CS 6375: Machine Learning
Assignment #1

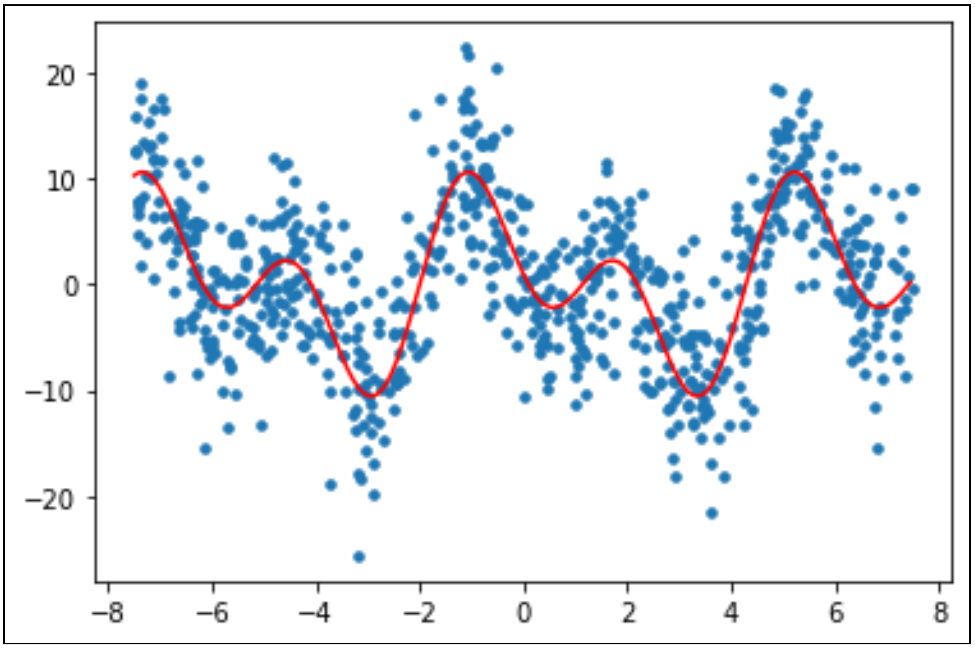
Prof. Anjum Chida

Name: Akash Biswal

NetID: axb200166

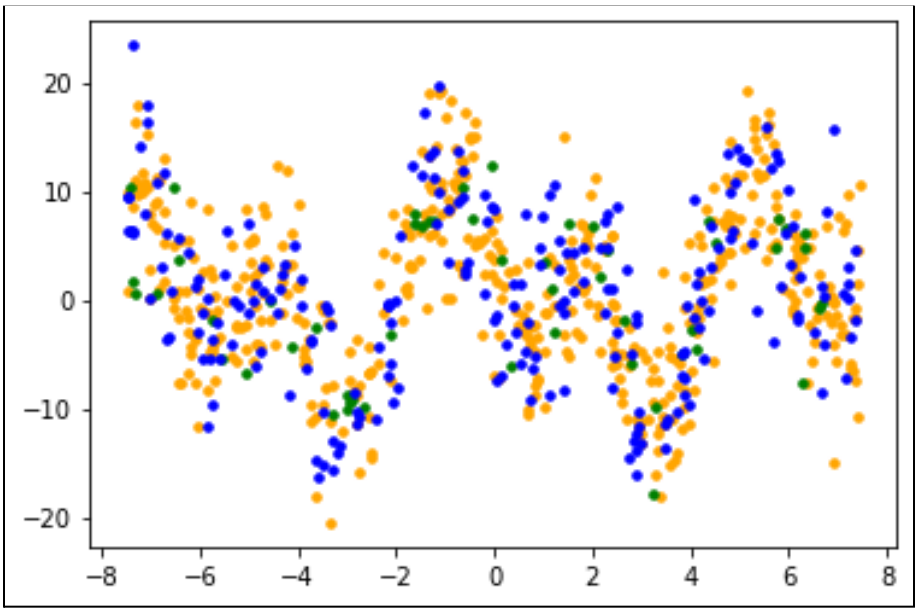
Plot of the generated synthetic data against the true function:

$$y_{\text{true}} = f_{\text{true}}(x) = 6(\sin(x+2) + \sin(2x+4))$$



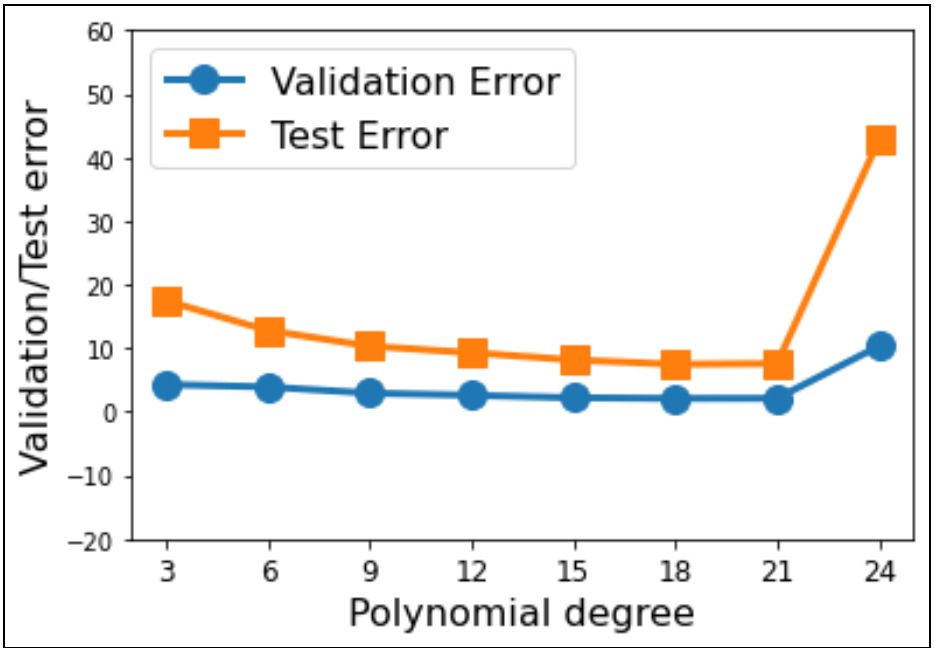
Data as visualized after splitting into Train, Validation and Test sets in a Scatter plot:

- Orange - Train set(60%)
- Green - Validation set(10%)
- Blue - Test set(30%)

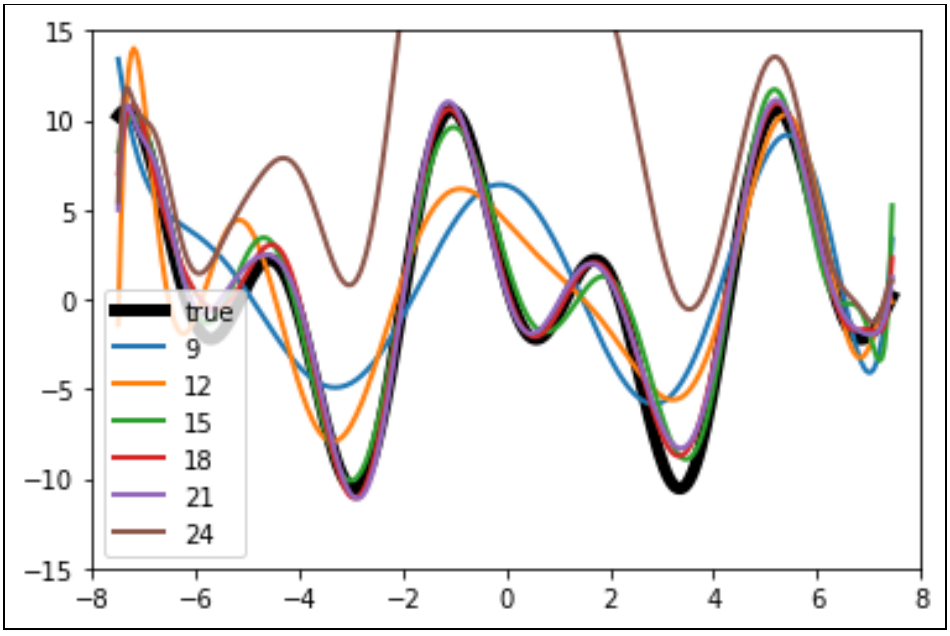


Problem 1:

d) Plot of validation error v.s dimension/degree

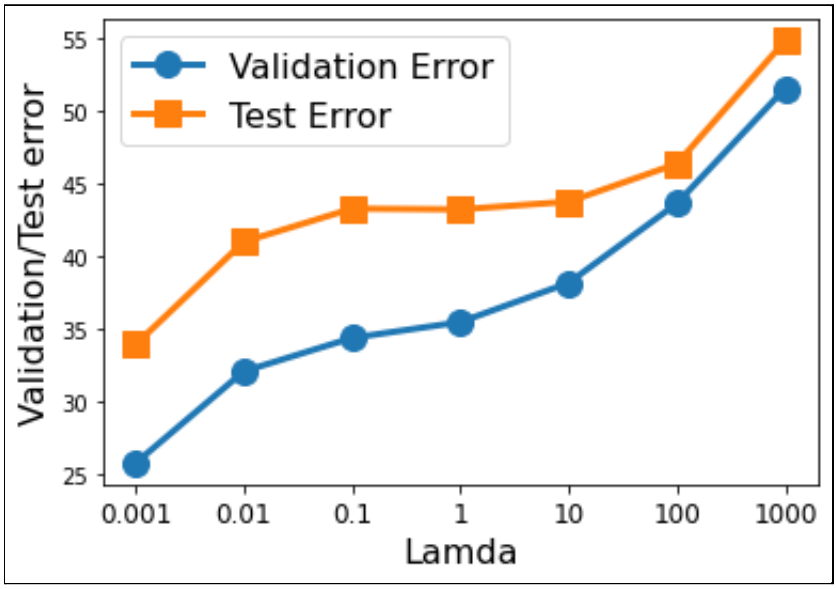


Based on the above image a polynomial of degree between 18-21 would best generalize the function. This can also be seen in the visualization of all learned models as curves 18 and 21 best fit/approximate the true function.



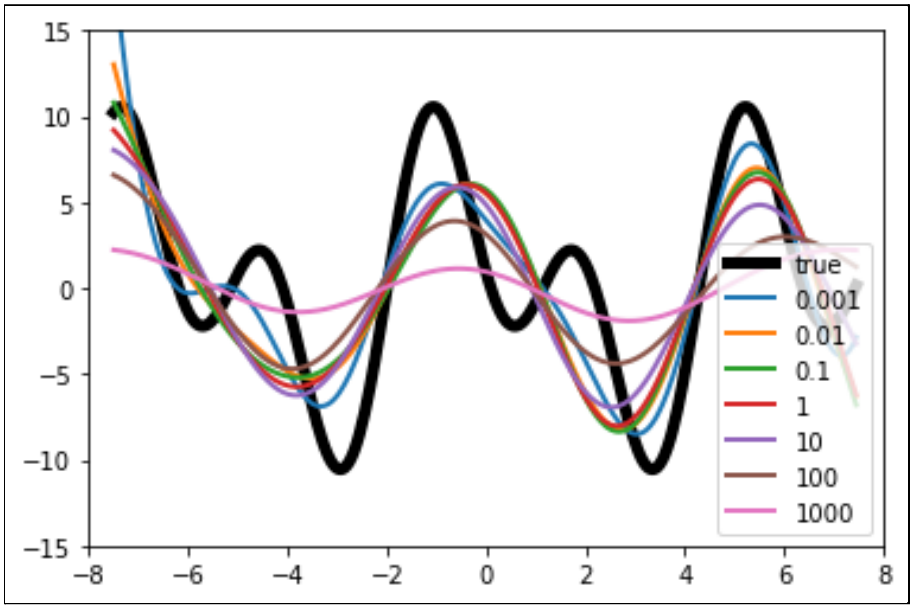
Problem 2:

c) Plot of validation/test error v.s regularization factor lamda.



From the above plot it can be concluded that the ideal values of λ are the small values 0.001-0.1 as the observed test and validation error are smaller than when the value of λ is higher.

d) Plots of the learned models as well as the true model



The linearity of the model can be said to be increasing with increase in **λ**, this is because the curves tend to become more like lines as **λ** increases as seen above. This is happening due to an increase in error. In contrast the lower values of **λ** best approximate the true curve.