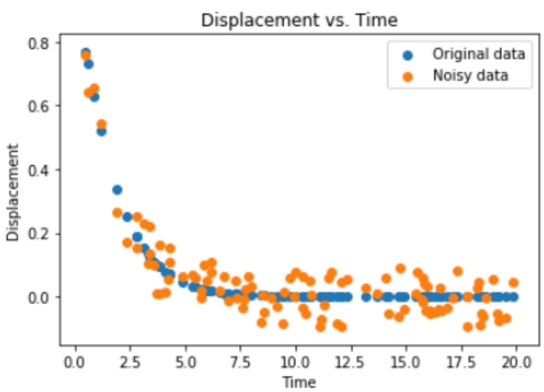
ASSIGNMENT 2

• **Verify your regression program by plotting the predicted values against the analytical solution.**

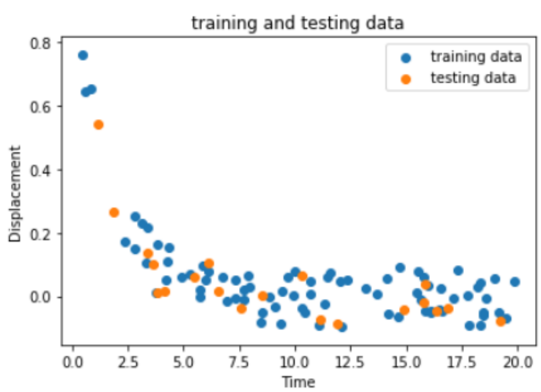
number of random time points=N=100

noise_level = 0.1 =10%

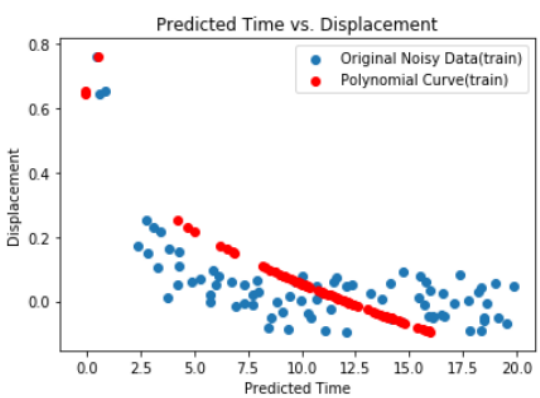
Degree of polynomial=n=2

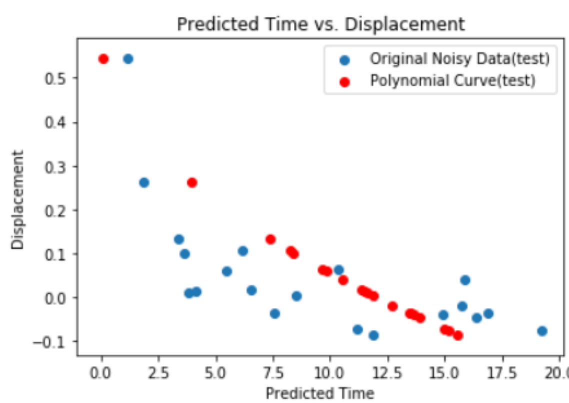


Splitting the data(noisy) into training and testing data in ratio 80:20 and plotting it.



For n=2 ie, 2nd degree polynomial plotting the predicted values against the analytical solution using the regression model.





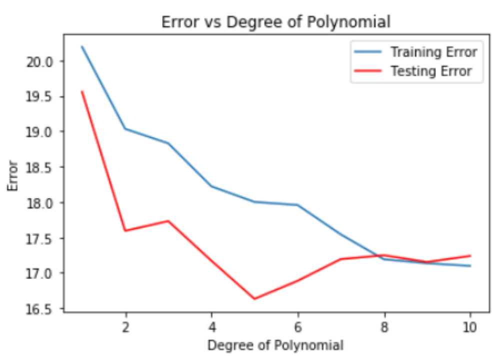
• Plot the training and testing error with respect to the degree of polynomial. From this plot identify the optimal degree of polynomial, underfitting region and over fitting region.

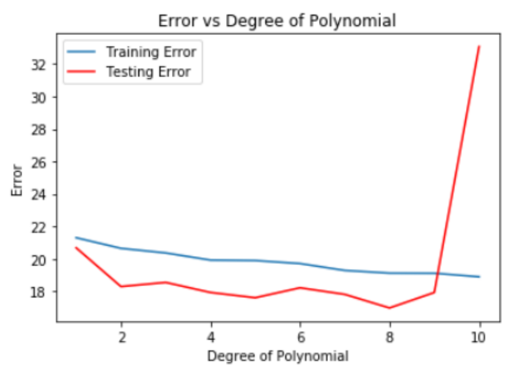
By running the program again and again and getting different data points the few graphs which match or are similar to our theoretical understanding are:

Note:

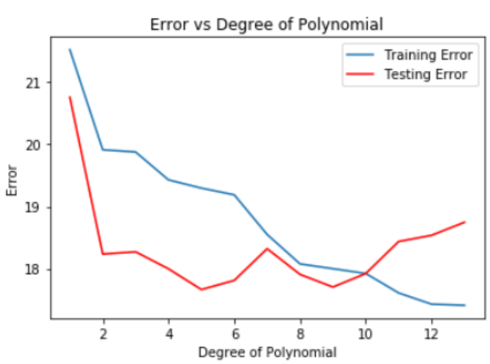
- ➔ we have taken 500 data points.
- ➔ The range of varying the degree of polynomial is 1 to 10
- ➔ noise_level = 0.1 =10%

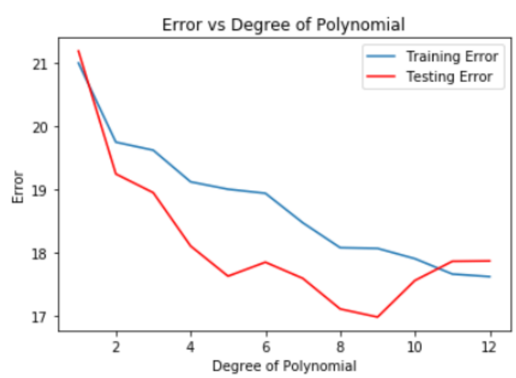
- For the below to graphs : The optimal degree of polynomial=9 ie, where the curve of training and testing error intersects





- If we vary the range of degree of polynomial upto 14, then for some cases higher optimal degree like n=10 or 11 is also recorded



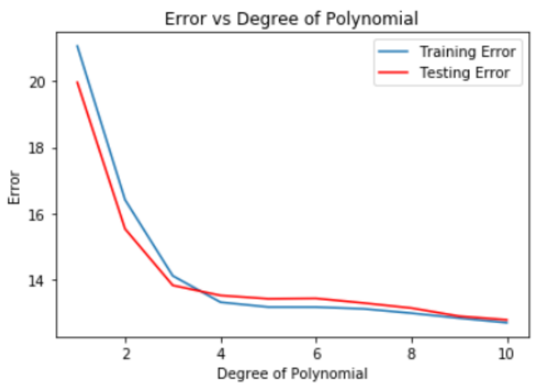


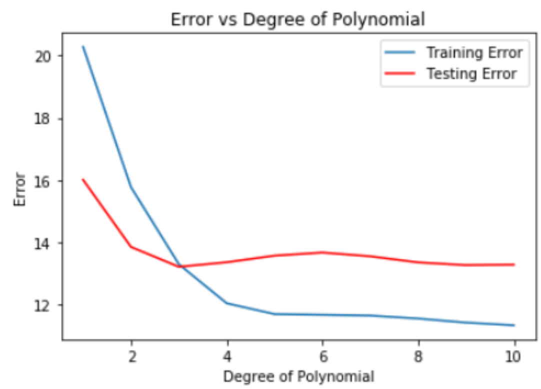
• Carry out the point no. 2 exercise for three different level of noise. Does the noise affects the conclusions of point-2.

Given: the range of noise is from 2 to 10%. For:

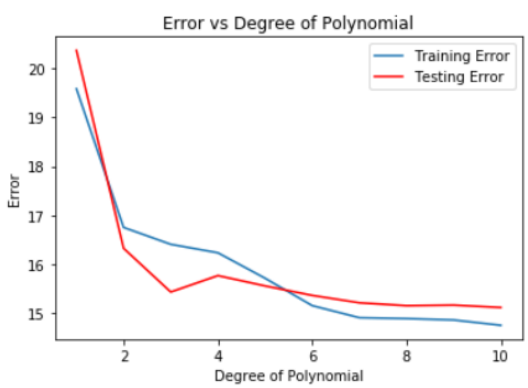
- Noise=0.02

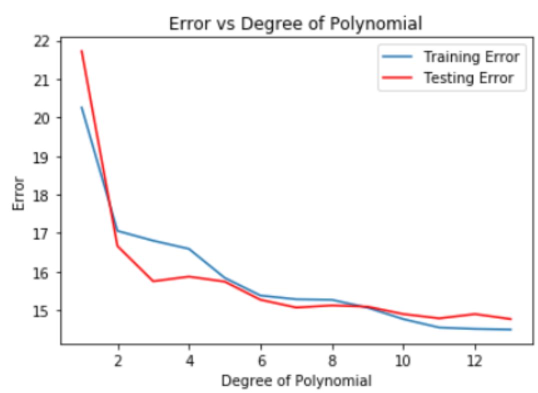
The overall Magnitude of error corresponding to both training and testing error has reduced considerably. Also the optimal value of degree of polynomial has reduced and comes out to be n=3



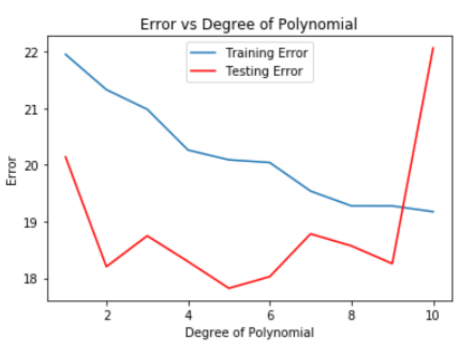


- Noise=0.05. Optimal value of degree of polynomial =5





- Noise=0.1.



- Optimal value of degree of polynomial =9

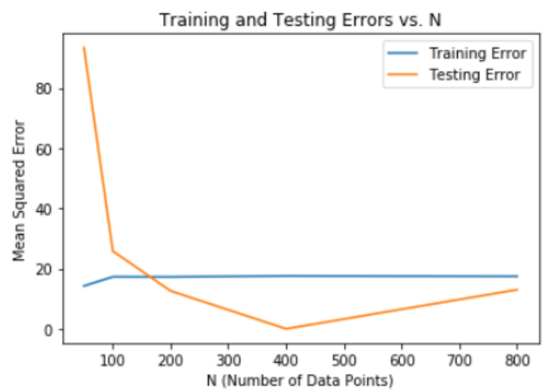
As the noise level increases both the magnitude of error and the optimal value of degree of polynomial increase.

• Take the case of the highest noise, considered in point-3. Further, take the polynomial degree that lies in the over fitting region. Now increase the data size in the geometric progression of two, and plot the training as well as testing error with respect to the data size.

Noise=0.1

Overfitting region=degree of polynomial “n”=11

Data size in geometric progression of 2 with base value of 50; as there should be considerable data size to be able to have training and testing data.



The training error should decrease as the data size increases because more data helps the model to learn better.

Since we are in the overfitting zone the testing error is greater than the training error. As the data points increases firstly the testing error tends to decrease because the model becomes more generalized with more data but after a certain point it starts increasing again because the model also learns noise from the training data due to larger data size and is not able to predict unseen testing data points properly.