Report Project 1:
Rutvij Dhotey
Pattern Recognition and Machine Learning

The Curve Fitting Problem:

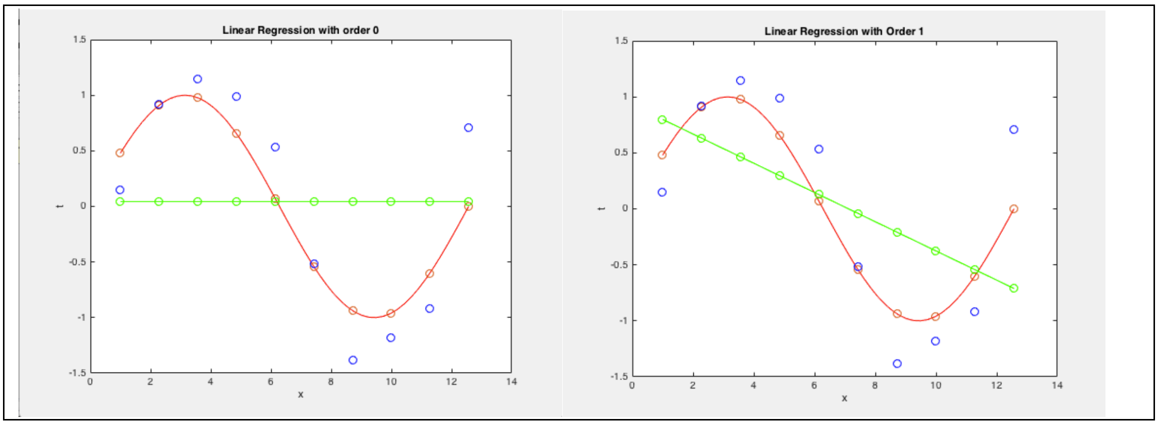
PART 1:

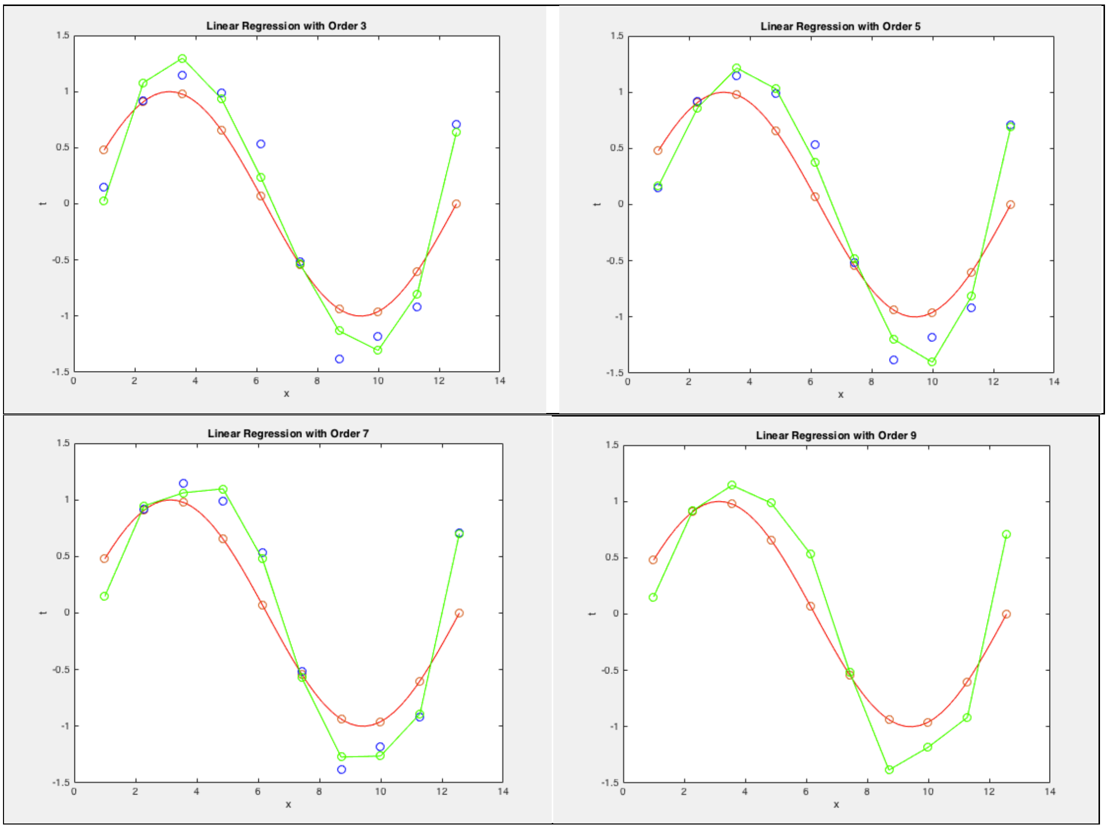
1. We first generate the training data points and Training Outputs by using the generateData.m Matlab file. These 10 points are randomised by adding noise .
2. Initially we solve the problem of linear regression by the technique of error minimisation. The minimising of sum of squares error function will give us a straight forward vector formula as calculated in the first homework.
   The Formula can be stated as: T is the training input and T is the training output.
   $$\mathbf{W^{*}= (X^{T}X)^{-1} * X^{T} * T}$$
3. Thus here W* is the coefficient parameters that we calculate to train wrt. the input / training data.
4. As we have to calculate the output y(X,W*)

   $$\mathbf{y(X,W^{*})}= \sum_{i=0}^{M} \boldsymbol{W0 + W1 * Xi + W2 * Xi^{2} \ldots\ldots WM * Xi\text{^}M}$$

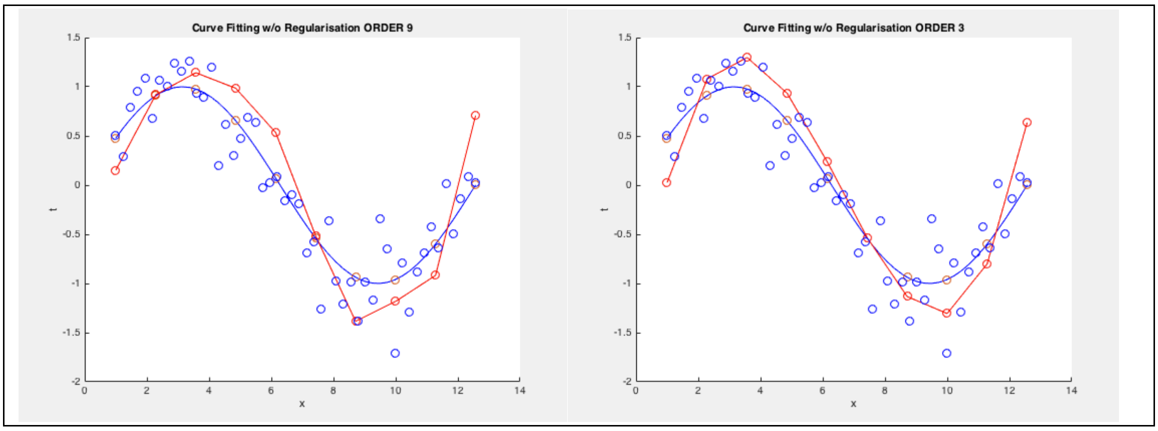
   By using y(X,W*) I have plotted the following figures T vs X:

   As we can see clearly , the training improves the curve fit. Thus as we go higher in order the curve tries to fit to the training data provided. But as we go to the order of 8 and 9 we can clearly make the conclusion that this is a case of OVER-FITTING. : Figures of the Orders 0,1,3,5,7,9 are attached below :





When we plot the curve of the testing input we can observe that the curve with the order 3 has a much better fit rather than the curve of order 9 as order 9 is working perfectly only for the input training provided and not a more generalized approach. Thus the error goes to almost 0 for the 9th Order but increases significantly when plotted for testing inputs which are more than the 10 training data points provided. Figures x and Y.



Thus we can finally conclude that the order to be used for training the data can be between 3-6 with 3 being close to the ground truth curve.

The Table of Values of W* for different Orders is shown Below.

	M=0	M=1	M=3	M=9
W0	0.044	0.9292	-1.5319	5.4088
W1		-0.1305	1.9271	-13.6146
W2			-0.3884	13.0246
W3			0.0198	-5.993
W4				1.539
W5				-0.2312
W6				0.02
W7				-0.0009
W8				0
W9				0

PART 2:

1. In the 1st Part we solved the problem of linear regression by the technique of error minimisation on sum of squares error function. The minimising of sum of squares error function coupled with regularisation will give us a much better understanding with respect to over fitting problem that we saw in the figures x and y. With the regularisation term we reduce this problem to some extent.

   The Vector X used in the example has elements of the order of 10^9.Thus regularisation term that will be added to the $X^T X$ is not going to make that big a difference for the higher ordered terms. Thus I have plotted with respect to the 7th order and how the Regularisation term when added tries to shift my graph more towards te ground truth.

   The Formula can be stated as: T is the training input and T is the training output. I is the eye matrix of the dimensions m+1,m+1.

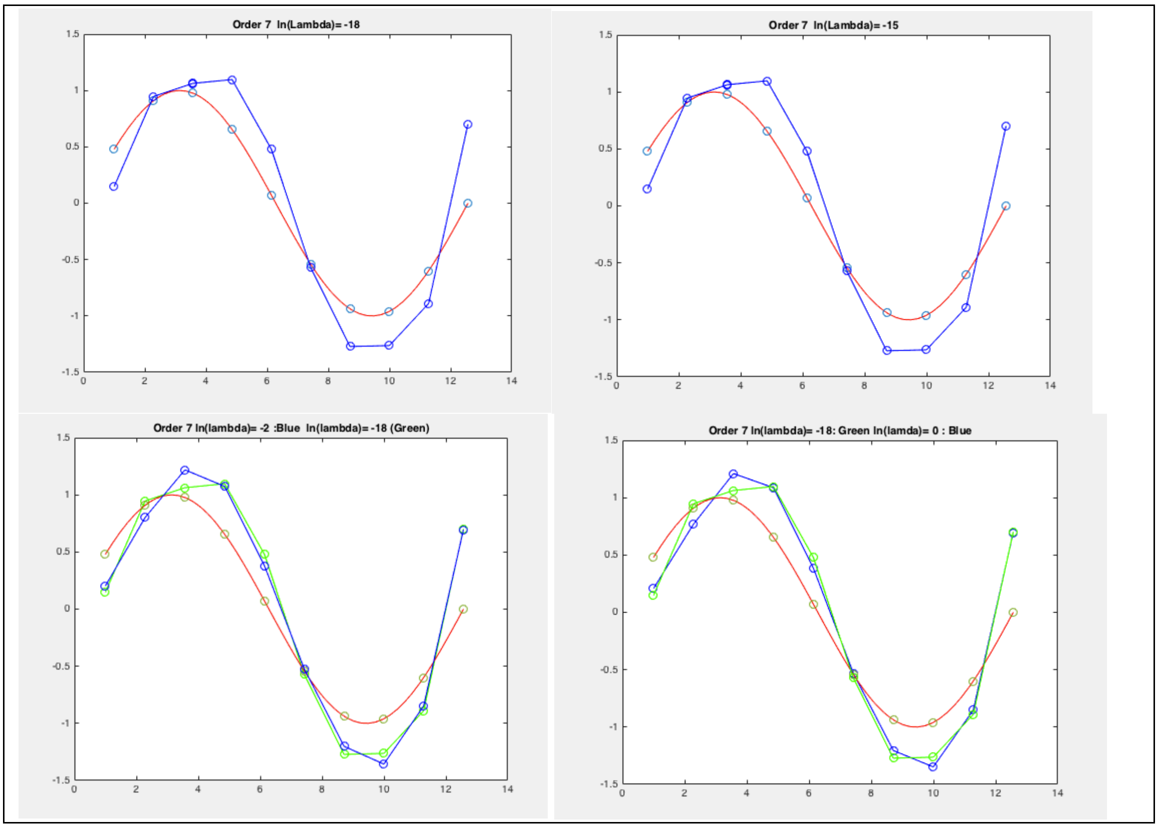
   $$\mathbf{W^{*} = (X^{T}X + \boldsymbol{\lambda} * \boldsymbol{I})^{-1} * X^{T} * T}$$

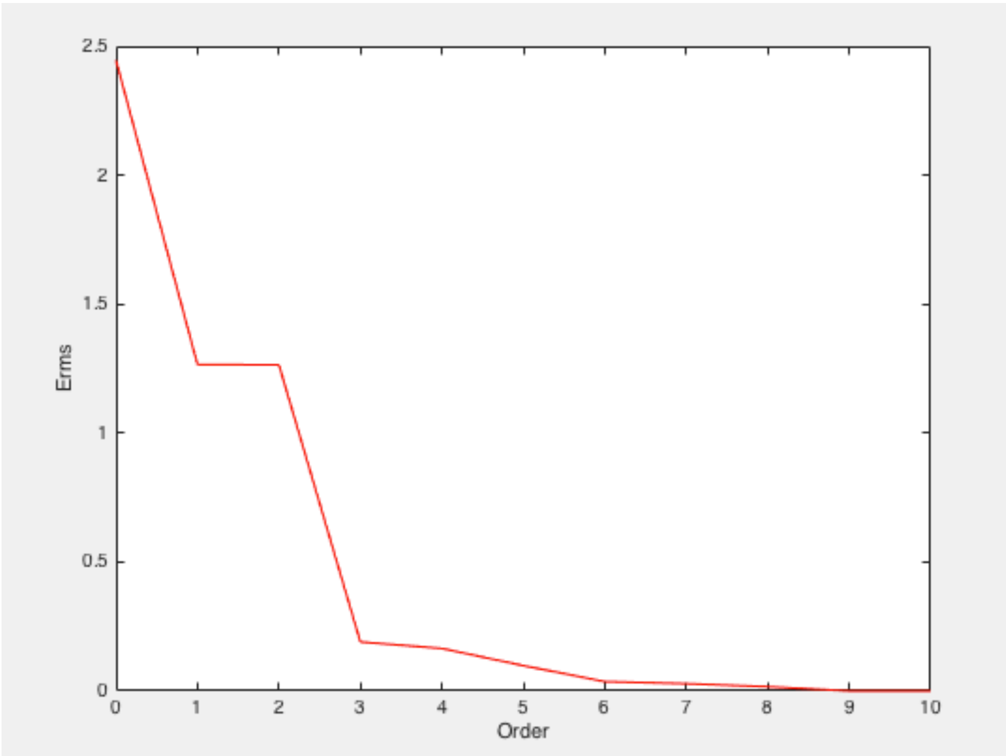
   Thus here W* is the coefficient parameters that we calculate to train wrt. the input / training data with the regularization parameter $\boldsymbol{\lambda}$.
   As we have to calculate the output y(X,W*)

   $$\mathbf{y(X,W^{*}) = \sum_{i=0}^{M} W0 + W1 * Xi + W2 * Xi^{2} \ldots\ldots WM * Xi\hat{\ }M}$$

2. Thus from the following figures its clear that when we increase the value of $\ln(\lambda)$ the curve goes more toward the ground truth. The graphs aren't as visibly different as the book because of the X vector being very large. However if we increase $\ln(\lambda)$ above 0 we do see some significant shift.
   Thus we can use λ to curb the OVER FITTING problem as it regularizes the terms and makes sure that over fitting is avoided.

Figures: Extra Credit Graphs :





PART 3:

Now in Bayesian Probability, we can solve the problem of Curve fitting using the Maximum Likelihood Equations . We can express the uncertainty over the value of the target variable using probability distribution. Using Gaussian Distribution with a mean value of y(x,w) of the polynomial curve fit, we get

$$p(t \mid x, w, \boldsymbol{\beta}) = \mathbb{N}(\boldsymbol{t} \mid \boldsymbol{y}(\boldsymbol{x}, \boldsymbol{w}), \boldsymbol{\beta}^{-1})$$

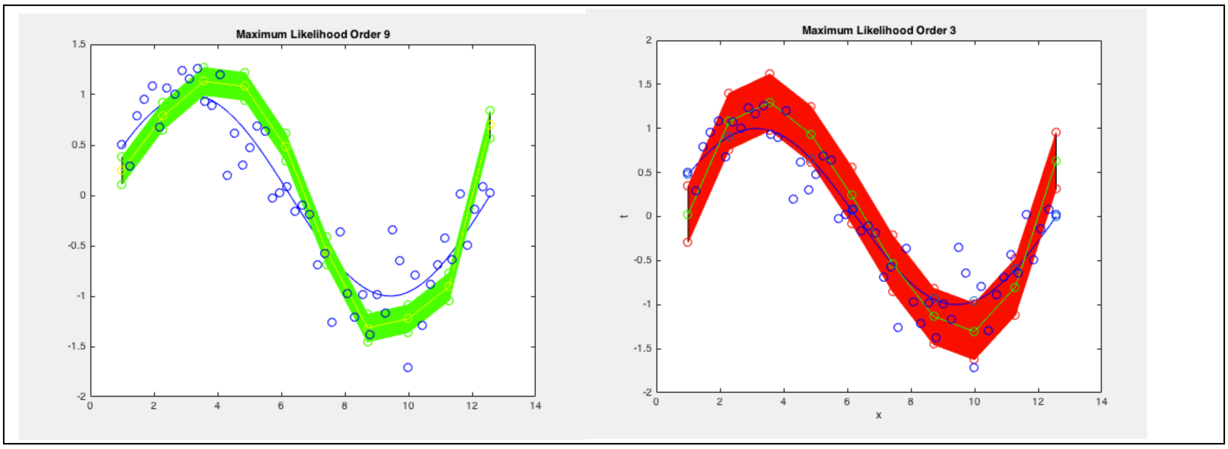
Here the mean is y(x,w) and β^{-1} is the variance of our distribution. So when we further right down the equation we get:

$$p(t \mid x, w, \boldsymbol{\beta}) = \prod_{\boldsymbol{n=1}}^{\boldsymbol{N}} \mathbb{N}(\boldsymbol{tn} \mid \boldsymbol{y}(\boldsymbol{xn}, \boldsymbol{w}), \boldsymbol{\beta}^{-1})$$

After solving this equation so as to maximizing it, thus finding the coefficients of the mean (W^*) we see that this is going to be the same as minimizing the $-_{ve}$ log of the equation and thus simplifying it down to minimizing the sum of squares error we found in Part 1.

After solving for W* we can substitute it back and find the Variance.

Thus these are the graphs that are plotted using the variance.
For the *ORDER 3*.
The points in **blue circles** are the Test Data, the **Green** plot is the mean. And the **BLUE** sinewave is the ground truth.
For the *ORDER 9*.
The points in **blue circles** are the Test Data, the **Yellow** plot is the mean. And the **BLUE** sinewave is the ground truth.



Part 4:

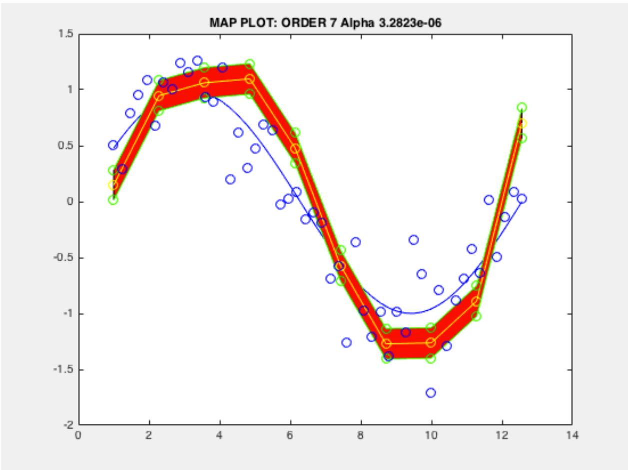
In this we use the equation 1.67 from the book to come up with a formula for optimal W*. We have included alpha in this to regularize the weight values and so that the problem of overfitting can be solved. We get an equation for W* which is only dependent on X, T and Alpha. After Deduction we can prove that :

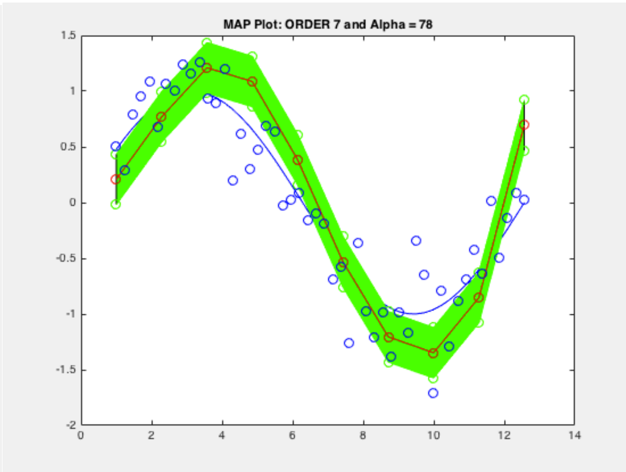
$$\boldsymbol{\lambda} = \boldsymbol{\alpha}/\boldsymbol{\beta}$$

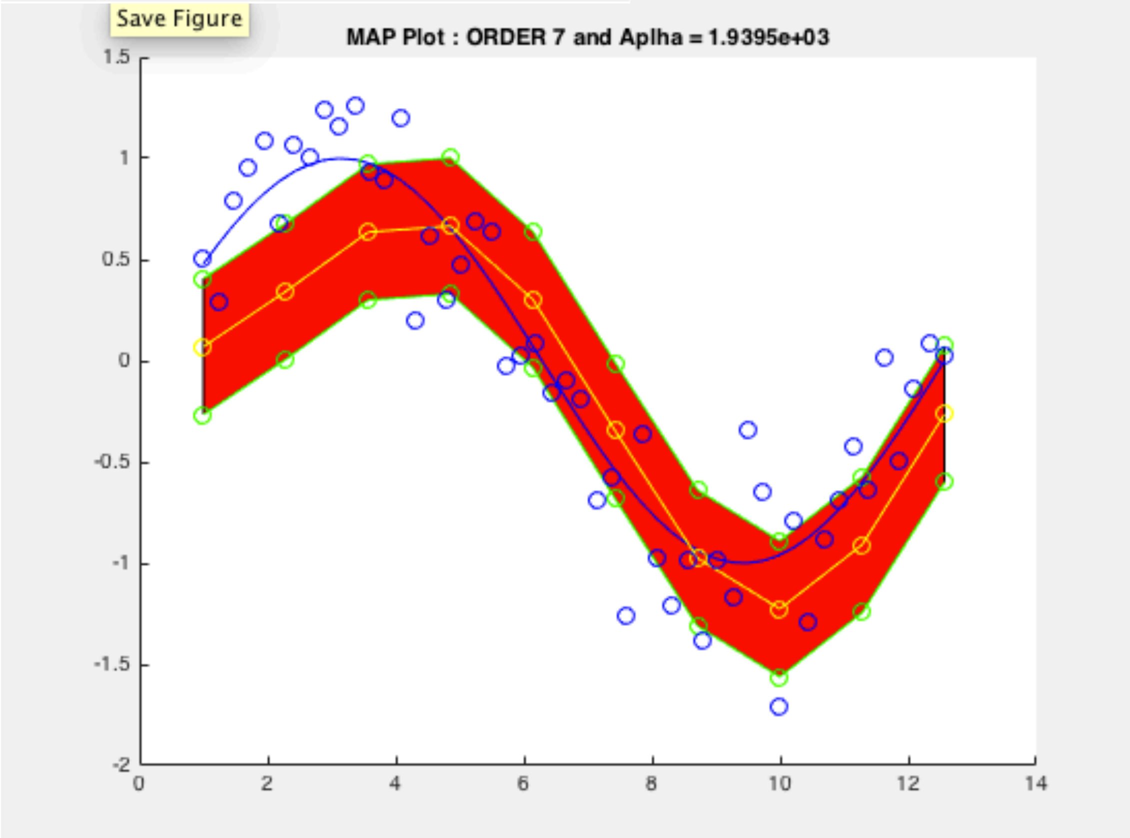
This equation is the one given in the book after Equation 1.67. Wnow we can calculate $\boldsymbol{\beta}$.

And then we calculate Y and standard deviation and plot them for different values of Alpha.
We can clearly see the effect of $\boldsymbol{\alpha}$ on the variance. Thus $\boldsymbol{\alpha}$ is directly related to variance as we find from the equation and prove it with results.

Figures For order 7.







Thus from the figures we can conclude that variance is directly related to Alpha. Thus higher the alpha higher the variance.

Conclusion:

Thus after performing the following experiments, we can derive two main Conclusions

1. The Order of the Polynomial Y decides the curve and higher the order greater is the problem of Over Fitting with respect to the number of training data points. In our observations we find this problem with order 9, 10.

2. The Regularization term tries to curb this problem . But as the values of X in our data are very large for Order 9 or 10. We cant see that much of a difference unless Lambda is increased significantly. Thus we have shown the effect on the Order 7 where we can see the difference.

3. The Variance increases as the alpha increases.