

ASSIGNMENT 3

Name: Vivek Kamboj

Roll No.: 2018IMT-109

Course: Machine Learning Lab

Course Code: ITIT-4107

Deadline: 02-Oct'21 (11:59 PM)

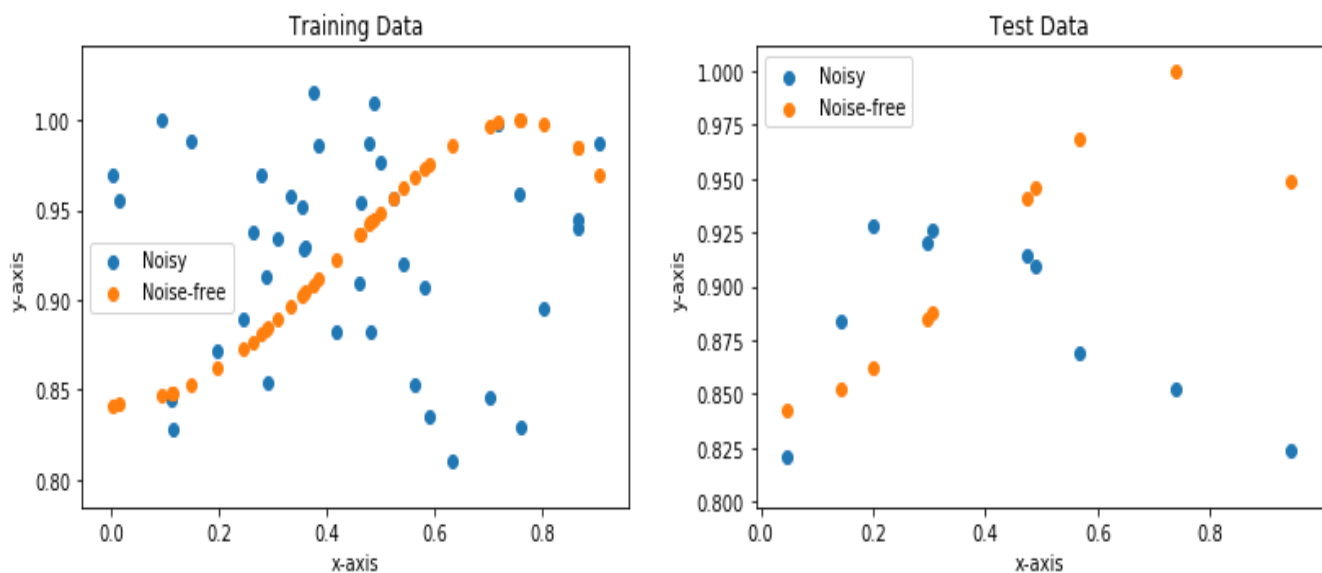
Objective:

1. Generate 50 data points (one-dimensional problem) with x uniform distance from (0, 1) and generate y from the formula $y = \sin(1+x^2) + \text{noise}$ 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.

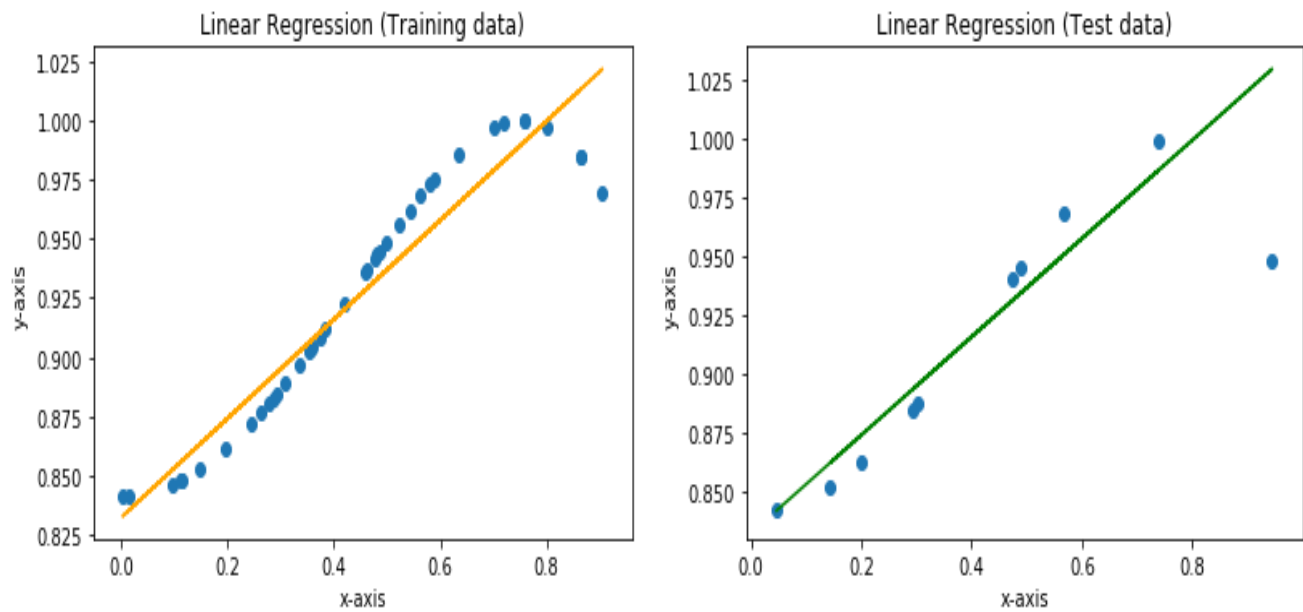
Implement the linear and polynomial regression models using the closed form solution as taught in the class on training data and explore their effectiveness in creating a predictor that best approximates the data using test data. Calculate the RMSE error and show the test and train results with varying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

Ans:

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:

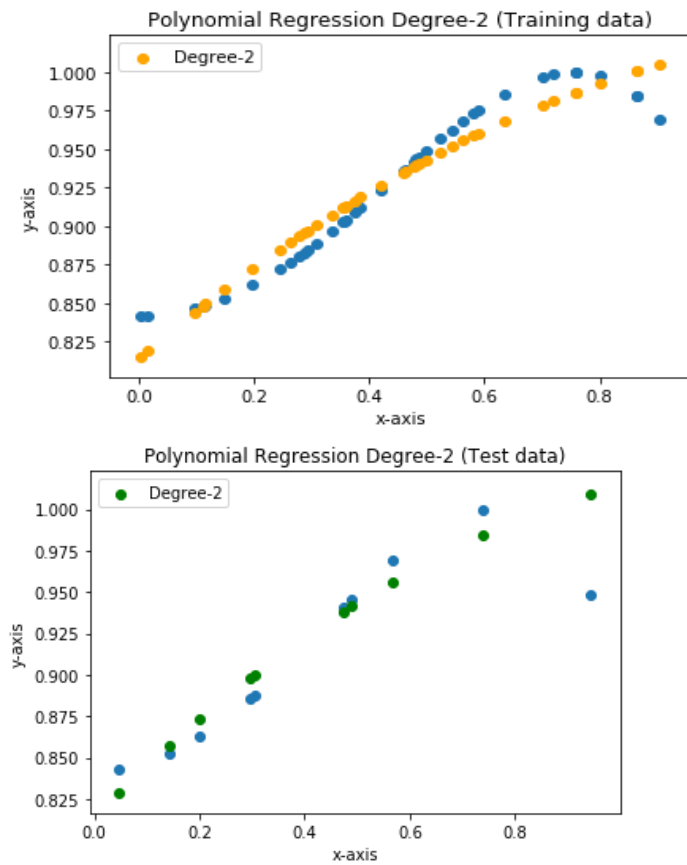


RMS error on training data = 0.014903765515524104

RMS error on test data = 0.027779168186686402

Polynomial Regression:

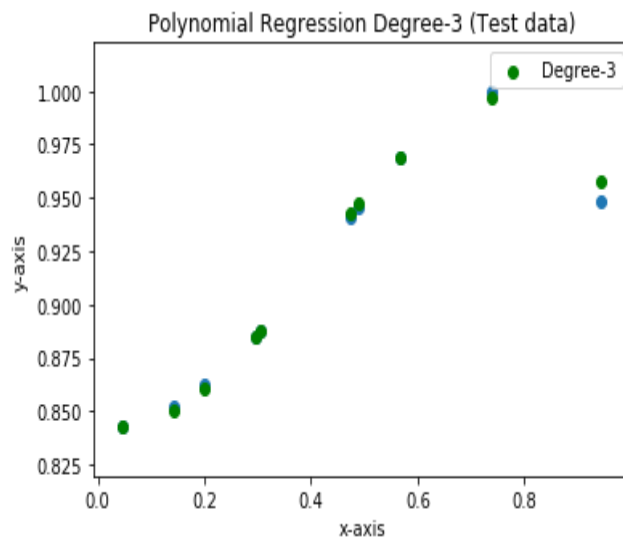
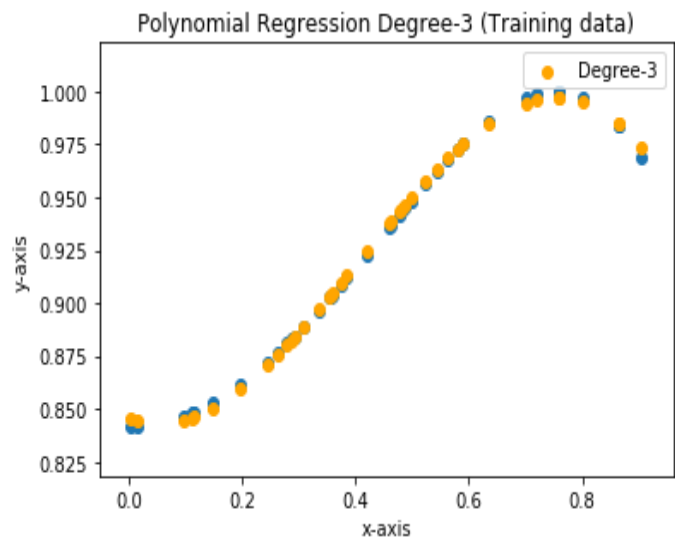
Degree-2:



RMS error on training data = 0.012927179019461443

RMS error on test data = 0.021612614529871073

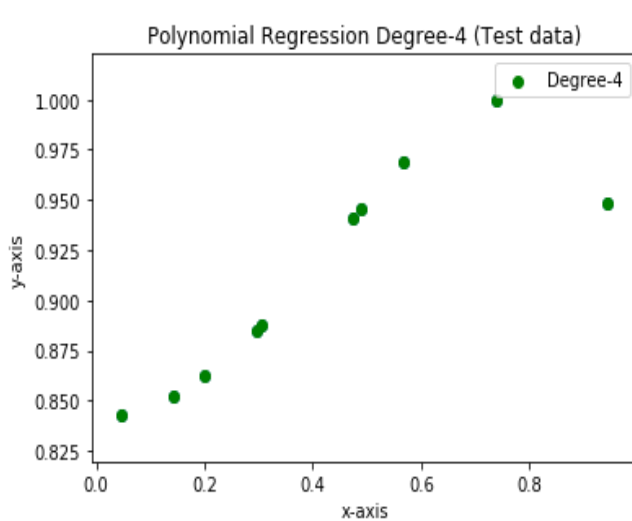
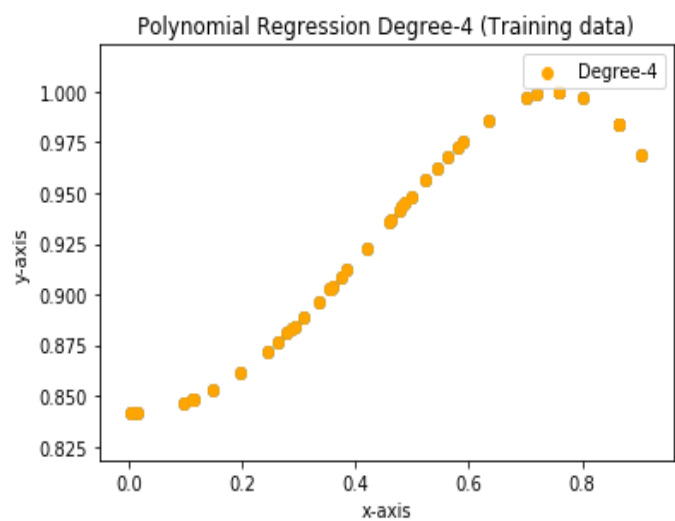
Degree-3:



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

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.

Here is the link to the code for reference:

<https://github.com/Vivek-Kamboj/ITIT-4103-2021/blob/main/Assignment3/Assignment3.ipynb>