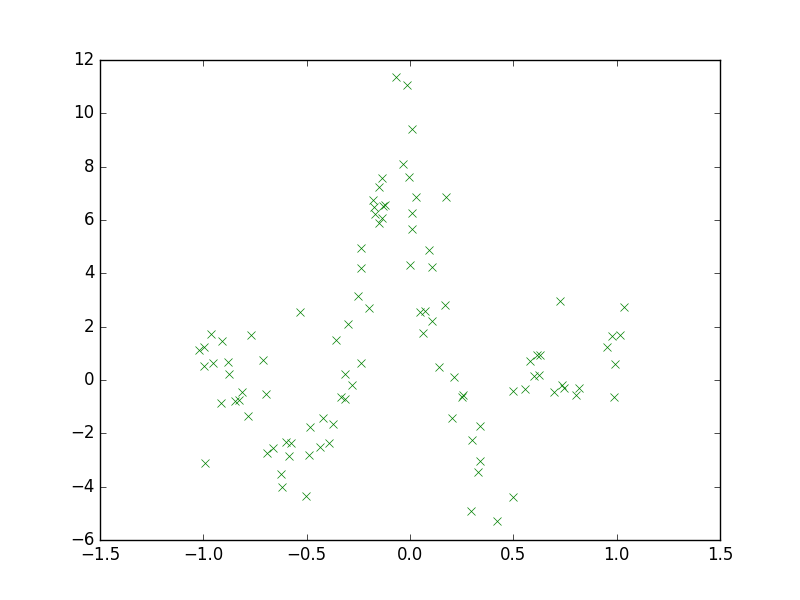
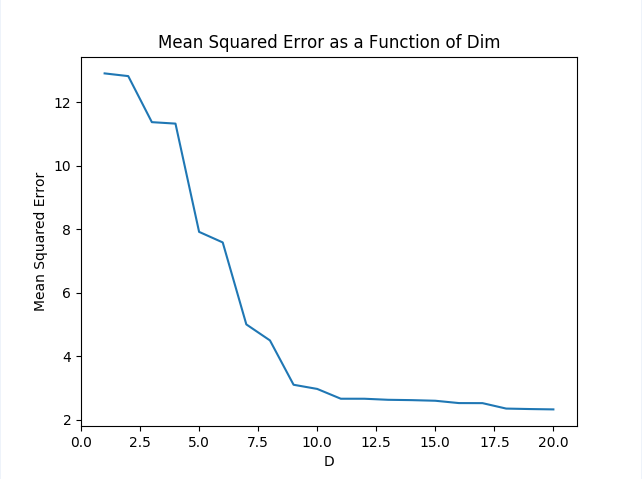
Assignment 2 Report

**Step 1**

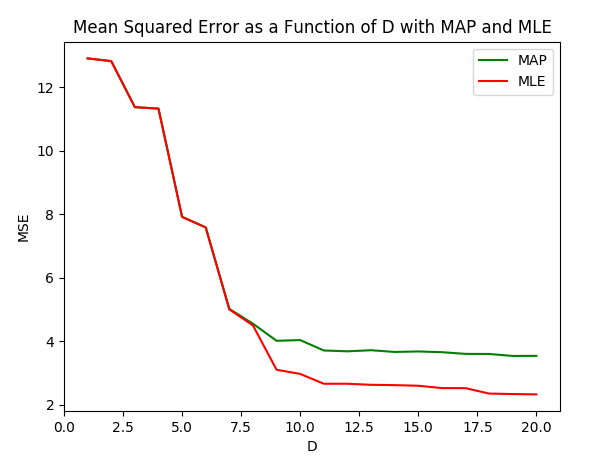


**Step 2**



We can see that the lowest mean squared error occurs when Dim = 20. This makes sense as by consequence of using such a high dimension, the curve we use to fit the data is the most accurate. However, it should be noted that using such a high number for out polynomial basis functions may cause issue as the curve may be over fitting. As discussed in class, it would be better to use simpler basis functions to avoid this. I recommend polynomial basis functions whose degree is 11. We can see that at beyond that point, as D changes, there is little change to the curve. As we want to avoid over fitting – which could occur by selecting a D that is too large, I choose 11.

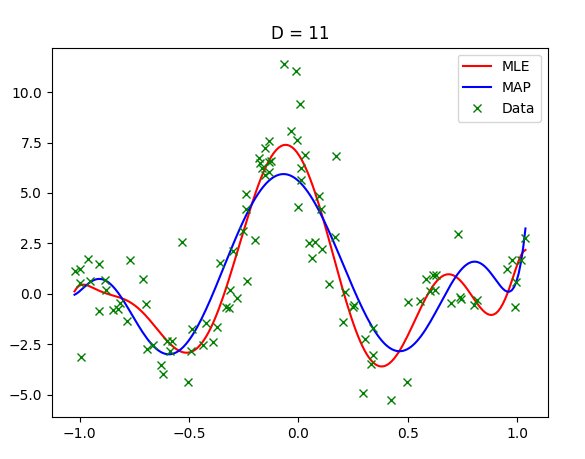
**Step 3**



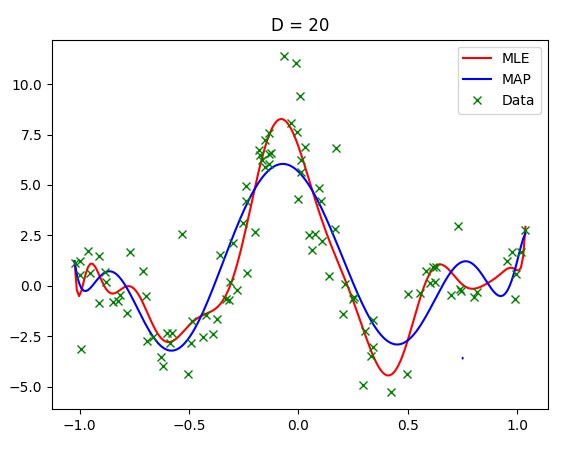
We can see that the mean squared error curve eventually diverges away from the MLE curve at approximately D = 6.5. This is because the purpose of a MAP estimate is to add a prior so as to ensure for smoother models. As D increases, the prior begins to have a greater effect on the quantities we want to estimate. Since the prior begins to contribute more of an effect on what we want to estimate, our mean squared error begins to increases. This is consistent with our findings.

**Step 4**

I would say D = 11 would provide for the best basis function for the MAP estimate. At this point, the data is fit very well and after looking at graphs for higher assignments to D, only minor improvements to the curve are observed (minor improvements with respect to how well the curve fits the data). As we want to fit the data well but not over fit the data, I believe that selecting D to be 11 is the best choice.

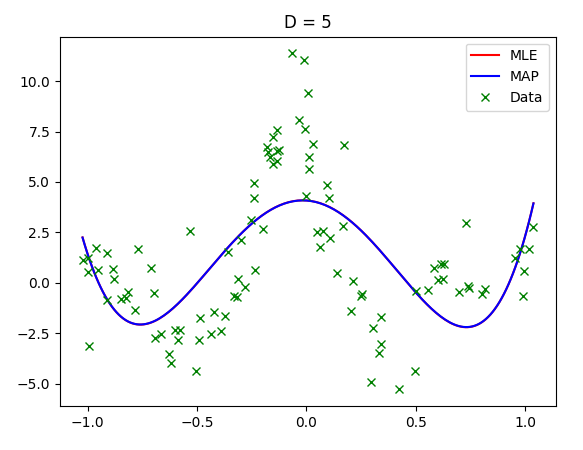


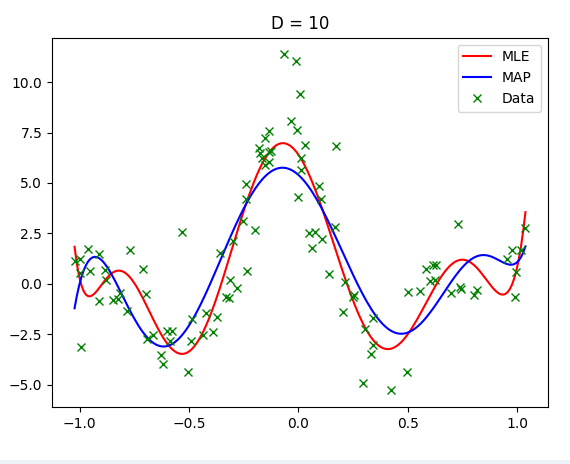
In step 2. I found that the Mean squared error was at its smallest with the Dimension D for the polynomial basis was assigned a value of 20. From the graph, we can see that the parameters that were chosen for both the MLE and the MAP as such that the function fits that data very well. However, it should be noted that over fitting is a possibility.

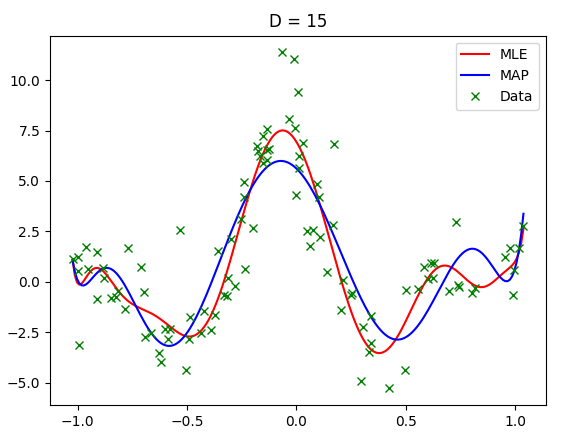


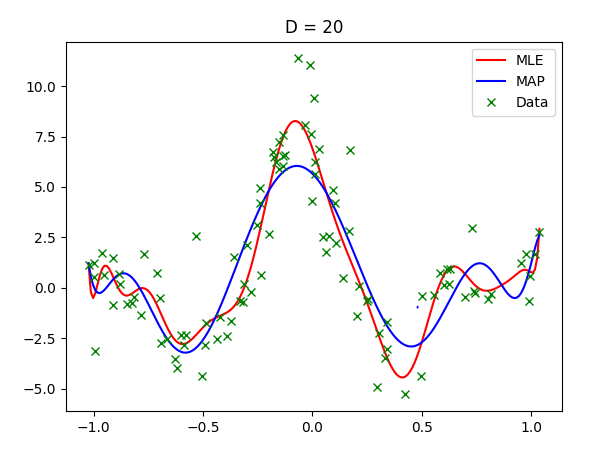
The differences between the MAP and ML estimated functions as D increases can be seen clearly from the graphs that we generate. By virtue of the fact the MAP estimates have contributions from a prior probability, the error is greater and the model becomes less complex. This is good as with less complex models, we are less likely to over fit.

**Below are plots for D = 5, 10, 15, and 20:**

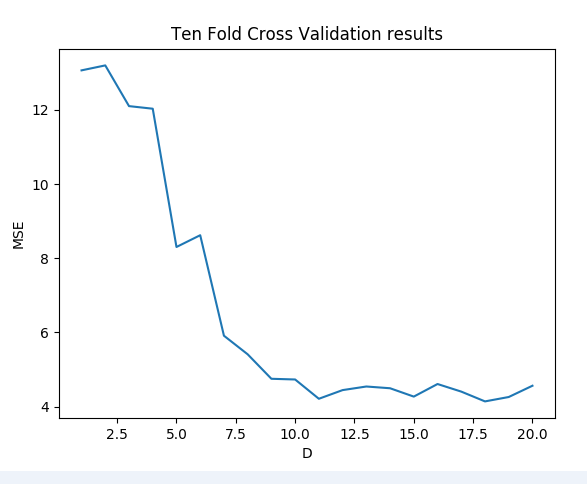




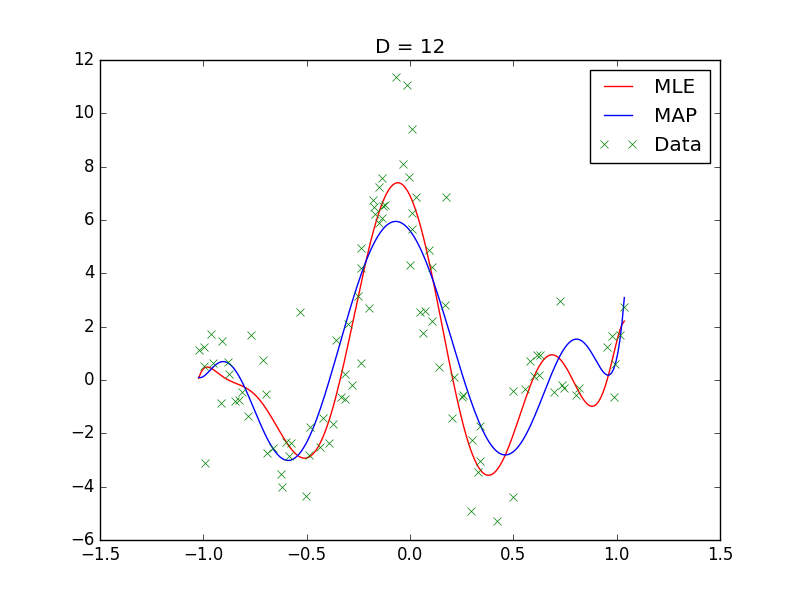




**Step 5**



Based on the curve I would pick D of about 11, 12 or 13. As we can see from the graph, the error is low and larger dimensions would only improve the error by an insignificant amount. We would like to shy away from choosing a larger D since we want to avoid the risk of over fitting.

****

**Step 6**

