

ASSIGNMENT 3

Name : Arsh Pratap

Roll no : 2018-IMT-021

Course : Machine Learning

Course Code : ITIT-4103

Code Link : <https://github.com/arshPratap/ML-Assignment/blob/main/Assignment3.ipynb>

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.

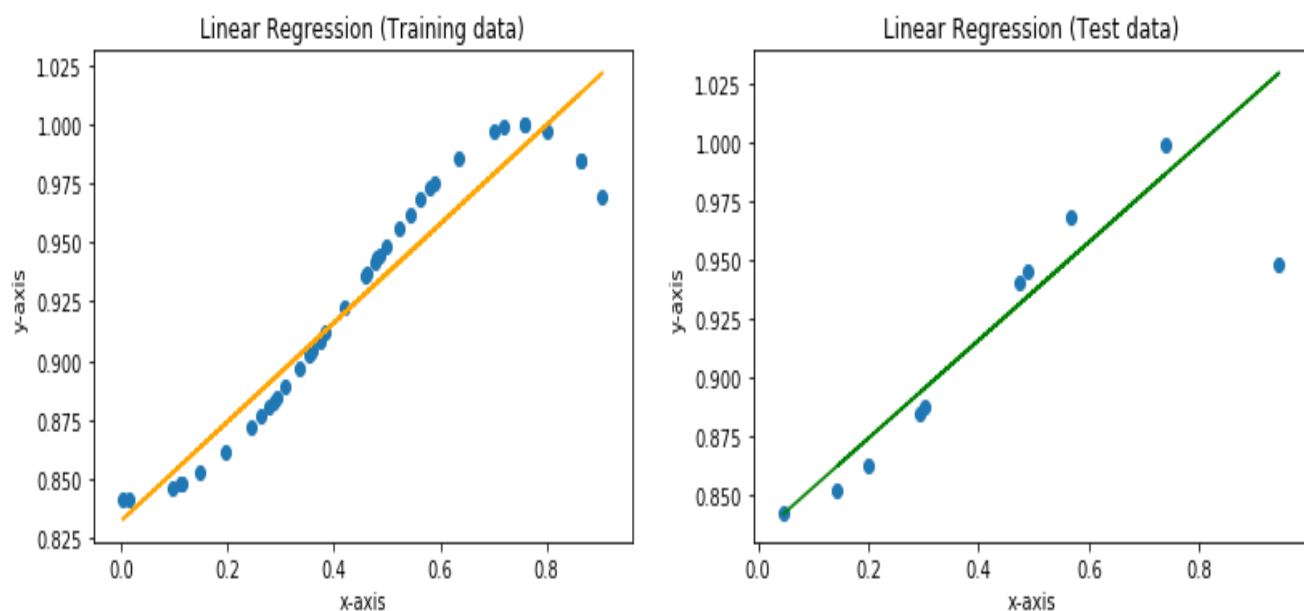
Implement the linear and polynomial regression models using the closed form solution as taught in the class on training data and explore their effectiveness in creating a predictor that best approximates the data using test data. Calculate the RMSE error and show the test and train results with varying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

Solution :

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:

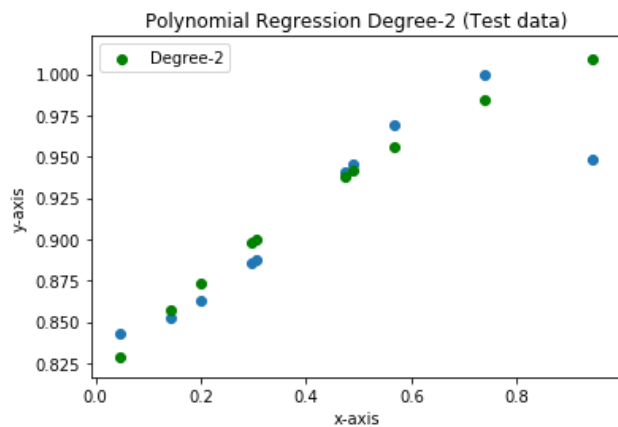
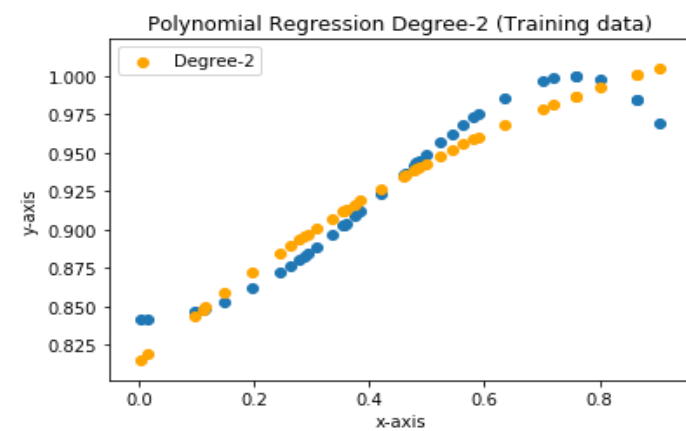


RMS error on training data = 0.03323156022291647

RMS error on test data = 0.03557615554980692

Polynomial Regression:

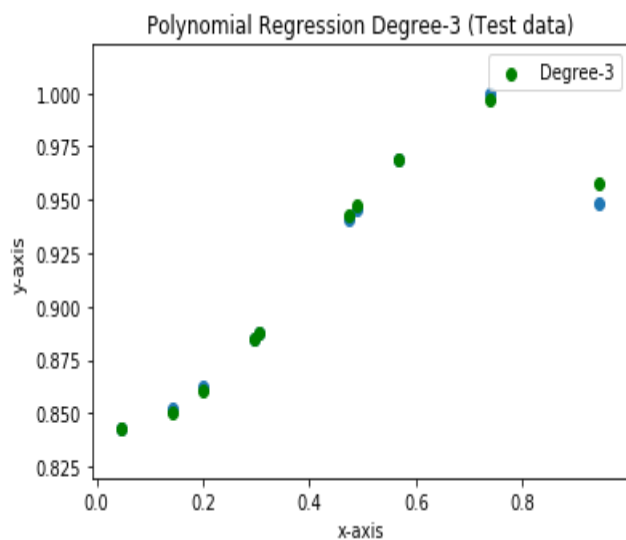
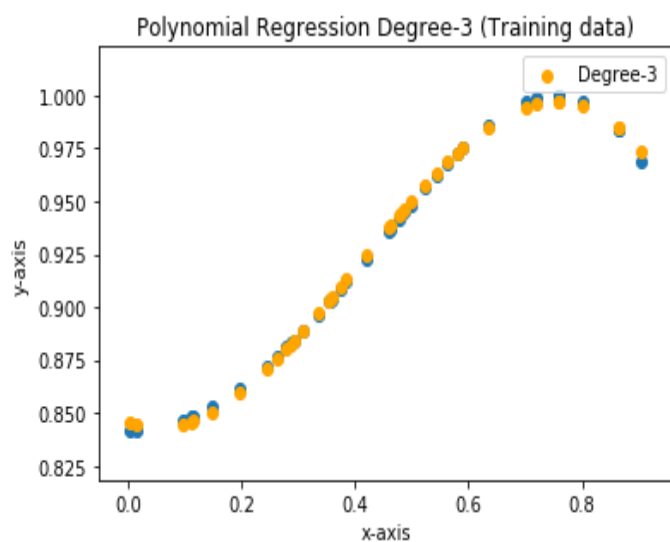
Degree-2 :



RMS error on training data = 0.01943684739300962

RMS error on test data = 0.024997030065501316

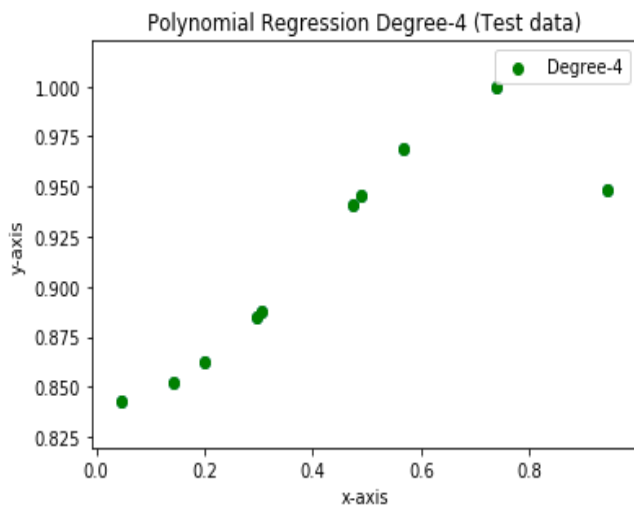
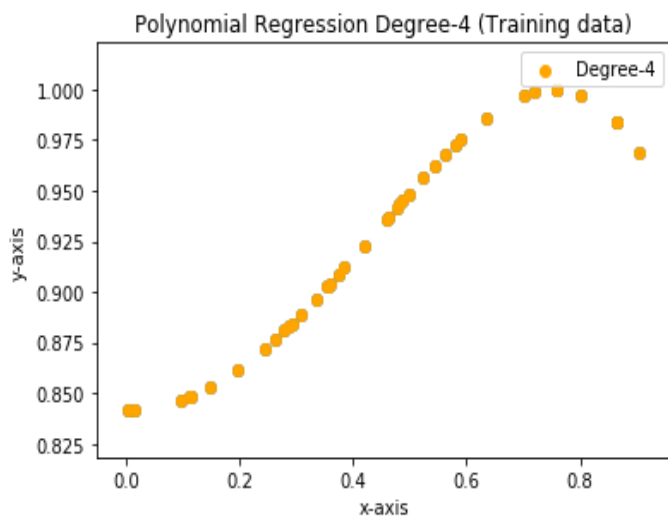
Degree-3 :



RMS error on training data = 0.002294967151908796

RMS error on test data = 0.0037108180703458457

Degree-4 :



RMS error on training data = 0.00012177970569231249
RMS error on test data = 0.00010426956217849425

Inference :

We can see that the RMS error for the 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.