# Name/Roll No: Himanshu Ruhela 2018IMT-039 Course: IMT Course Code: ITIT - 4107

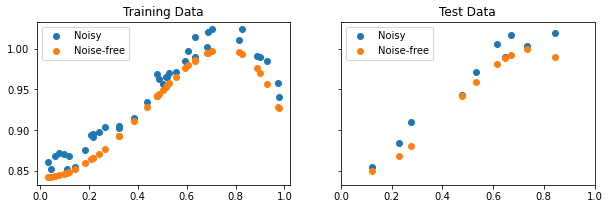
Assignment Number: 3  
Deadline for Submission: 23:59 2th October 2021  
  
**Aim**

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.

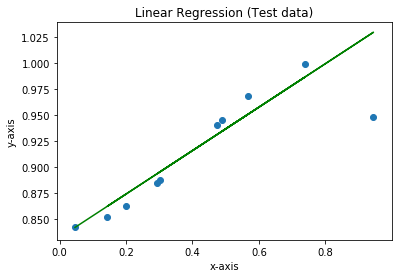
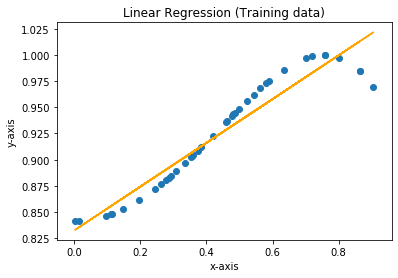
**Procedure/Algorithm**

The training and test data is generated and shown below –



**Results**

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

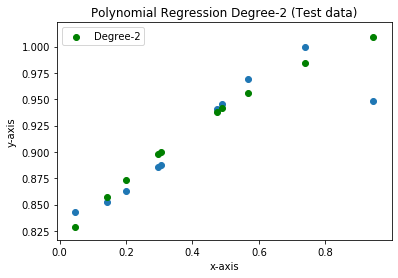
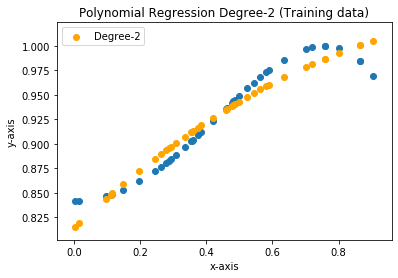


RMS error on training data = 0.014903765515524104

RMS error on test data = 0.027779168186686402

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

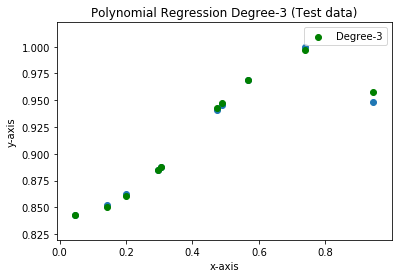
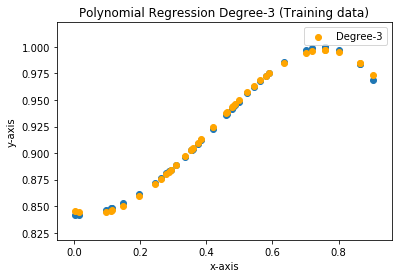
**Degree 2:**



RMS error on training data = 0.012927179019461443

RMS error on test data = 0.021612614529871073

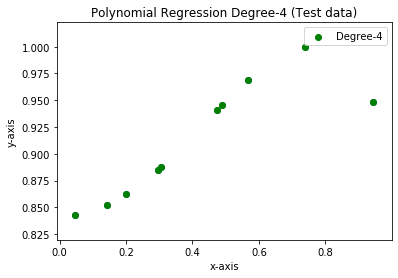
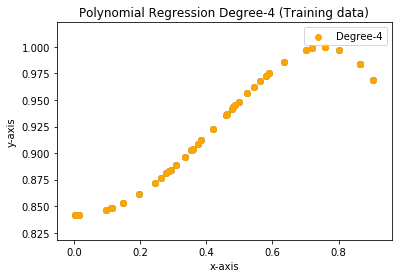
**Degree 3:**



RMS error on training data = 0.0018485269241123336

RMS error on test data = 0.0034585869955217285

**Degree 4:**



RMS error on training data = 2.747726398683098e-05

RMS error on test data = 0.00014384275155573045

Jupyter Notebook:

Repository:

<https://github.com/verdantfire/ITIT-4103-2021/tree/main/Assignment%203>

**Inference**

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.