Assignment 4

Dibyendu Roy Chaudhuri
MT19034

Q1)

Linear Regression

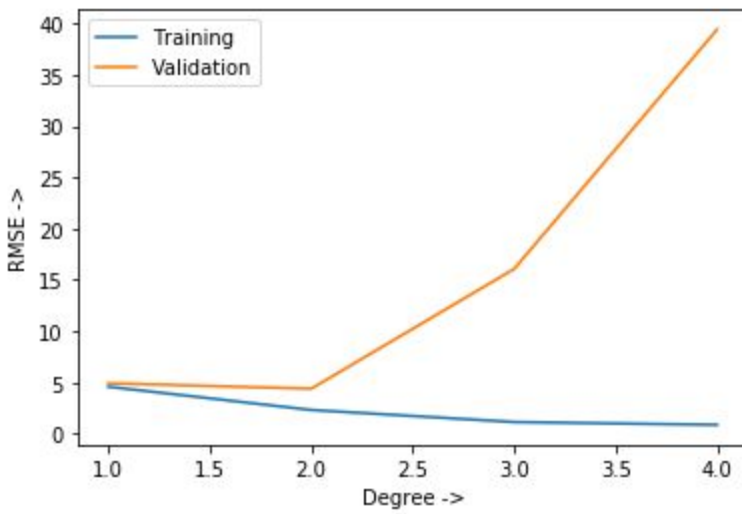**Training-Testing ratio- 80:20**

- Before splitting, the data set is shuffled so error is coming different each time.
- Training error -   **4.395243011302578**
- Testing error -   **5.0433708115698**

1) ## Polynomial Regression

All attributes polynomial

Degree	Total Attributes
1	14
2	105
3	560
4	2380

Least mean validation error is 4.384911247411972   **degree is** 2
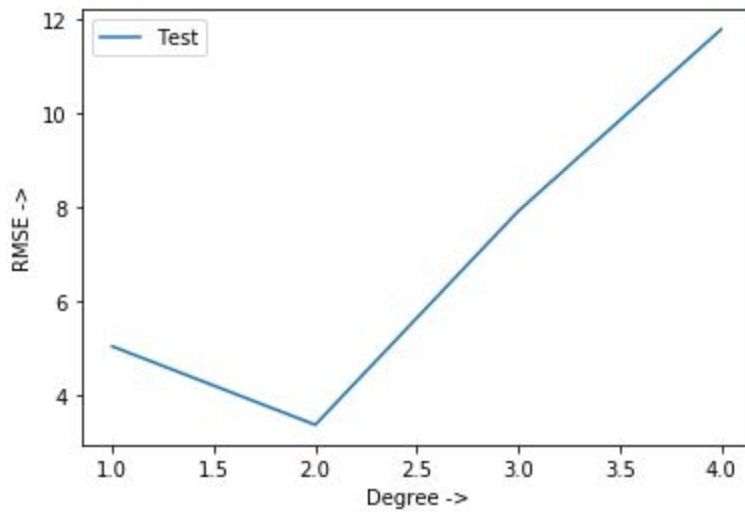
Average training RMSE

[4.581559179610763,2.3107480546603956, 1.133906016480121, 0.862004018671071]

Average validation RMSE

[4.916606245964653,4.384911247411972, 16.071265292695536, 39.43768024169051]
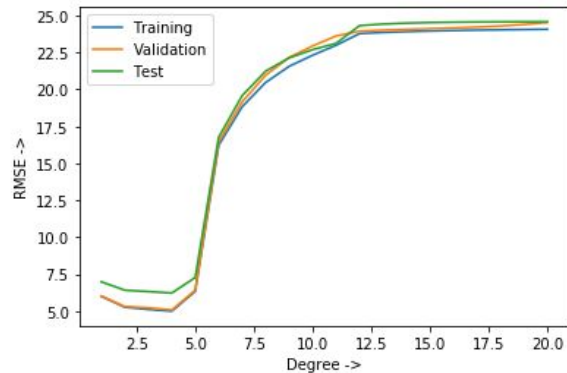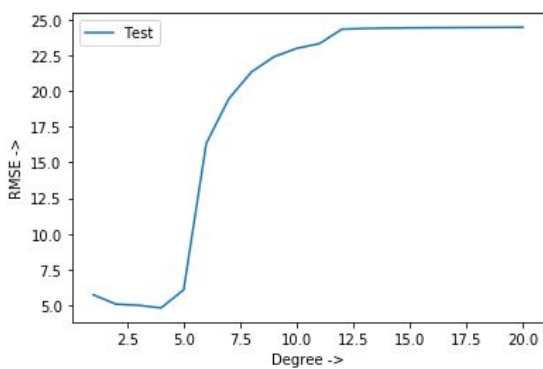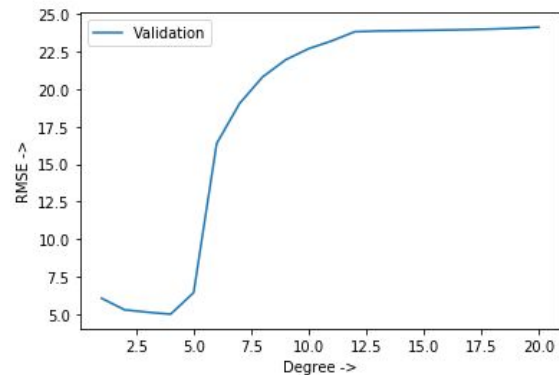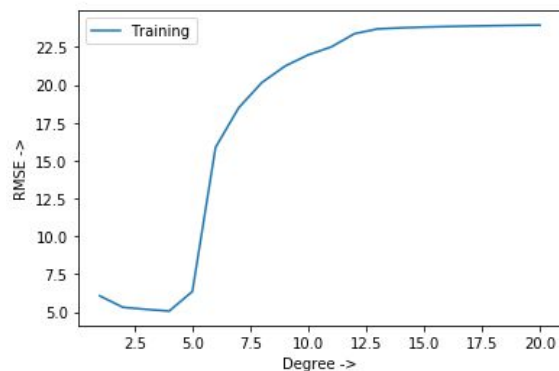
After degree 2, model overfit

Test RMSE at different degree (80:20)

[5.0433708115698, 3.37699011467745, 7.912480808694395, 11.7760264742906]

At best degree test RMSE(degree 2)- 3.376699011467745

2) LSTAT Polynomial

Least mean validation error is 5.38755129847791 **degree is** 4

Training-Testing ratio- 80:20
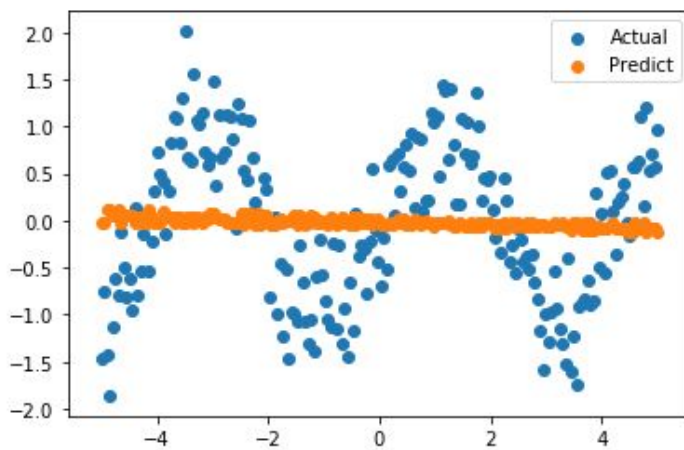
For degree 4 Test RMSE is 6.157160126204482

Q2)

Degree is 1

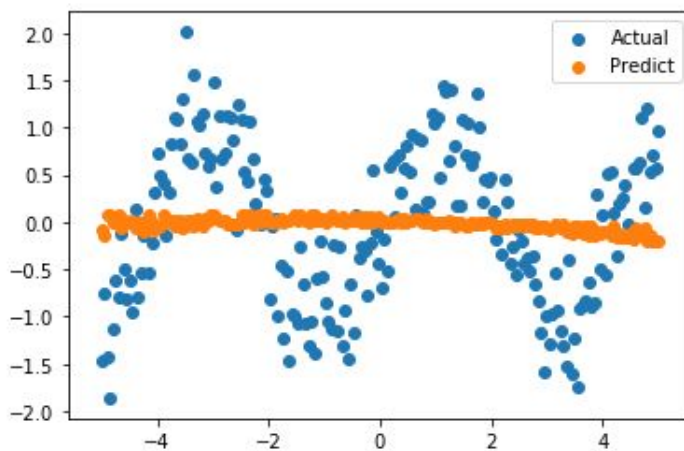train RMSE- 0.8179063496203307   test RMSE- 0.8222813568935567

Degree is 2

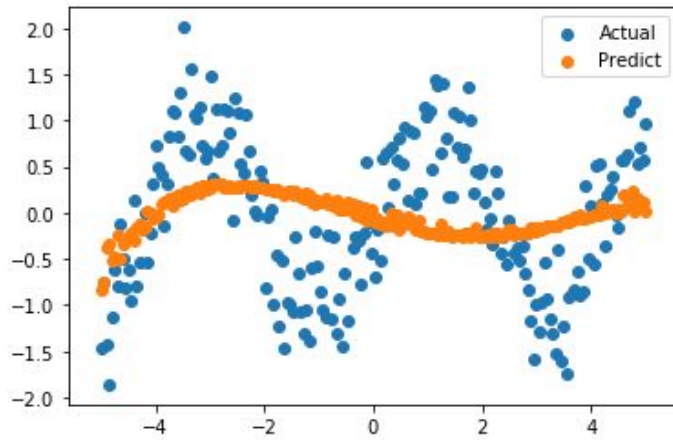train RMSE- 0.817104550029595   test RMSE- 0.8225897154264669

Degree is 4

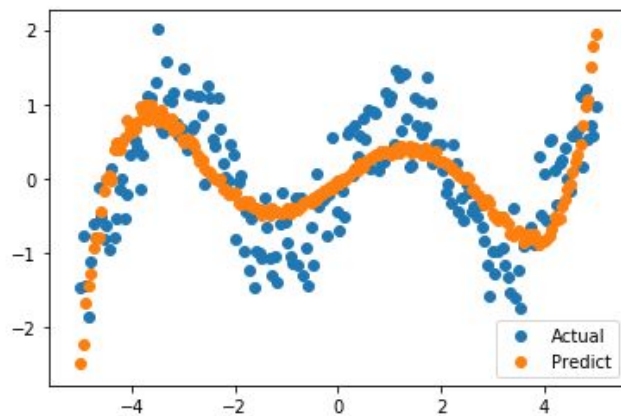train RMSE- 0.7909240225843871   test RMSE- 0.8015690135050845

Degree is 5

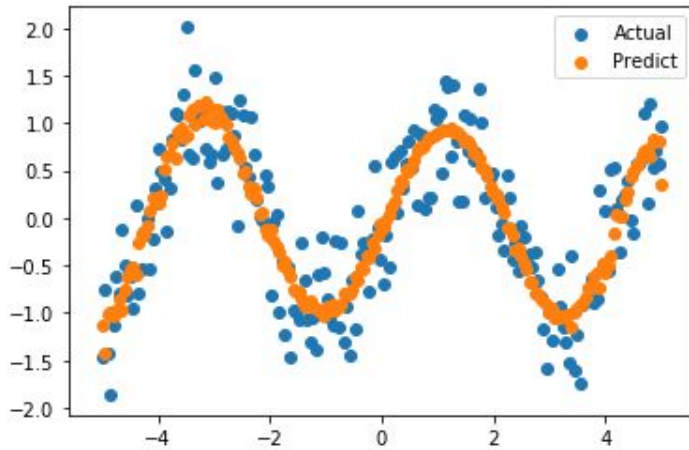train RMSE- 0.5698019850583831   test RMSE- 0.5932444181020593

Degree is 10

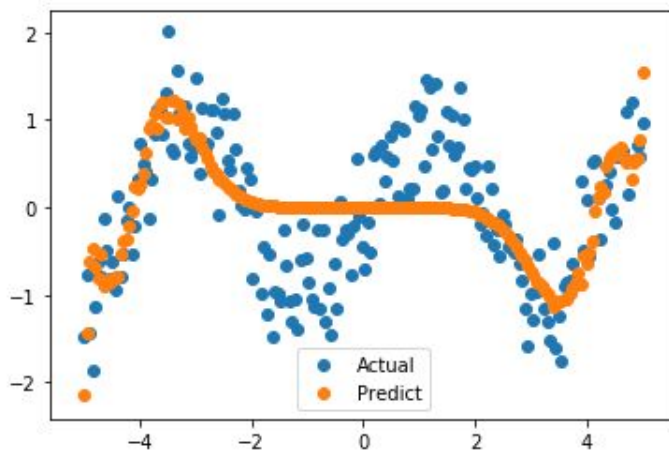train RMSE- 0.3873228248417007   test RMSE- 0.425896774726494
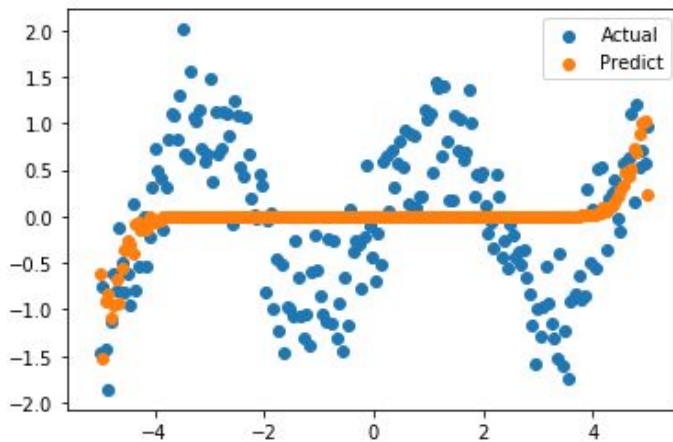

Degree is 15

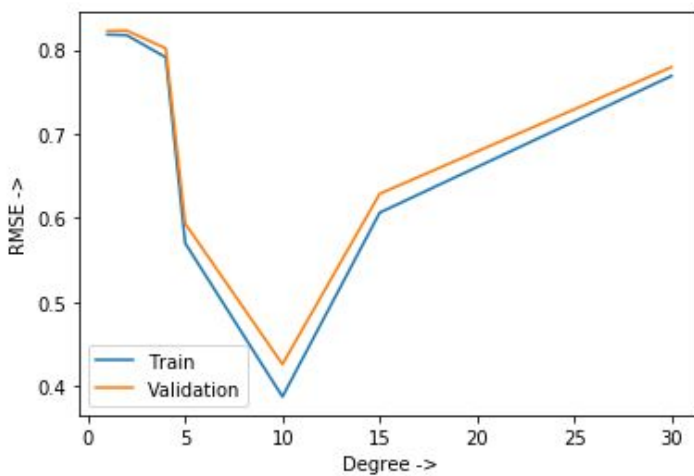train RMSE- 0.6061975304064193   test RMSE- 0.6286496193436244

Degree is 30

train RMSE- 0.7687294585556044   test RMSE- 0.7792314876940512

Plot RMSE at degree [1,2,4,5,10,15,30]

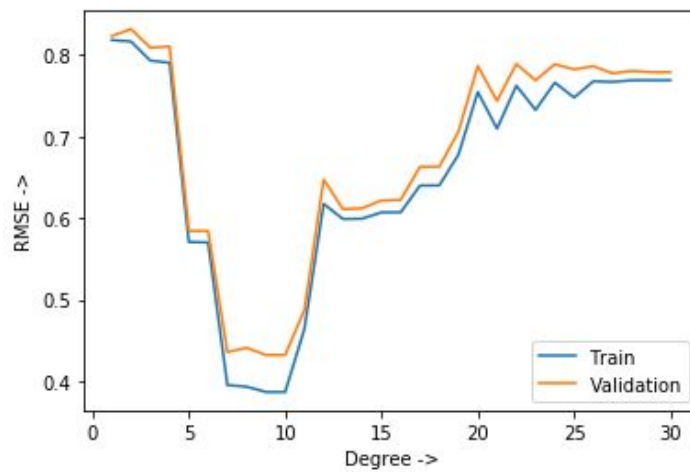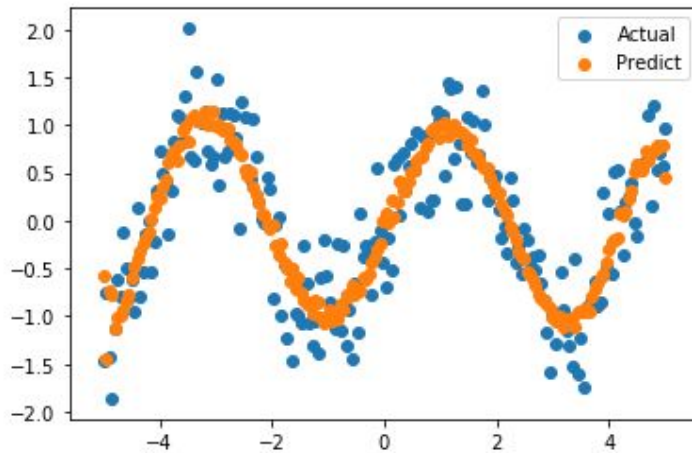Least mean validation error is 0.425896774726494   degree is 10

Plot RMSE at Degree (1-30)

Degree is 9

train RMSE- 0.3867926707260086   test RMSE- 0.4320451266876061

Least mean validation error is 0.4320451266876061   degree is 9