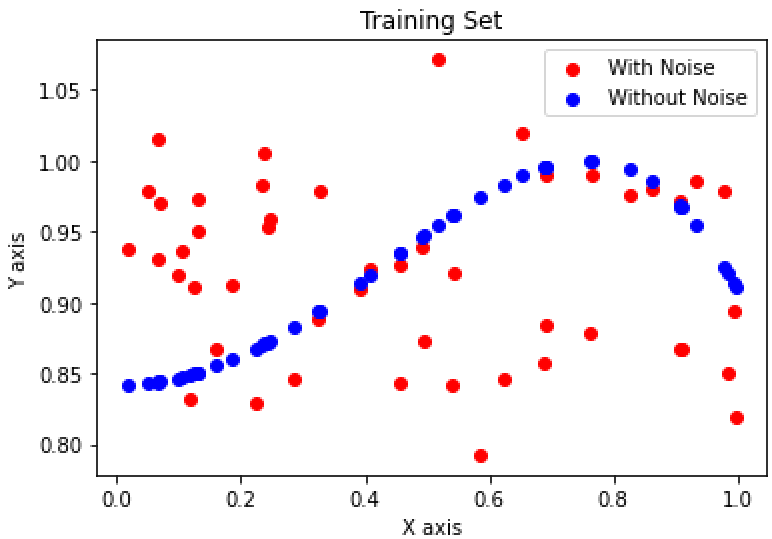
ASSIGNMENT 3

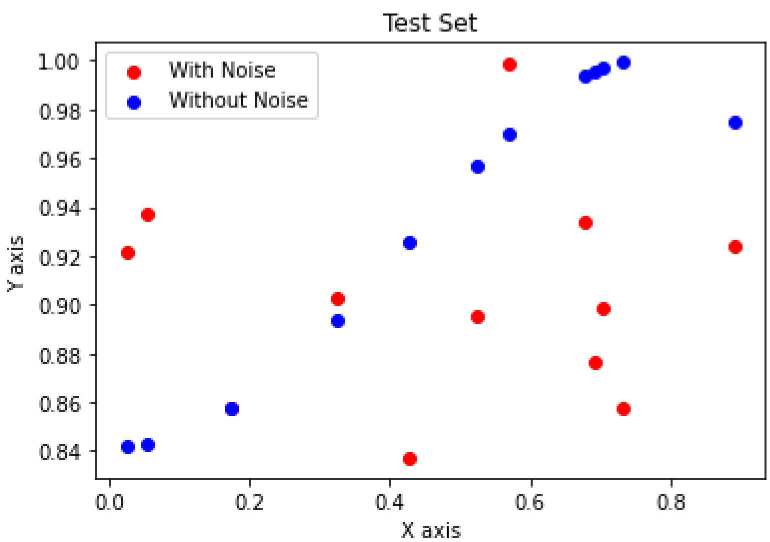
1. Generate 50 data points (one-dimensional problem) with x uniform distance from (0, 1) and generate y from the formula y=sin (1+x^2)+noise Where noise has (0, 0.032) distribution. Divide the data into training and test (8:2). Scatter plot the y values as well as plot the noise-free true function for train and test separately.

Implement the linear and polynomial regression models using the closed form solution as taught in the class on training data and explore their effectiveness in creating a predictor that best approximates the data using test data. Calculate the RMSE error and show the test and train results with varying degrees of a polynomial regression model to fit the data.

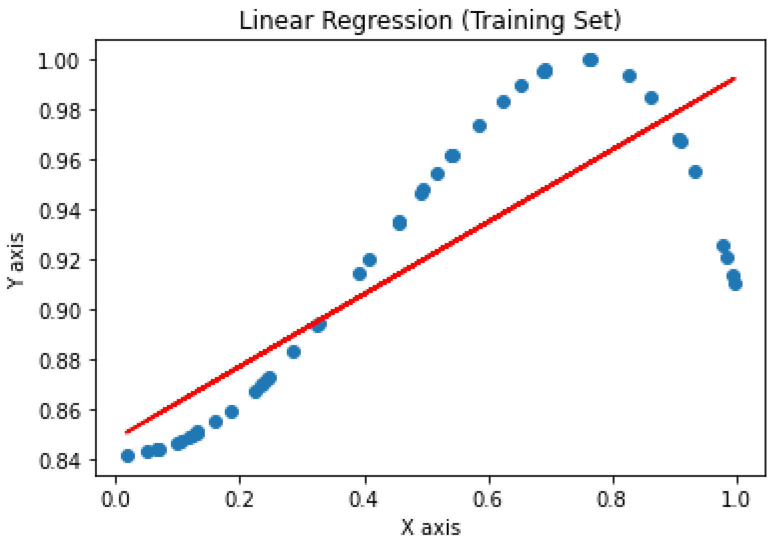
Soln:

The training and test data is generated and shown below –

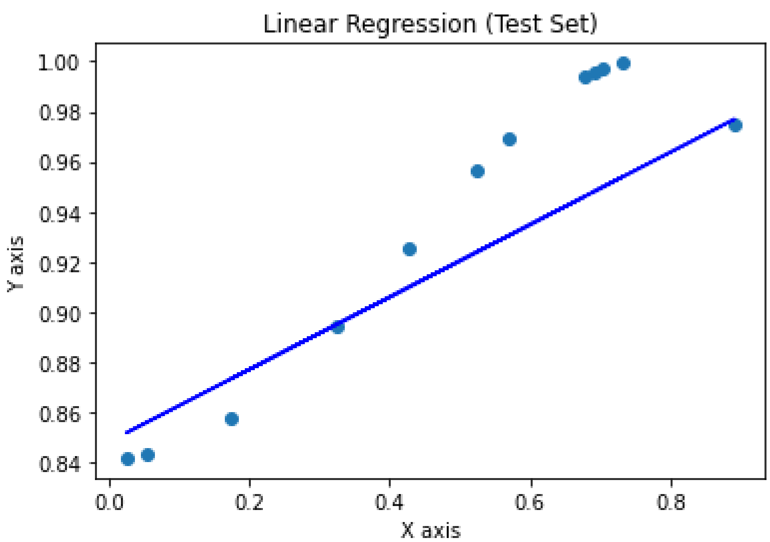




After applying linear regression, the output for the training and test dataset comes as follows:



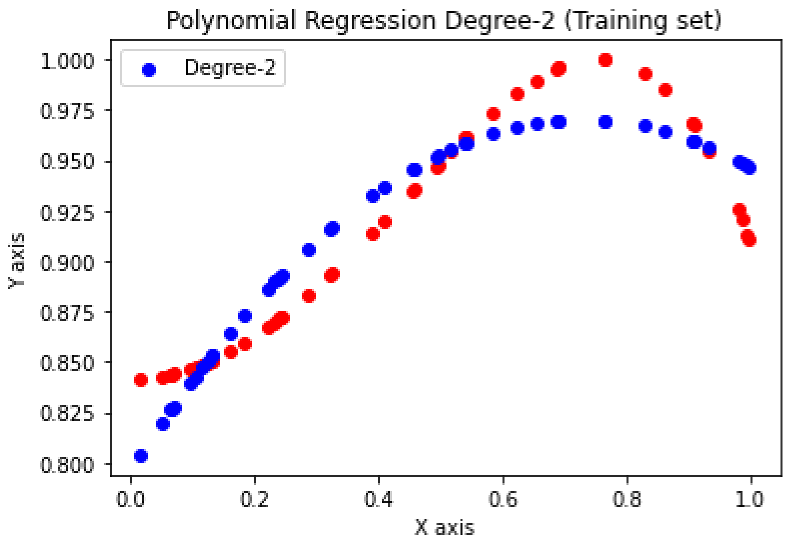
RMS error on training data = 0.03222560451258129



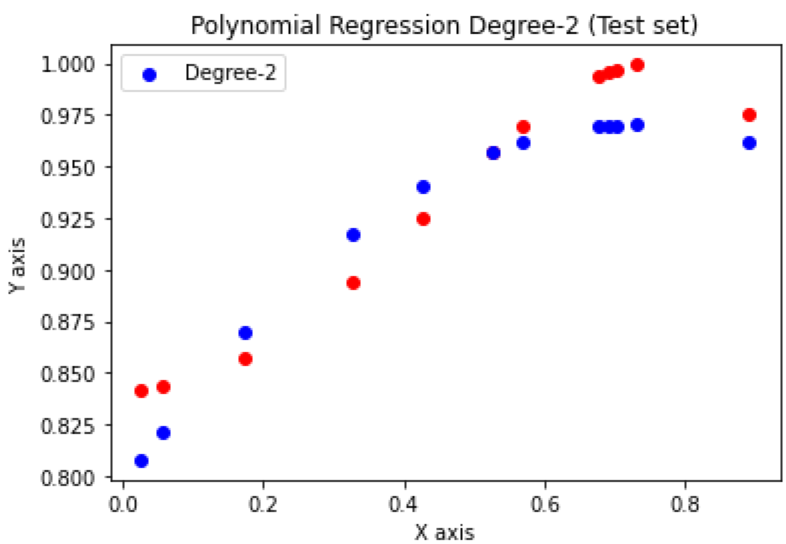
RMS error on test data = 0.031961700587547333

<u>Polynomial Regression</u>:

<u>Degree-2</u>:

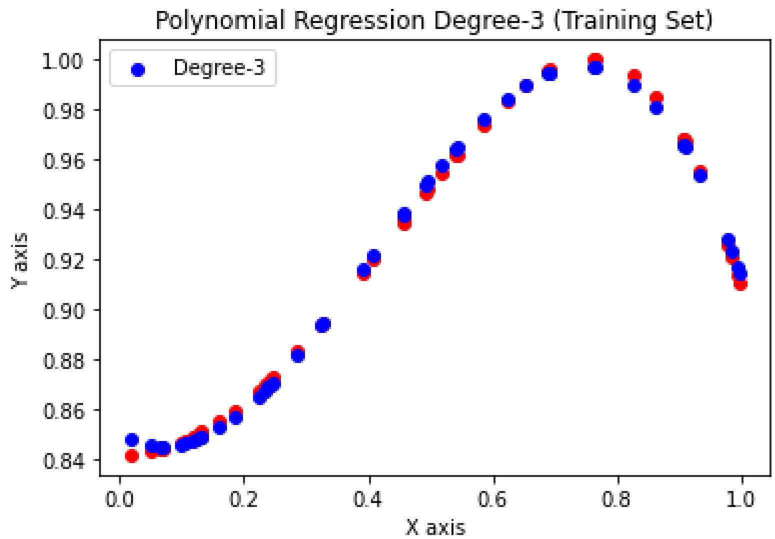


RMS error on training data = 0.019122887731053265

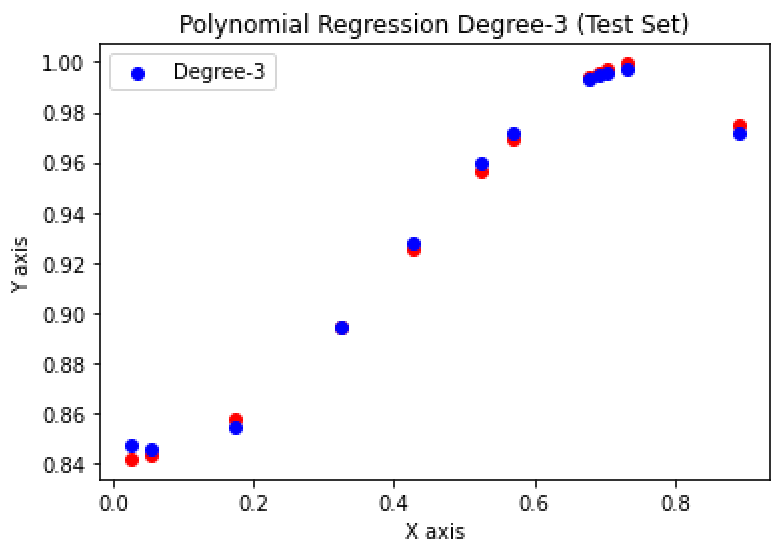


RMS error on test data = 0.021595484696380413

Degree-3:

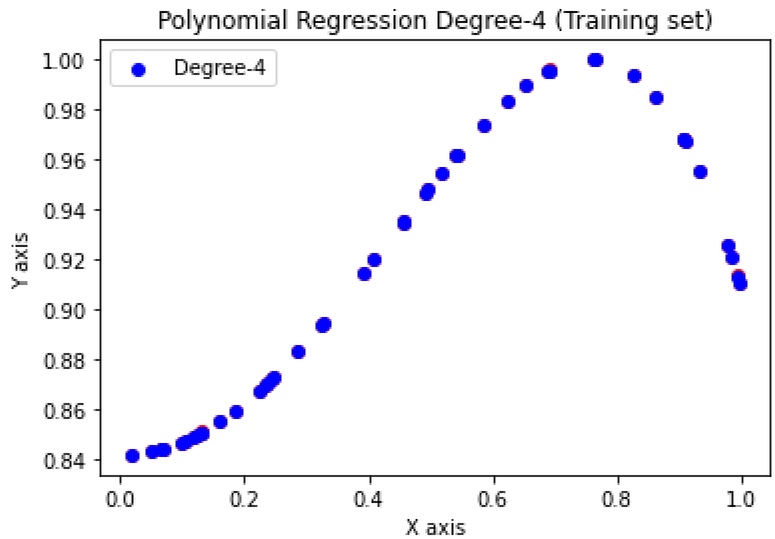


RMS error on training data = 0.0024947664669286063

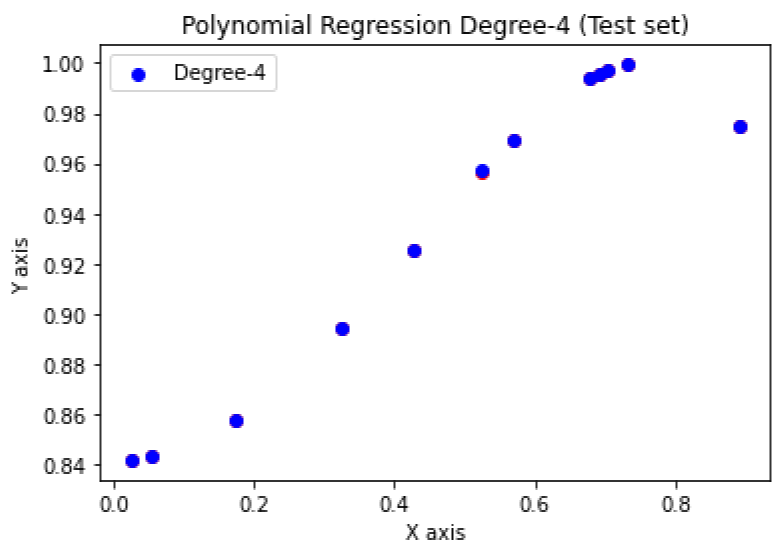


RMS error on test data = 0.002598245191685055 7

Degree-4:



RMS error on training data = 0.0001263749933763009



RMS error on test data = 0.00014398135013720108

Inference:

Root mean square error for:

Polynomial(degree-4) < Polynomial(degree-3) < Polynomial(degree-2) < Linear

We can see that the RMS error for polynomial regression model with degree 4 comes out to be the lowest among all. So, it would be the most effective for the given dataset. Also, linear regression is the least effective for the dataset.

Github Repository link: https://github.com/MehulJain-831/ITIT-4103-2021