represented by the blue curve and with the light grey curve we introduce the Hypothesis ($\theta \cdot X$) or in other words the predicted line for various values of λ . As the λ increases, the slope is starting to minimize, getting closer and closer with the x axis. We can also see that progressively for the various λ s, the changes in the variable x do not affect so much the changes in the variable y , so by introducing a small amount of bias we have a significant drop in variance.

Next, we have the plots of the Ridge coefficients and the coefficient error. Each color in the left plot represents one different dimension of the coefficient vector and this is displayed as a function of the regularization parameter. The right plot shows how exact the solution is.



As λ tends towards zero, the coefficients found by Ridge regression stabilize towards the randomly sampled vector w . For a big λ (strong regularization) the coefficients are smaller (eventually converging at 0) leading to a simpler and biased solution. These dependencies can be observed in the left plot.

The right plot shows the mean squared error between the coefficients found by the model and the chosen vector w . Less regularized models retrieve the exact coefficients (the error is equal to 0), while stronger regularized models increase the error.