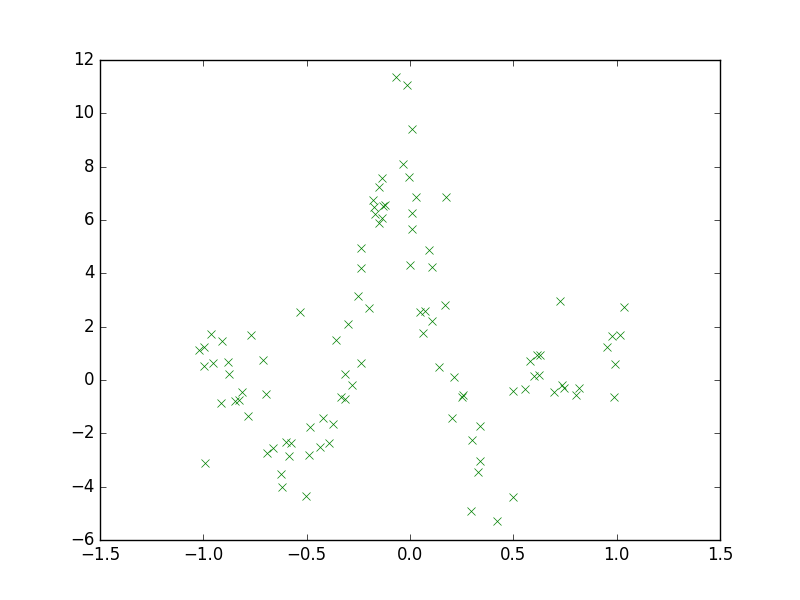
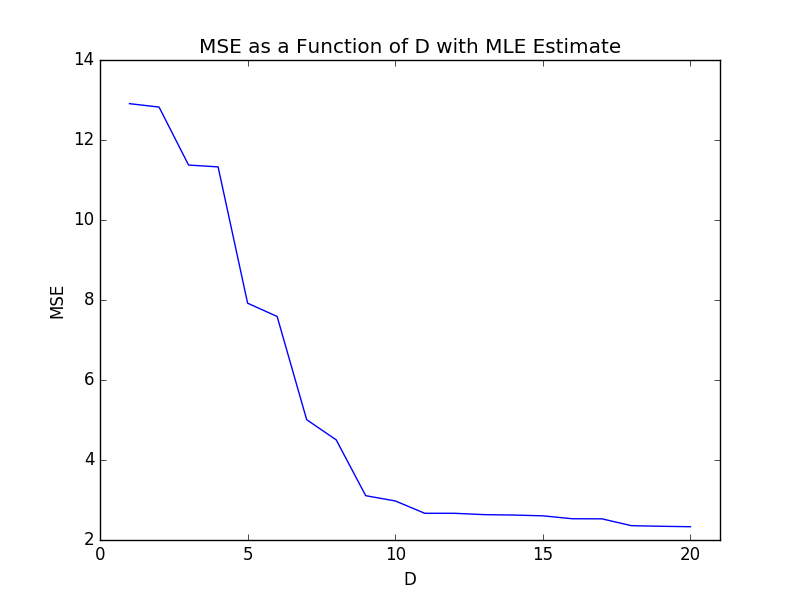
Assignment 2 Report

Anton Sitkovets

**Step 1**

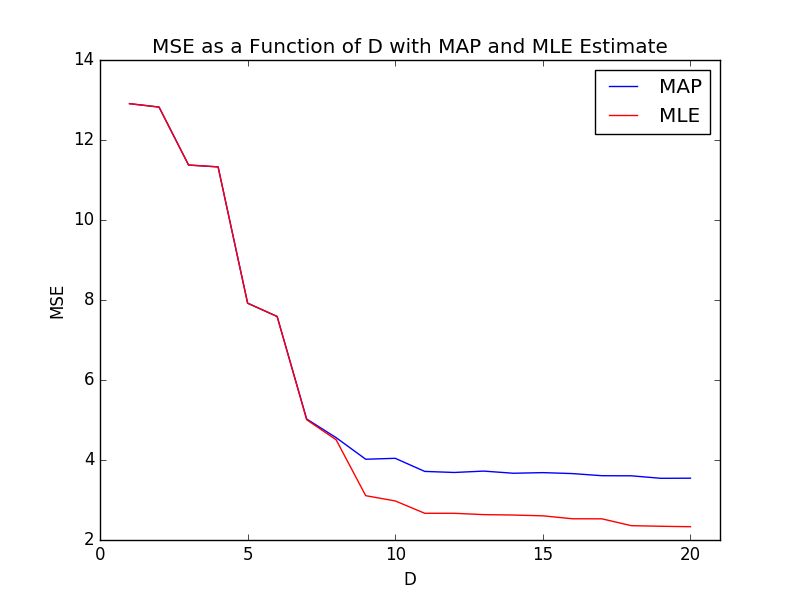


**Step 2**



Based on the values we can see that when D = 20 gives the best result as it has the smallest error when comparing the predicted values to the actual values. Although the data shows that this is the best value, I would say that using a polynomial of (20 – 1) = 19, to be very complicated and could potentially cause overfitting. As discussed in class, a good way to avoid overfitting is to use simpler basis functions, so I think the best bet would be to use D = 11 to 16.

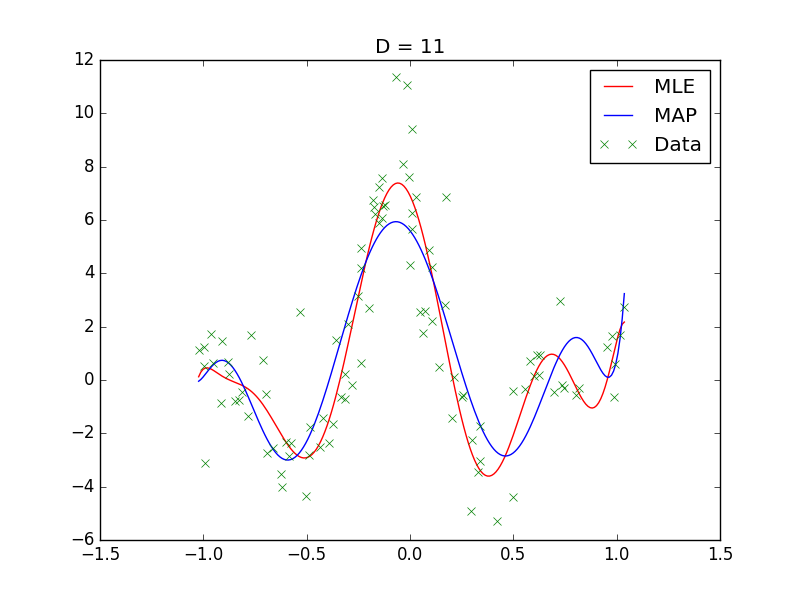
**Step 3**



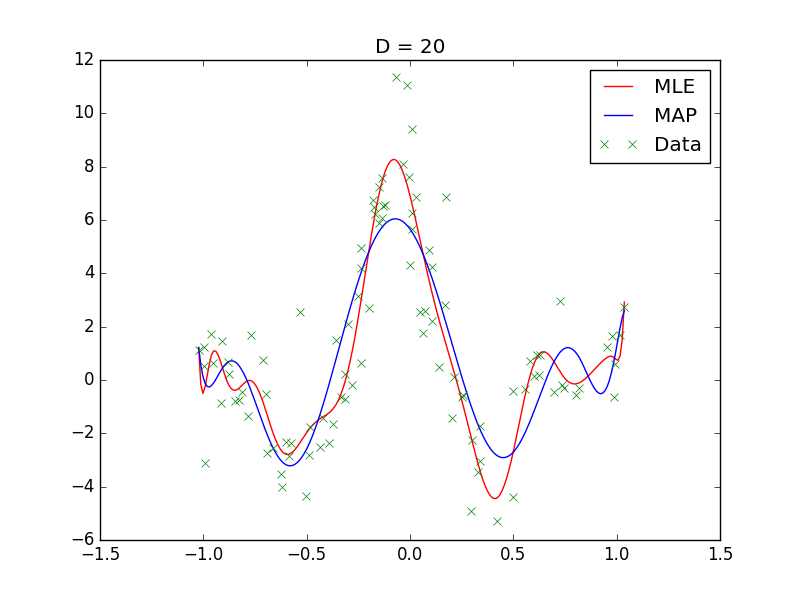
The MSE for the MAP estimate stops decreasing after a point while the ML case continues to decrease because the purpose of the MAP estimate is to add a prior to encourage smooth models and penalize more complex models. So as D increases, the prior begins to have more and more effect on the quantities that one wants to estimate. Meaning that as we increase D, the resulting estimates begin to rely more on the prior rather than the actual distribution of data, making its predictions less accurate and its mean squared error large. Hence as can be seen from the result the MAP estimate discourages models with high complexity as unlike the ML estimate, the mean squared error doesn’t continue to decrease as rapidly after a certain value of D.

**Step 4**

Since the MAP estimate function is used to penalize complex models, using the results from step 3 I would say that D = 11 would give the best result for the MAP estimate. After looking at the results I can see this is the case, as when D increases after 11, there is only minor improvements. But we consider these minor improvements worth the added complexity. Also, we can assume that real world phenomena are mostly smooth, so we can say that adding a prior for the MAP estimate adds those smoothness assumptions. Since the data should be smooth, the plot should not have all these short curves as you can see in higher order MLE functions. Below we can see the MAP estimate fits the data fairly well.

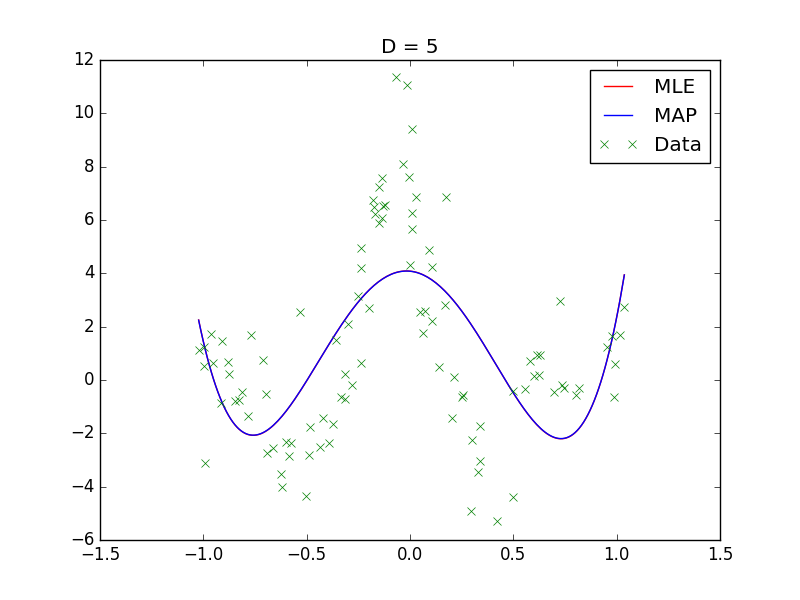


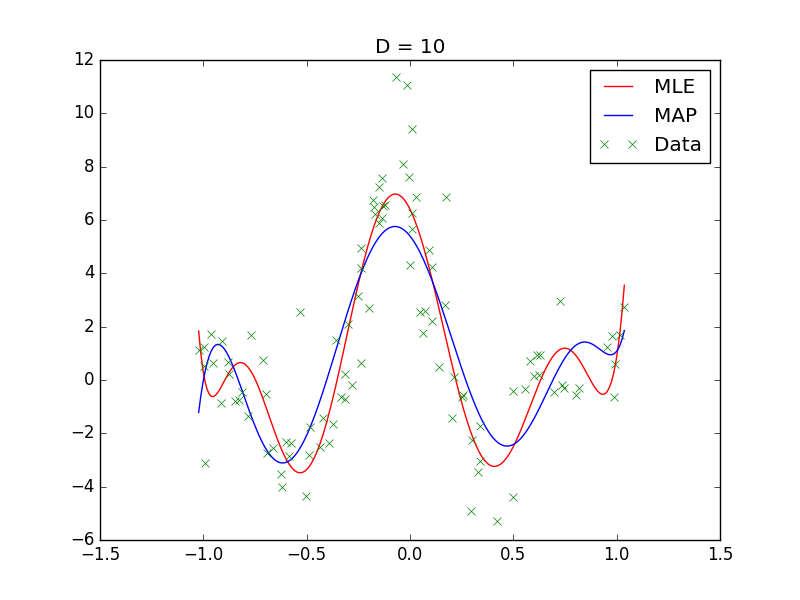
From Step 2, I found that when D = 20, the MSE is the lowest for the MLE and looking at the plot we can see that the results are quite good for the dataset. From the plot below we can see that the plot is a good fit for the MLE as well as the MAP estimate. But although this seems like the best fit, as mentioned in step 2, to prevent over fitting the data we should use a less complex model as real world data is generally smooth and doesn’t need this complexity of curves as seen below for the MLE.

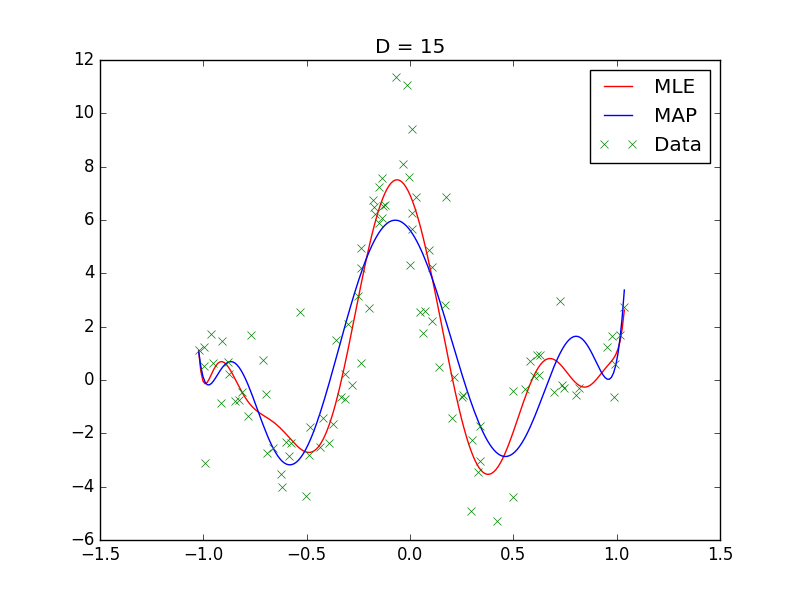


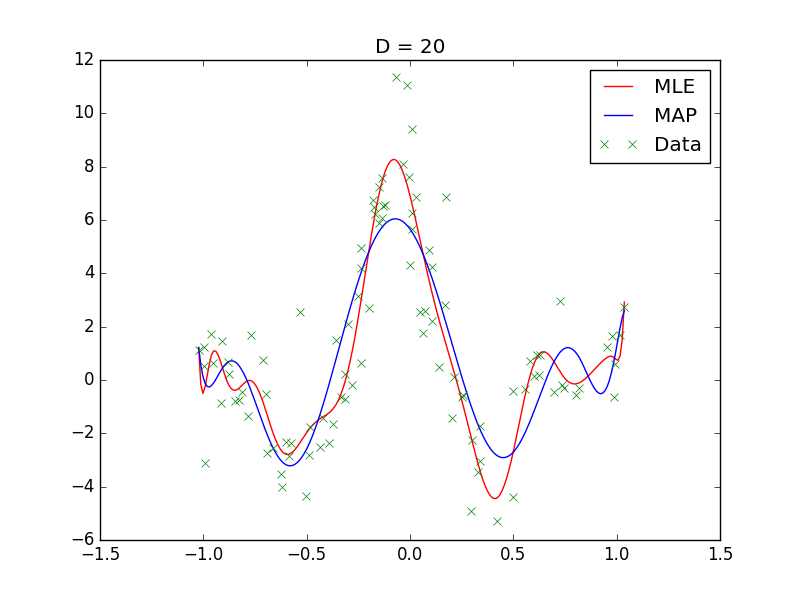
The differences between the MAP and ML estimated functions as D increases can be seen clearly from the results of step 3. As D increases, the MAP and ML estimates tend to decrease in the mean squared error, but the MAP will cut off at a certain point and will stop decreasing. This can be attributed to the prior knowledge added to the model with the MAP, as we are trying to enforce the use of smoother and simpler models in order to simulate the phenomena that real-world situations are typically smooth. This is all done to prevent overfitting the data and making sure that we are not using a very complex model that fits our training data perfectly, but poorly with test data. Hence, MAP estimates should be used over ML estimates to prevent the use of complex models and overfitting.

Below are plots for D e {5, 10, 15, 20}:

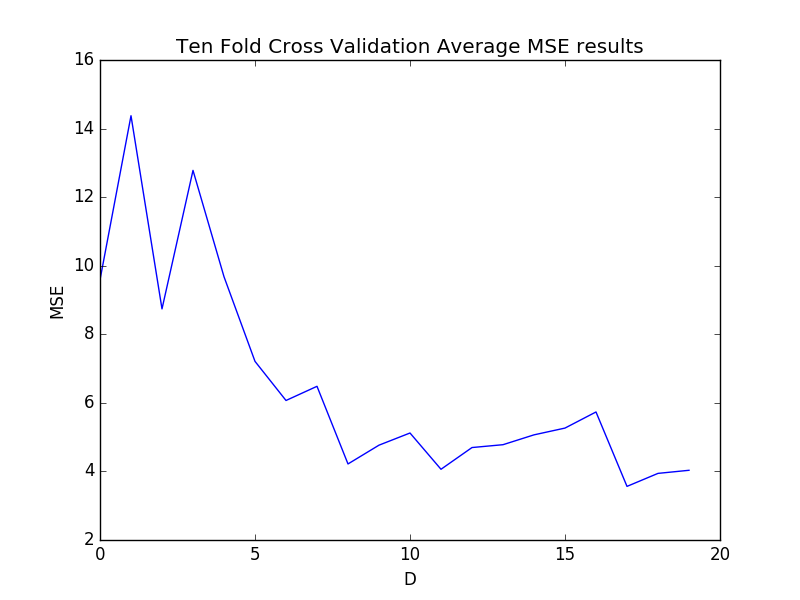




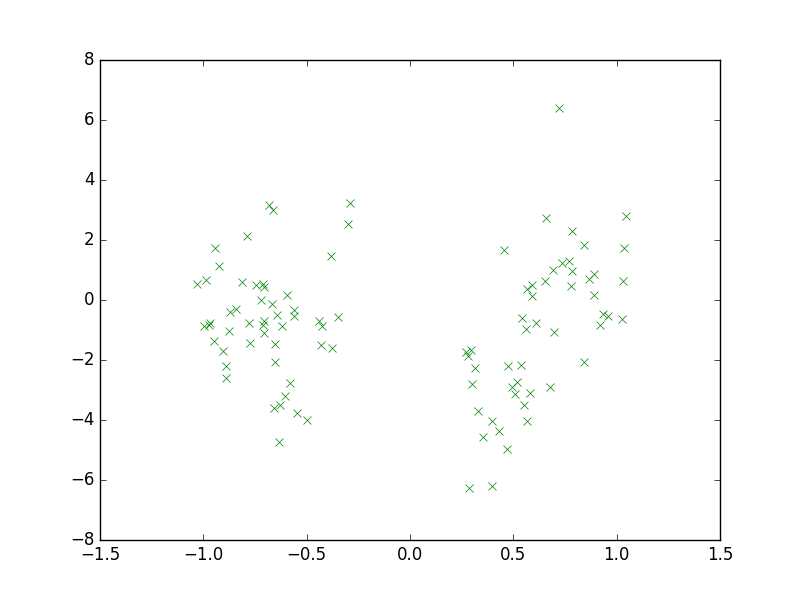




**Step 5**

****

**Step 6**



**Step 7**

