Name: Vaibhav Agrawal
Roll Number: 15CE10057
Assignment 1: Linear Regression
Machine Learning (CS60050)

---

1. **Synthetic data generation and simple curve fitting [10 + 5 + 25 = 40 marks]**

a. Generate a synthetic dataset as follows. The input values {x} are generated uniformly i in range [0, 1] , and the corresponding target values {yi} are obtained by first computing the corresponding values of the function sin(2Πx) , and then adding a random noise with a Gaussian distribution having standard deviation 0.3. Generate 10 such instances of (xi , yi) . [You can use any standard module to generate random numbers as per a gaussian / normal distribution, e.g., numpy.random.normal for python.]

b. Split the dataset into two sets randomly: (i) Training Set (80%) (ii) Test Set (20%).

c. Write a code to fit a curve that minimises squared error cost function using gradient descent (with learning rate 0.05), as discussed in class, on the training set while the model takes following form $y = W^T \Phi_n(x)$, $W \in R^{n+1}$ $\Phi_n(x) = [1, x^1, x^2, x^3, \dots, x^n]$. Squared error is defined as $J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(W^T \Phi_n(x) - y)^2$. In your experiment, vary n from 1 to 9. In other words, fit 9 different curves to the training data, and hence estimate the parameters. Use the estimated W to measure squared error on the test set, and name it as test error on test data.

2. **Visualisation of the dataset and the fitted curves [10 + 10 = 20 marks]**

a. Draw separate plots of the synthetic data points generated in 1 (a), and all 9 different curves that you have fit for the given dataset in 1 (c).

b. Report squared error on both train and test data for each value of n in the form of a plot where along x-axis, vary n from 1 to 9 and along y-axis, plot both train error and test error. Explain which value of n is suitable for the synthetic dataset that you have generated and why.

3. **Experimenting with larger training set [10 marks]**

Repeat the above experiment with three other datasets having size 100, 1000 and 10,000 instances (each dataset generated similarly as described in Part 1a). Draw the learning curve of how train and test error varies with increase in size of datasets (for 10, 100, 1000 and 10000 instances).

4. **Experimenting with cost functions [20 + 10 = 30 marks]**

a. Solve the problem by minimizing different cost functions (Do not use any regularization, Use gradient descent to minimize the cost function in each case) :

   i. Mean absolute error i.e. $J(\Theta) = \frac{1}{2m}\sum_{i=1}^{m}|W^T \Phi_n(x) - y|$

   ii. Fourth power error i.e. $J(\Theta) = \frac{1}{2m}\sum_{i=1}^{m}(W^T \Phi_n(x) - y)^4$

b. Plot the test RMSE vs learning rate for each of the cost functions. Vary the learning rates as 0.025, 0.05, 0.1, 0.2 and 0.5. Which one would you prefer for this problem and why?

Language Used: Python
Files submitted as .ipynb (Python Notebook)

The report consists of graphs and the required result data mentioned in the assignment questioner.

The report is divided into four following sections:

1. Squared Error Cost Function
$$J(\Theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(W^T \Phi_n(x)-y\right)^2$$
2. Mean Absolute Error Cost Function
$$J(\Theta) = \frac{1}{2m}\sum_{i=1}^{m}\left|W^T \Phi_n(x)-y\right|$$
3. Fourth power Error Cost Function
$$J(\Theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(W^T \Phi_n(x)-y\right)^4$$
4. Learning curve of how train and test error varies with increase in size of datasets

First three section are further divided into 10 cases. Each case considers the different form the model takes, starting from N=0 to N=9.

The equation is given as follows:

$$y = W^T \Phi_n(x)$$

$$W \in R^{n+1},\ \Phi_n(x) = [1, x^1, x^2, x^3, x^4, x^5, x^6, x^7, x^8, x^9]$$

For each case a plot of the synthetic data points and the best curve fit is shown. Also the plot of cost function error(train and test) vs n is shown, followed by the plot of test RMSE vs learning rate for the respective cost functions.

Last section give the learning curve of how train and test error varies with increase in size of datasets. The Squared Error cost function is considered for model

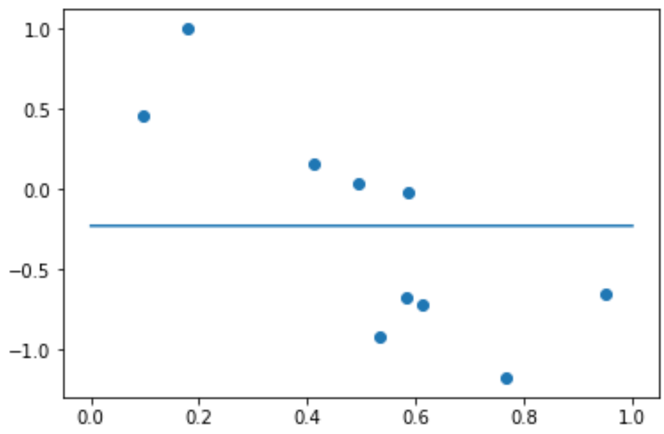
$$Y = \beta_0 + \beta_1{*}X + \beta_2{*}X^2 + \beta_3{*}X^3 + \beta_4{*}X^4$$

Also, for each cost function and the a particular datasets the best fitting model is identified. And the Beta matrix is described in the report.

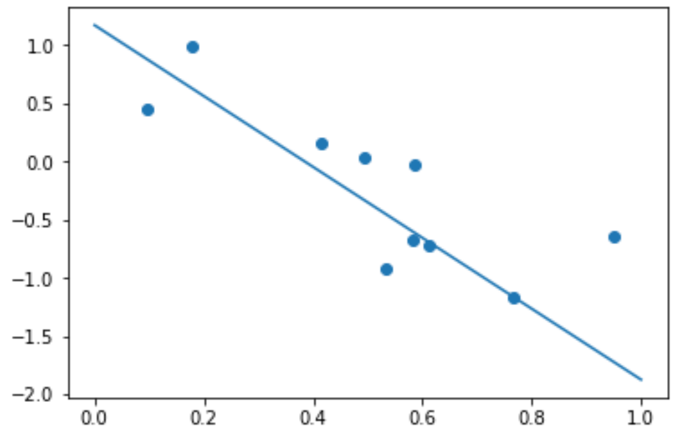
- **<u>Cost Function 1:</u> Squared Error**

I. <u>Number of data points = 10</u>

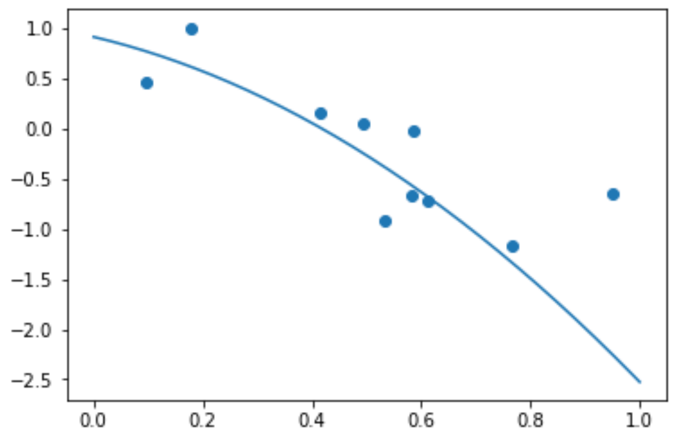
$Y = \beta_0$



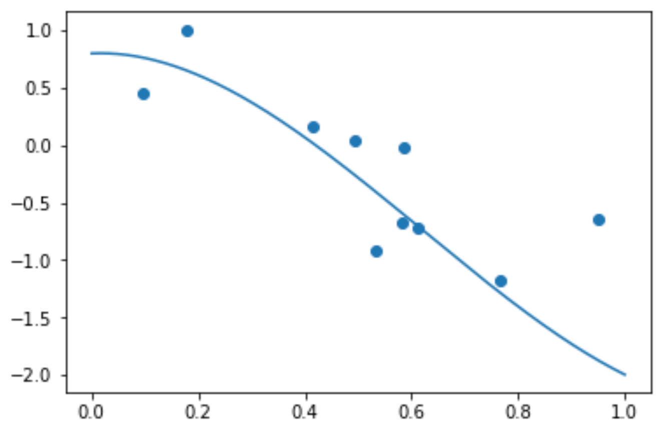
$Y = \beta_0 + \beta_1 * X$



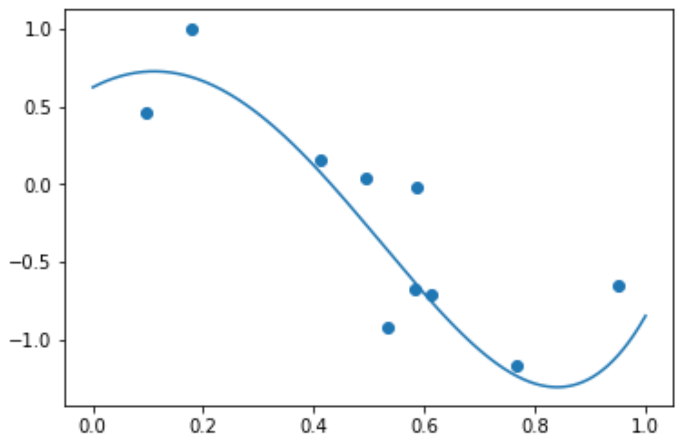
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2$



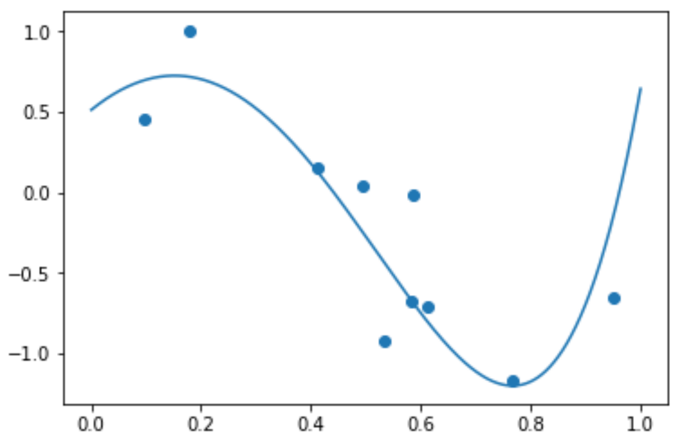
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3$



$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4$

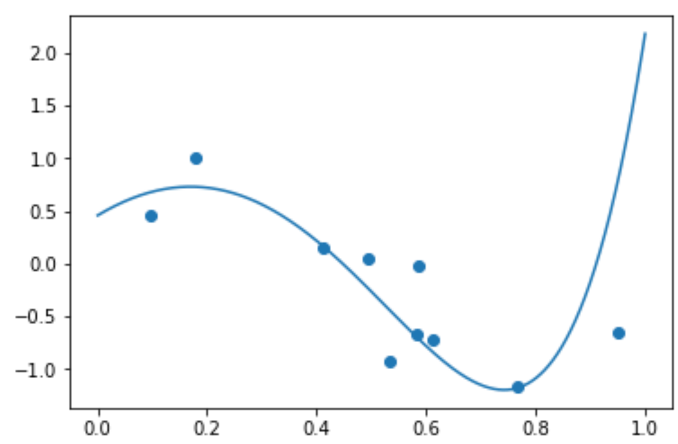


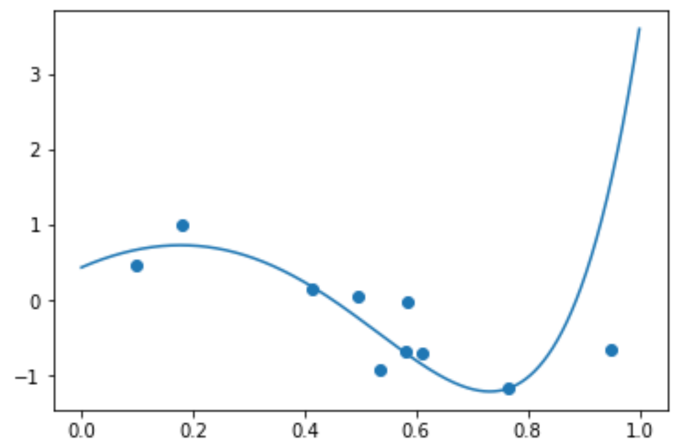
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5$



$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6$

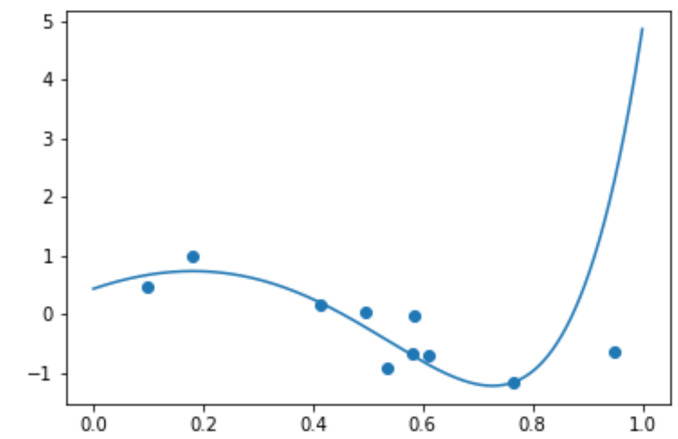
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7$

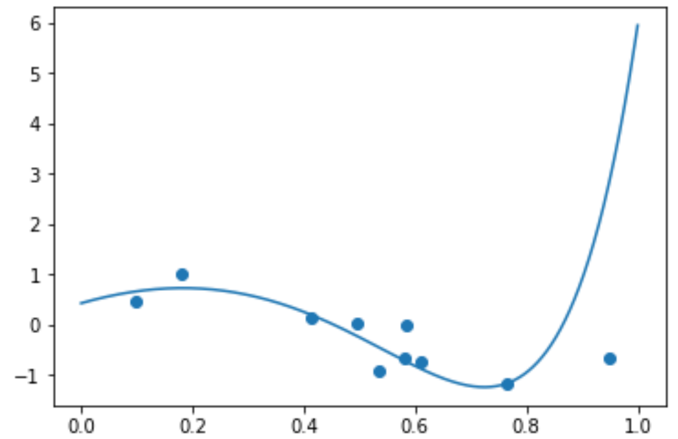




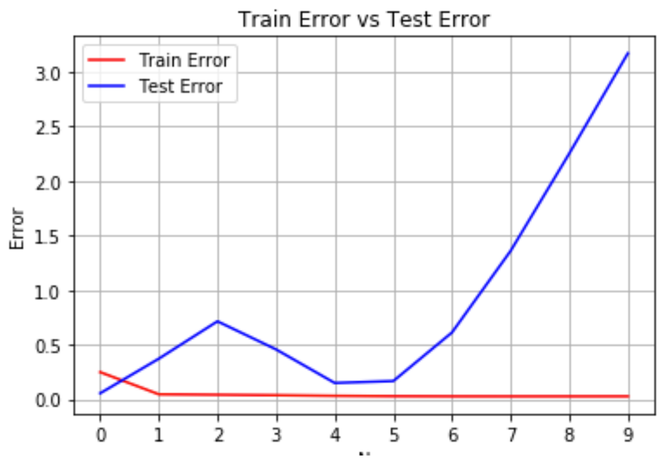
$$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7 + \beta_8 * X^8$$

$$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7 + \beta_8 * X^8 + \beta_9 * X^9$$





Train Error vs Test Error:



From the plot we can conclude that N = 4 is the best fit because the test error is minimum. For N>4 there is overfitting and as the test error is increasing thus N>4 is ignored.

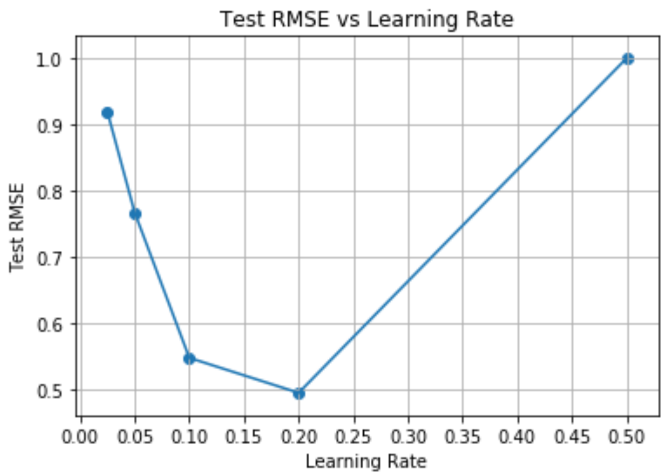
Therefore,

`Y = B0 + B1*X1 + B2*X2 + B3*X3 + B4*X4`

Learned Value of the parameter i.e. beta matrix =

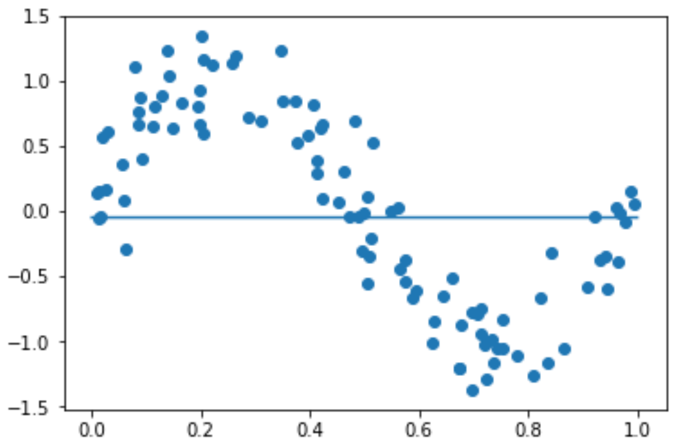
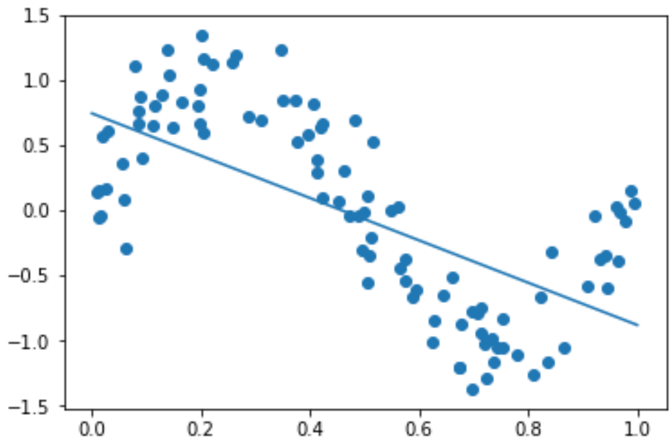
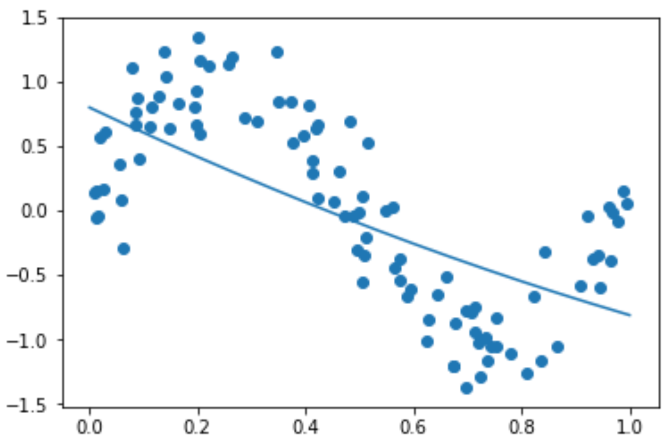
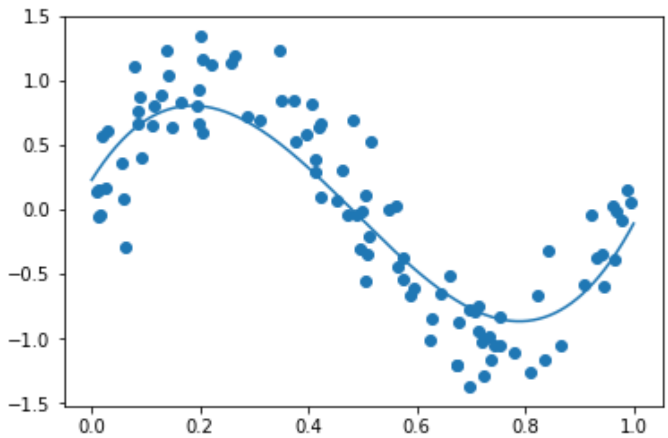
```
array([ 0.62480915,  1.858861  , -8.3162825 , -1.1383047 ,  6.12220294])
```

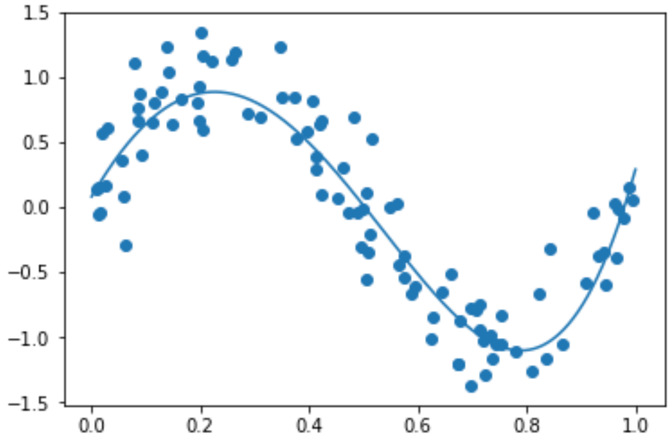
Test RMSE vs Learning Rate:

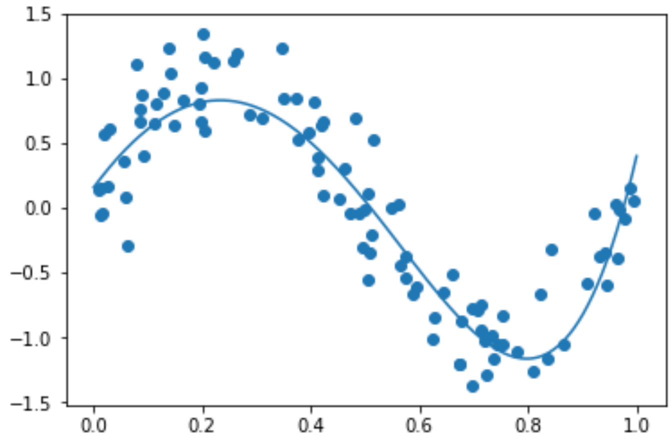


Lower values of RMSE indicate better fit. Thus from the plot we can conclude that α = 0.2 is the best fit because the test RMSE is minimum.

II. Number of data points = 100

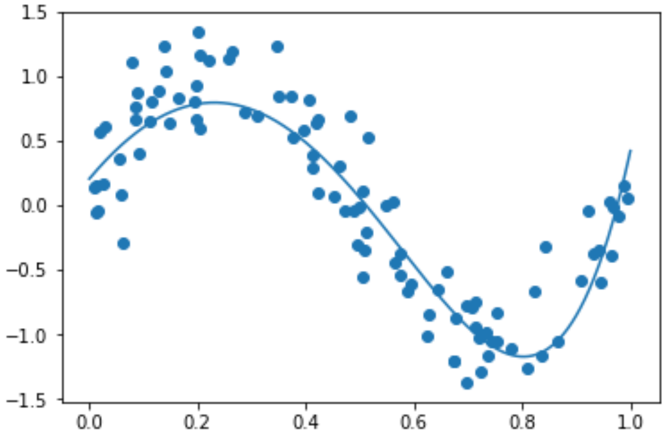
$Y = \beta_0$	$Y = \beta_0 + \beta_1 * X$
	
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2$	$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3$
	
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4$	$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5$

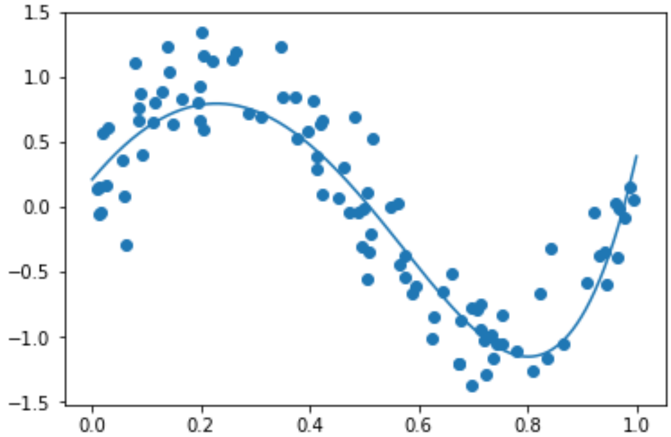




$$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6$$

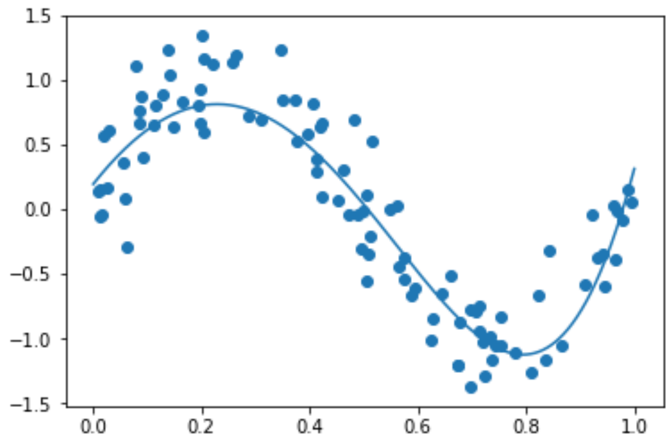
$$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7$$

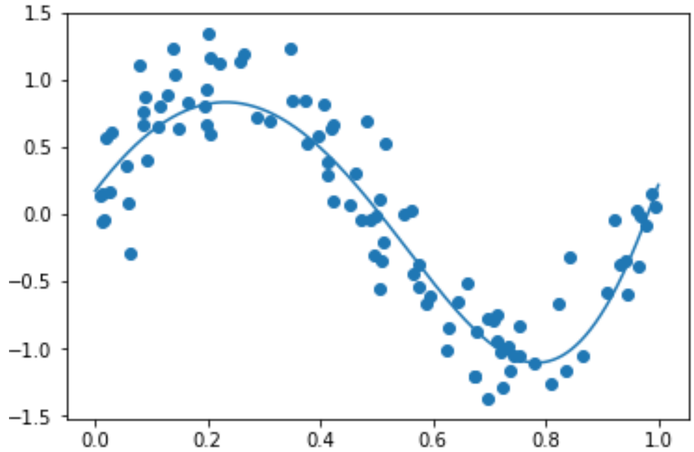




$$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7 + \beta_8 * X^8$$

$$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7 + \beta_8 * X^8 + \beta_9 * X^9$$

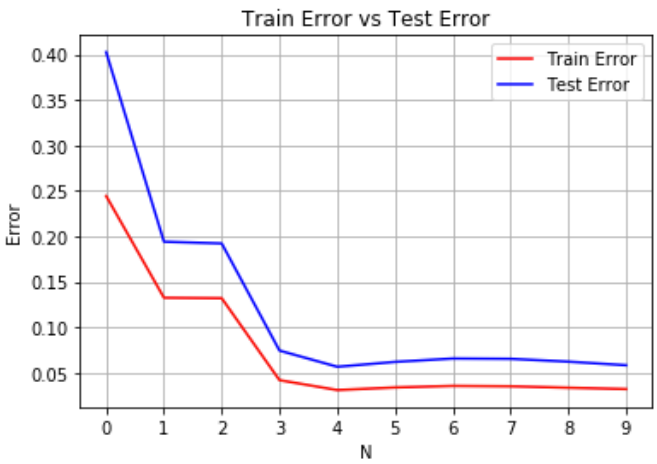




Learned Value of the parameter i.e. beta matrix =

```
array([0.08016677, 7.23077021, -15.94562914, -4.04716708, 12.97169312])
```

Train Error vs Test Error:

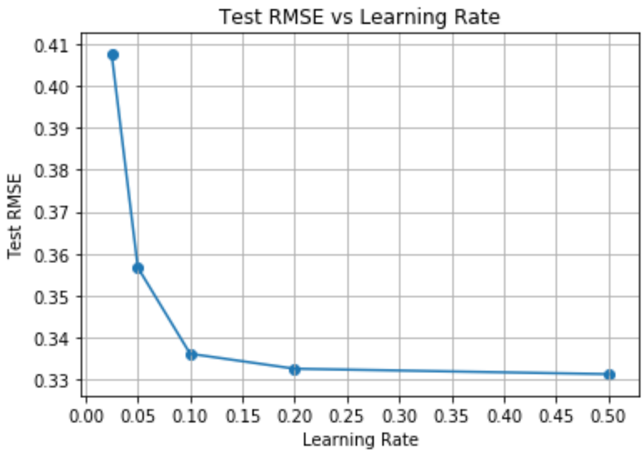


From the above plot we can conclude that N = 4 is the best fit because the test error is minimum. For N>4 there is overfitting and as the test error is increasing thus N=9 is ignored.

Therefore,

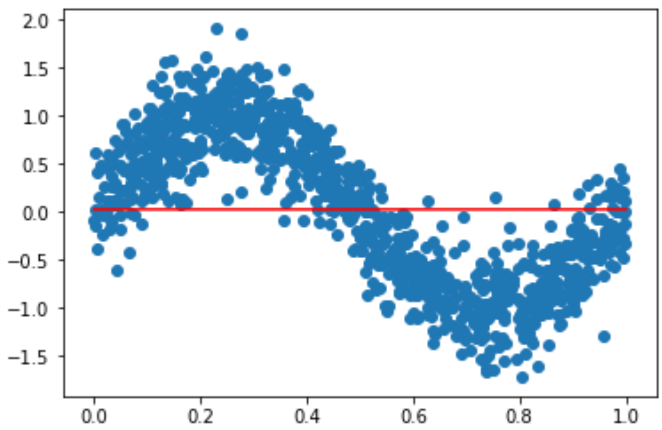
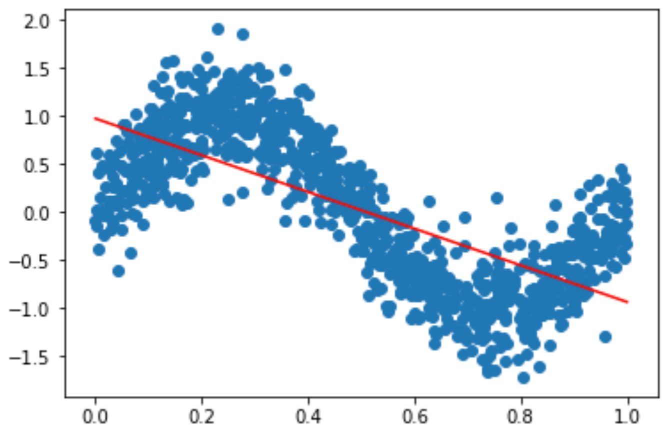
```
Y = B0 + B1*X1 + B2*X2 + B3*X3 + B4*X4
```

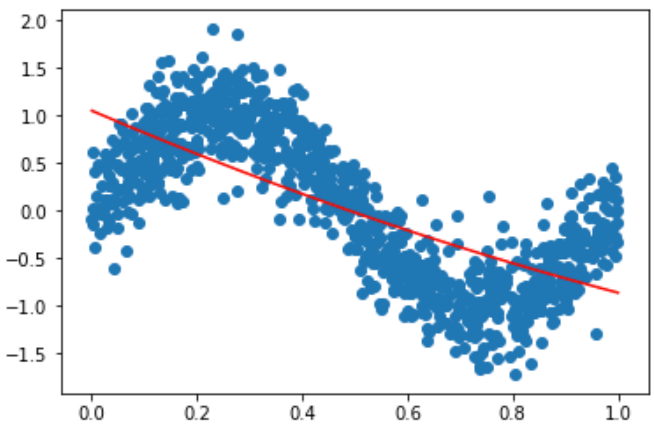
Test RMSE vs Learning Rate:



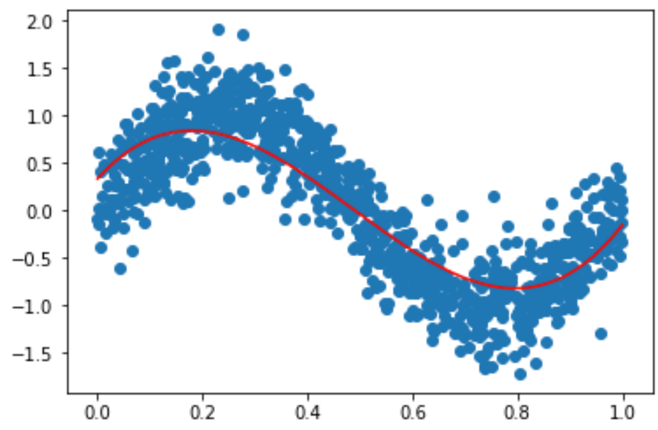
Lower values of RMSE indicate better fit. Thus from the plot we can conclude that α = 0.5 is the best fit because the test RMSE is minimum.

III. Number of data points = 1000

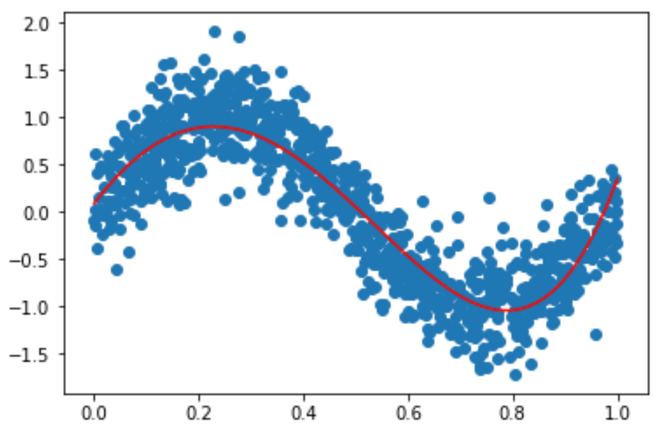
$Y = \beta_0$	$Y = \beta_0 + \beta_1 {*} X$
	
$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2$	$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2 + \beta_3 {*} X^3$



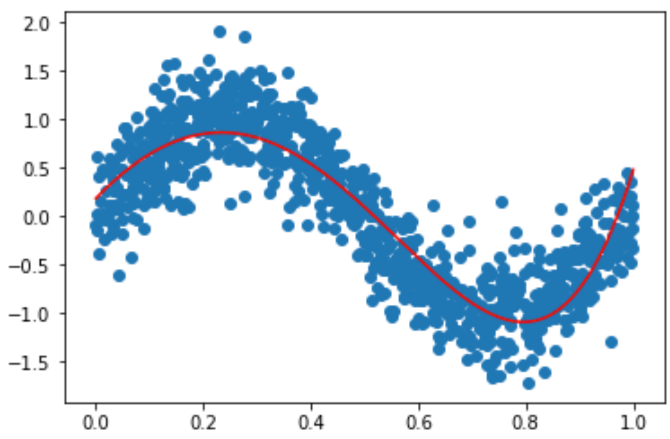
$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2 + \beta_3 {*} X^3 + \beta_4 {*} X^4$



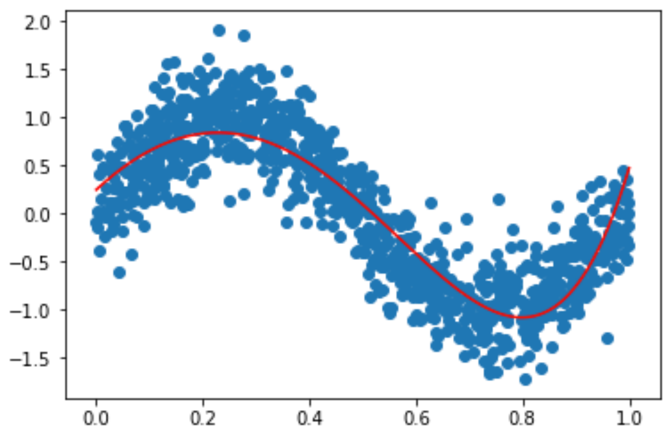
$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2 + \beta_3 {*} X^3 + \beta_4 {*} X^4 + \beta_5 {*} X^5$



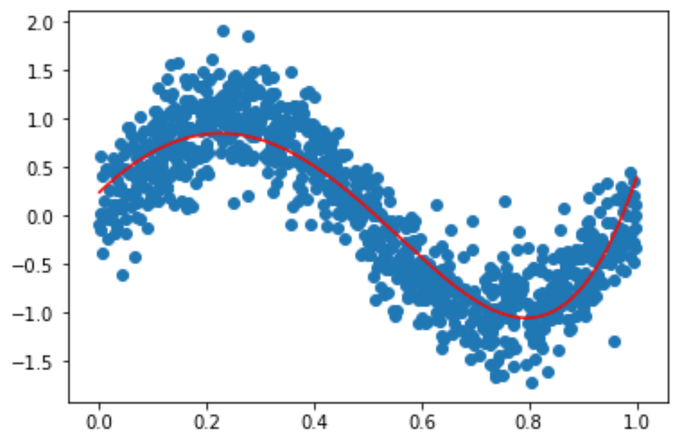
$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2 + \beta_3 {*} X^3 + \beta_4 {*} X^4 + \beta_5 {*} X^5 + \beta_6 {*} X^6$



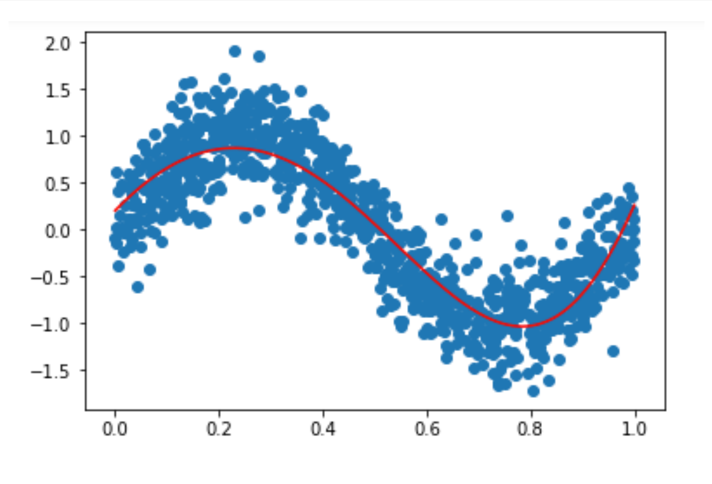
$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2 + \beta_3 {*} X^3 + \beta_4 {*} X^4 + \beta_5 {*} X^5 + \beta_6 {*} X^6 + \beta_7 {*} X^7$

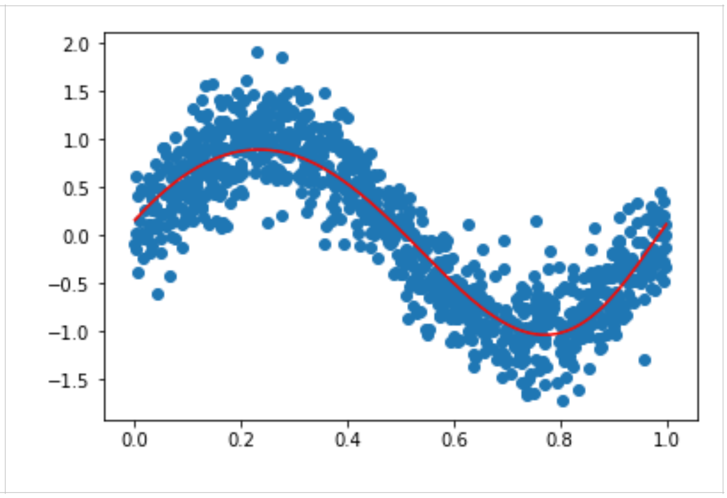


$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2 + \beta_3 {*} X^3 + \beta_4 {*} X^4 + \beta_5 {*} X^5 + \beta_6 {*} X^6 + \beta_7 {*} X^7 + \beta_8 {*} X^8$



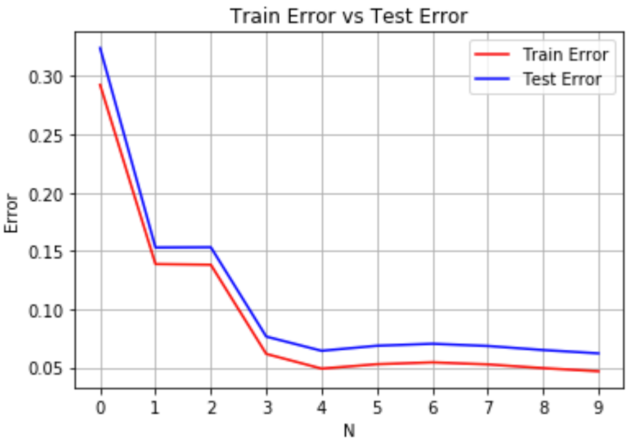
$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2 + \beta_3 {*} X^3 + \beta_4 {*} X^4 + \beta_5 {*} X^5 + \beta_6 {*} X^6 + \beta_7 {*} X^7 + \beta_8 {*} X^8 + \beta_9 {*} X^9$





Learned Value of the parameter i.e. beta matrix =
`array([ 0.08891846, 7.23738216, -15.85199066, -3.94442207, 12.81661213])`

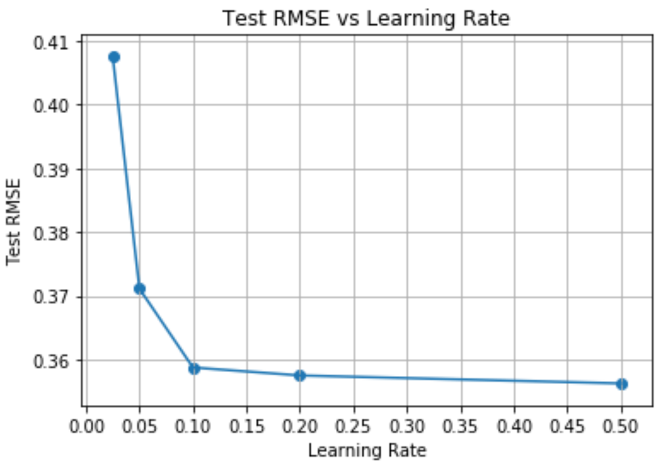
Train Error vs Test Error:



From the above plot we can conclude that N = 4 is the best fit because the test error is minimum. For N>4 there is overfitting and as the test error is increasing thus N=9 is ignored.

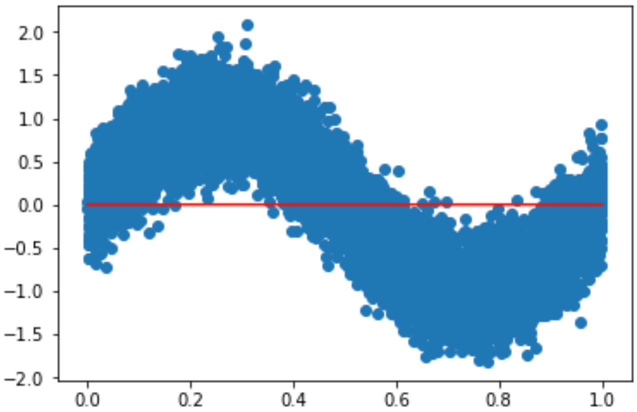
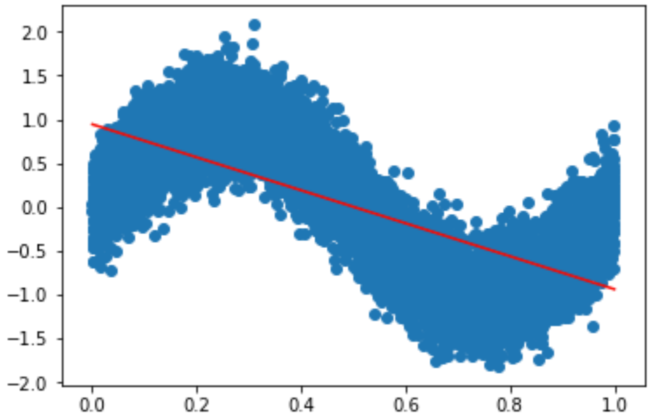
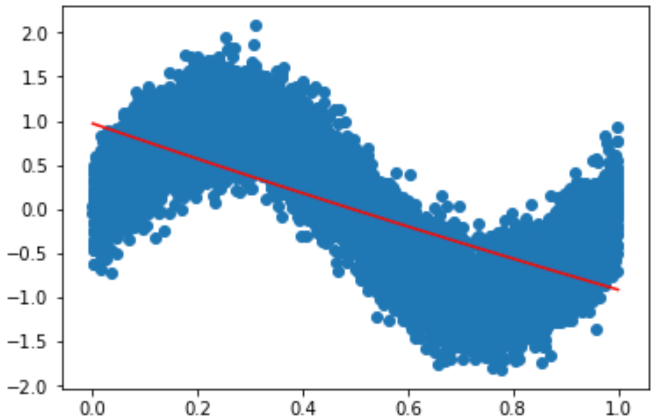
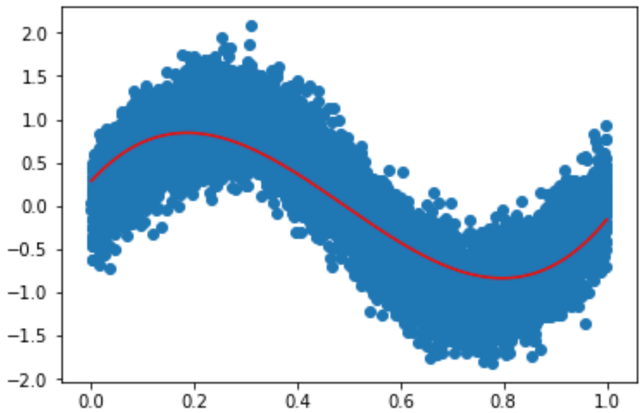
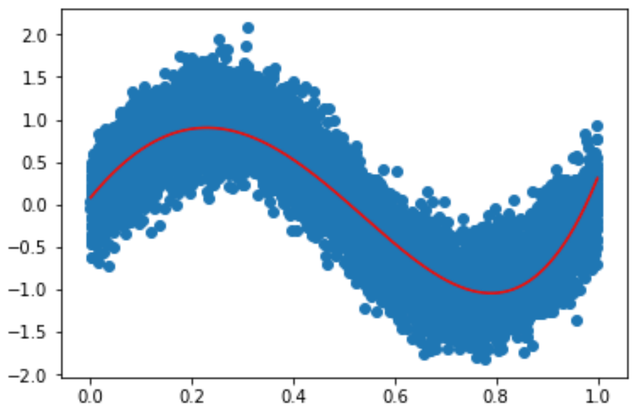
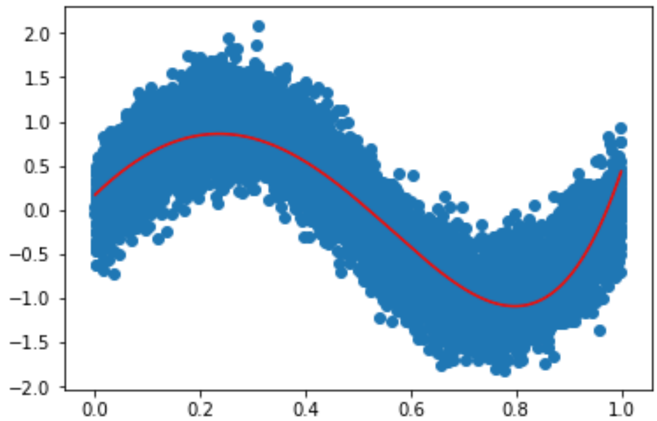
Therefore,
`Y = B0 + B1*X1 + B2*X2 + B3*X3 + B4*X4`

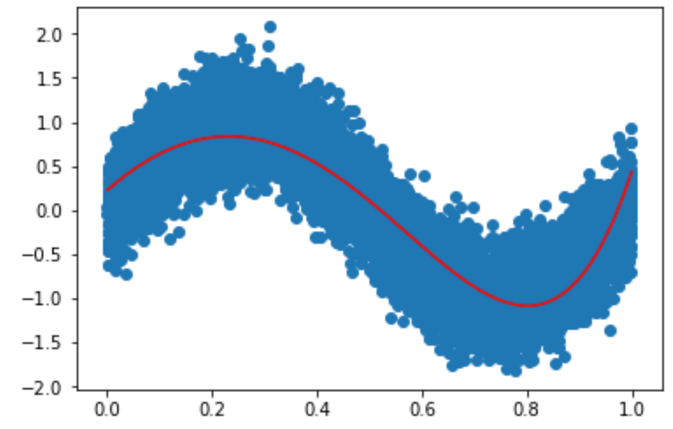
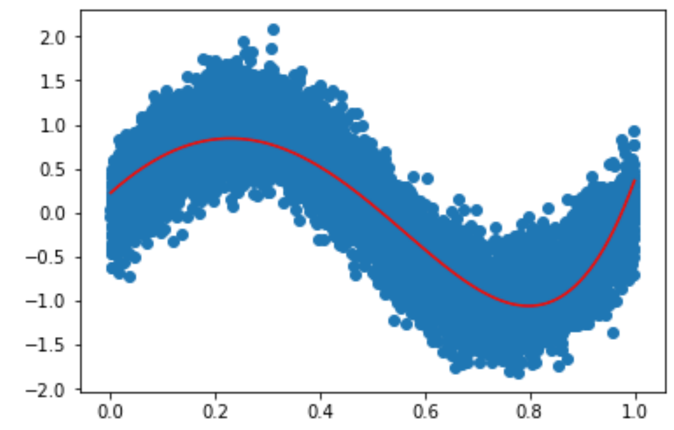
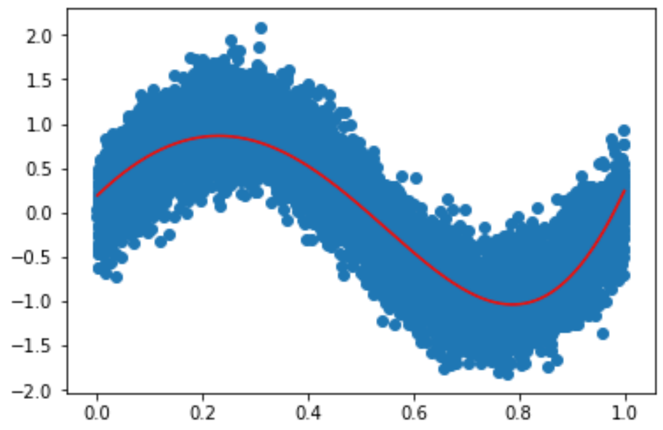
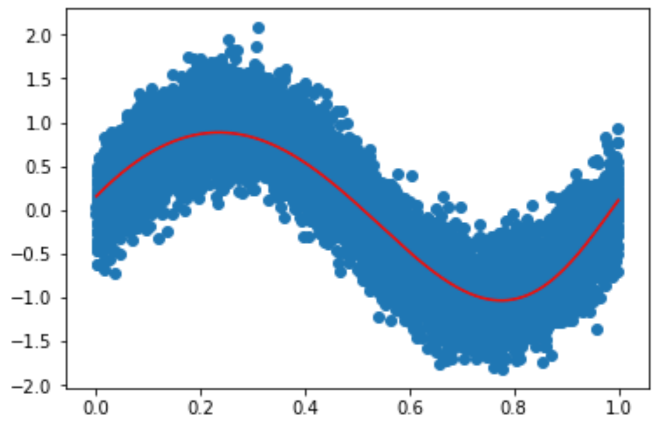
Test RMSE vs Learning Rate:



Lower values of RMSE indicate better fit.
Thus from the plot we can conclude that α = 0.5 is the best fit because the test RMSE is minimum.

IV. Number of data points = 10000

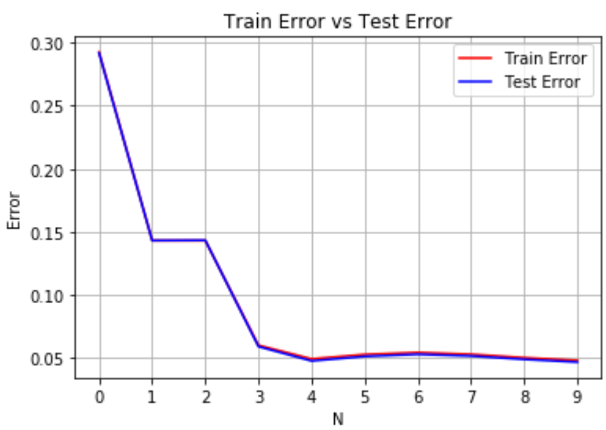
$Y = \beta_0$	$Y = \beta_0 + \beta_1 * X$
	
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2$	$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3$
	
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4$	$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5$
	
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6$	$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7$

	
$Y = \beta_0 + \beta_1{*}X + \beta_2{*}X^2 + \beta_3{*}X^3 + \beta_4{*}X^4 + \beta_5{*}X^5 + \beta_6{*}X^6 + \beta_7{*}X^7 + \beta_8{*}X^8$	$Y = \beta_0 + \beta_1{*}X + \beta_2{*}X^2 + \beta_3{*}X^3 + \beta_4{*}X^4 + \beta_5{*}X^5 + \beta_6{*}X^6 + \beta_7{*}X^7 + \beta_8{*}X^8 + \beta_9{*}X^9$
	

Learned Value of the parameter i.e. beta matrix =

```
array([ 0.08055777, 7.27913518, -15.88369935, -3.89211768, 12.72452468])
```

Train Error vs Test Error:

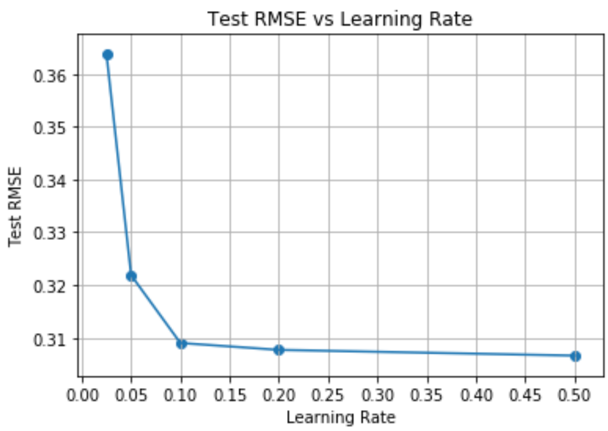


From the above plot we can conclude that N = 4 is the best fit because the test error is minimum. For N>4 there is overfitting and as the test error is increasing thus N=9 is ignored.

Therefore,

```
Y = B0 + B1*X1 + B2*X2 + B3*X3 + B4*X4
```

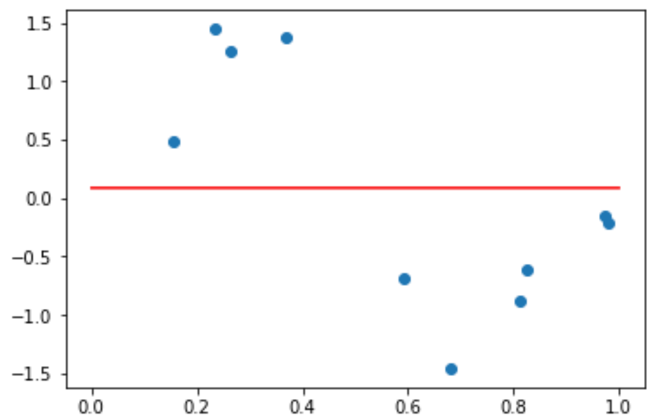
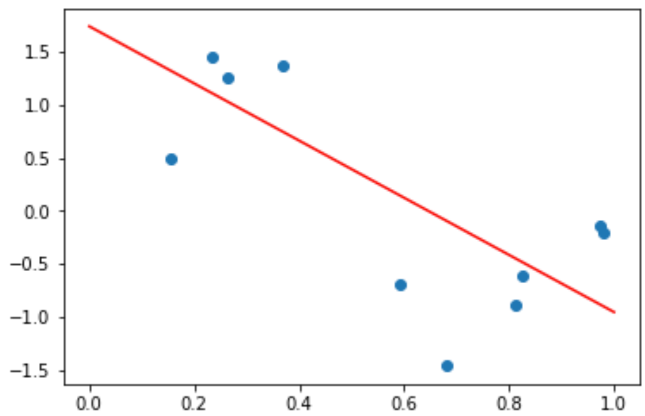
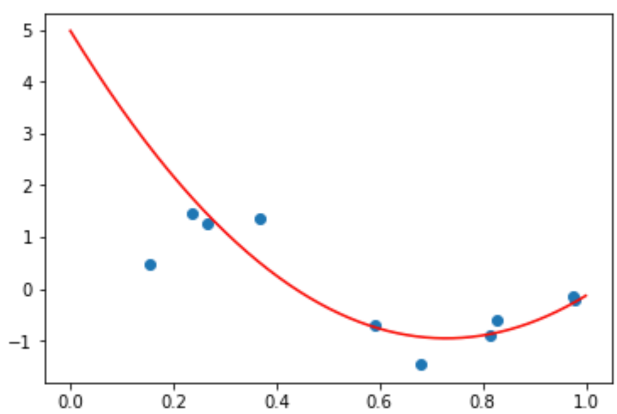
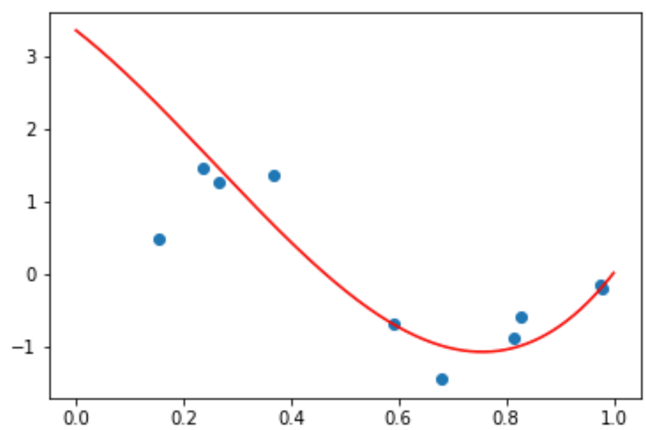
Test RMSE vs Learning Rate:

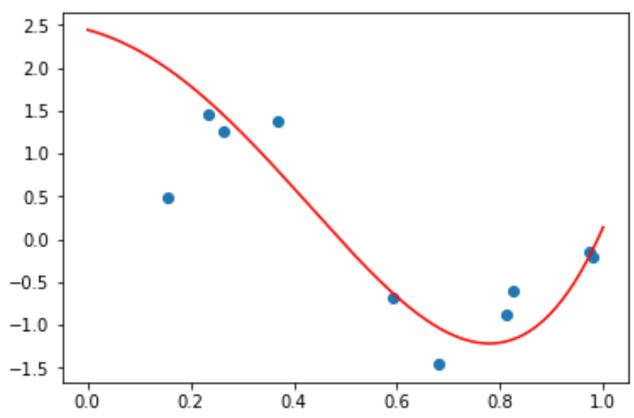


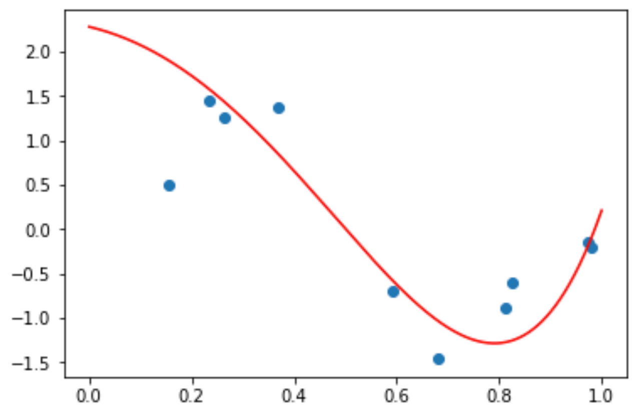
**Lower values of RMSE indicate better fit.
Thus from the plot we can conclude that α = 0.5 is the best fit because the test RMSE is minimum.**

- **Cost Function 2: Mean Absolute Error**

I. Number of data points = 10

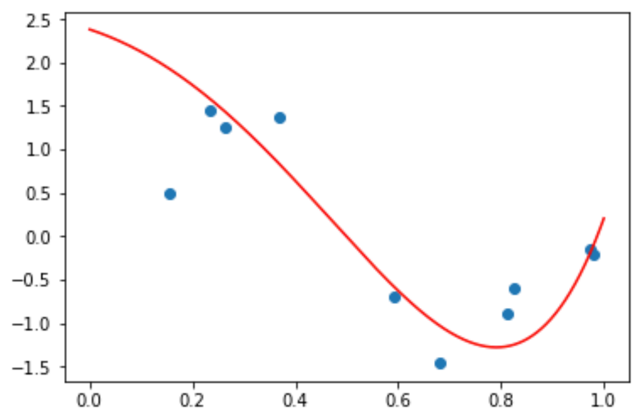
$Y = \beta_0$	$Y = \beta_0 + \beta_1 {*} X$
	
$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2$	$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2 + \beta_3 {*} X^3$
	
$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2 + \beta_3 {*} X^3 + \beta_4 {*} X^4$	$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2 + \beta_3 {*} X^3 + \beta_4 {*} X^4 + \beta_5 {*} X^5$

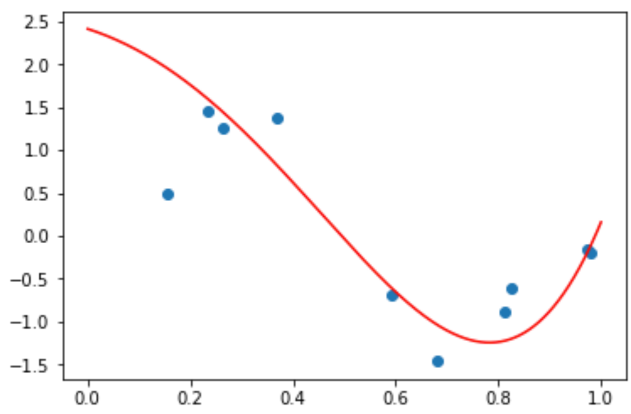




$$Y = \beta_0 + \beta_1{*}X + \beta_2{*}X^2 + \beta_3{*}X^3 + \beta_4{*}X^4 + \beta_5{*}X^5 + \beta_6{*}X^6$$

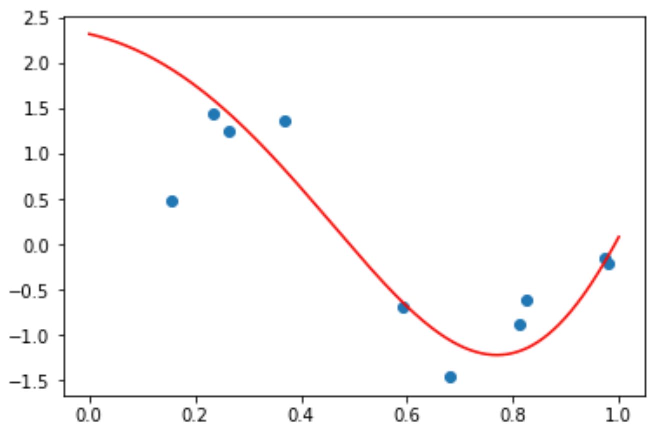
$$Y = \beta_0 + \beta_1{*}X + \beta_2{*}X^2 + \beta_3{*}X^3 + \beta_4{*}X^4 + \beta_5{*}X^5 + \beta_6{*}X^6 + \beta_7{*}X^7$$

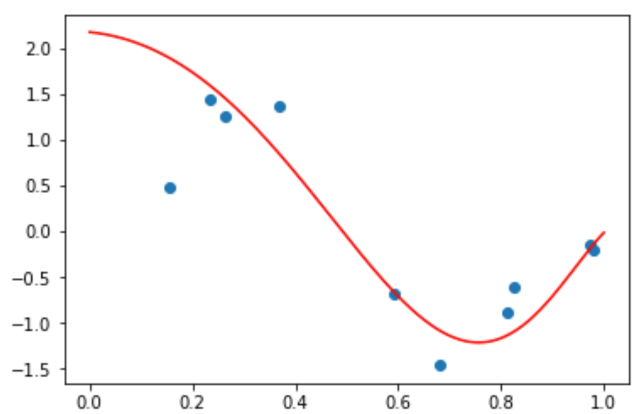




$$Y = \beta_0 + \beta_1{*}X + \beta_2{*}X^2 + \beta_3{*}X^3 + \beta_4{*}X^4 + \beta_5{*}X^5 + \beta_6{*}X^6 + \beta_7{*}X^7 + \beta_8{*}X^8$$

$$Y = \beta_0 + \beta_1{*}X + \beta_2{*}X^2 + \beta_3{*}X^3 + \beta_4{*}X^4 + \beta_5{*}X^5 + \beta_6{*}X^6 + \beta_7{*}X^7 + \beta_8{*}X^8 + \beta_9{*}X^9$$

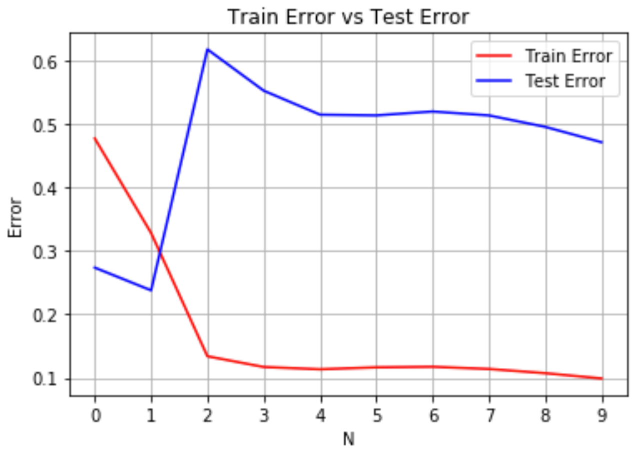




Learned Value of the parameter i.e. beta matrix =

```
array([ 1.74200466, -2.69666284])
```

Train Error vs Test Error:

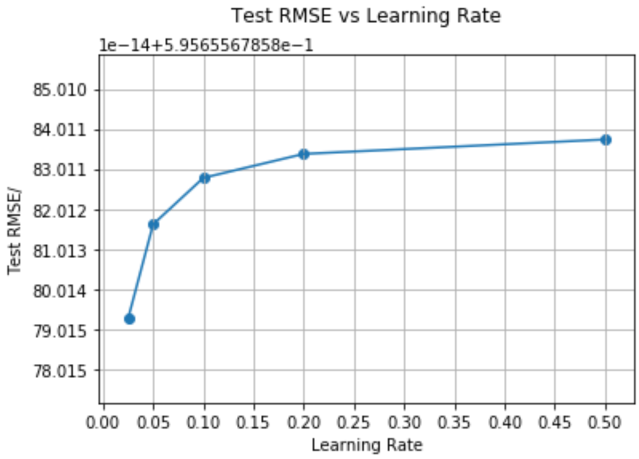


From the above plot we can conclude that N = 1 is the best fit because the test error is minimum. For N>1 there is overfitting and as the test error is increasing thus N>1 is ignored.

Therefore,

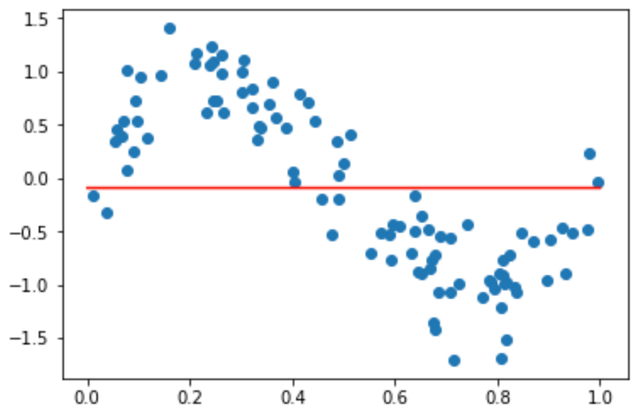
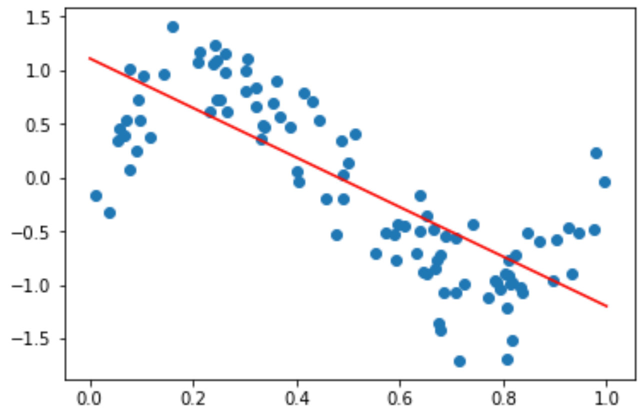
`Y = B0 + B1*X1`

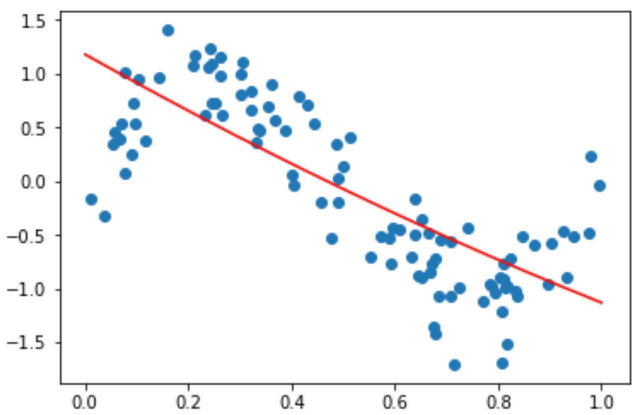
Test RMSE vs Learning Rate:



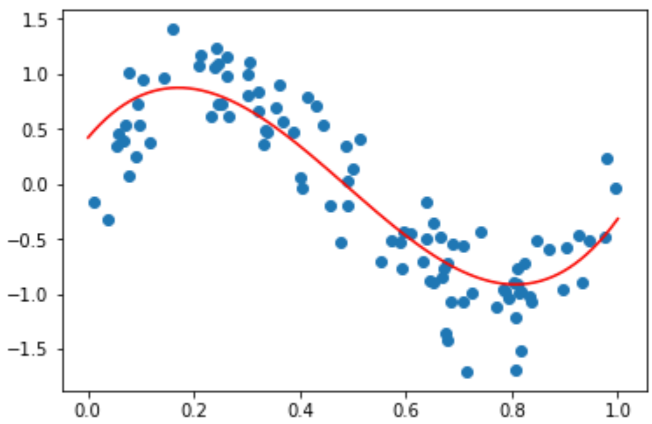
Lower values of RMSE indicate better fit. Thus from the plot we can conclude that α = 0.025 is the best fit because the test RMSE is minimum.

II. Number of data points = 100

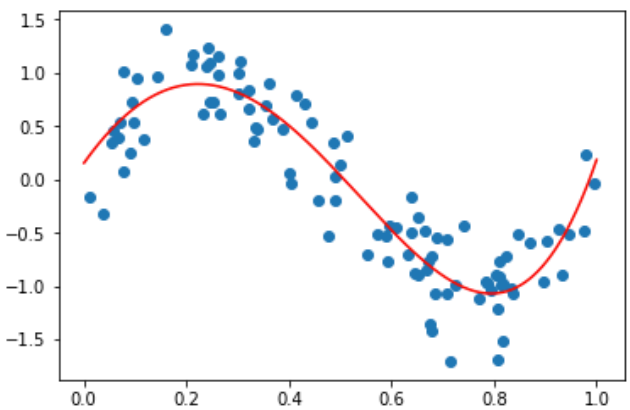
Y = β₀	***Y = β₀ + β₁*X***
	
Y = β₀ + β₁*X + β₂*X²	***Y = β₀ + β₁*X + β₂*X² + β₃*X³***



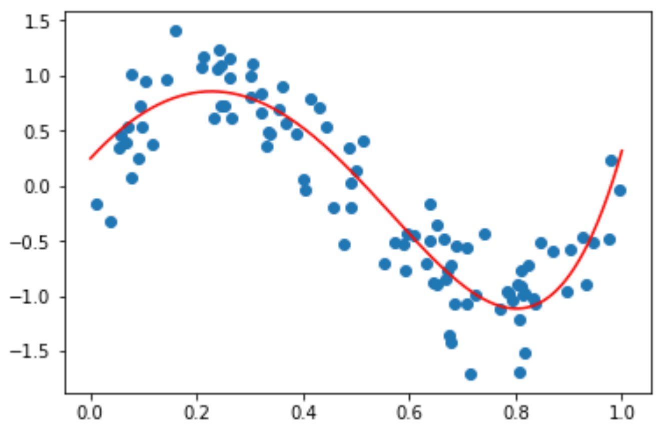
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4$



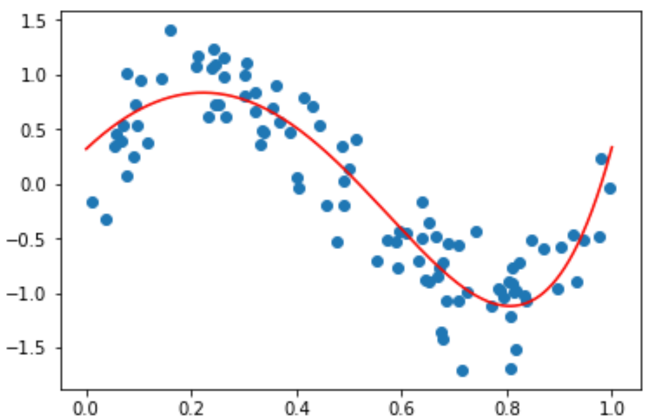
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5$



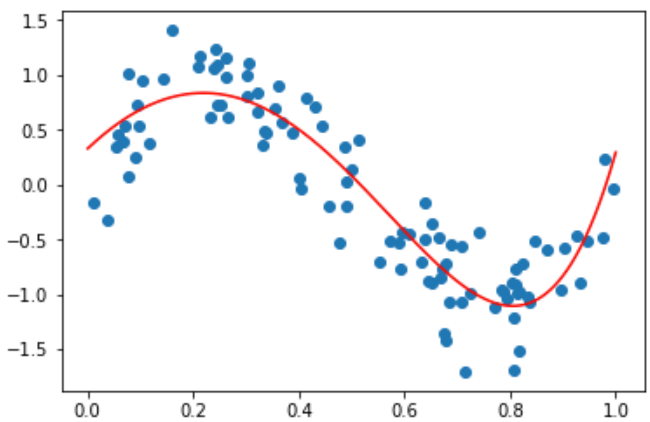
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6$



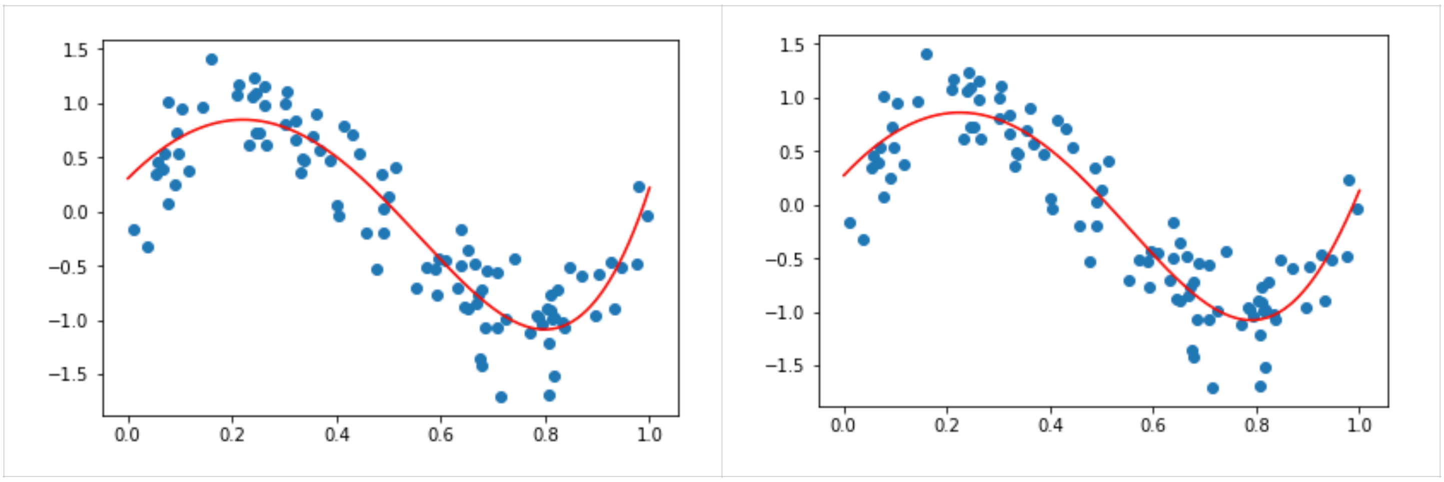
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7$



$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7 + \beta_8 * X^8$

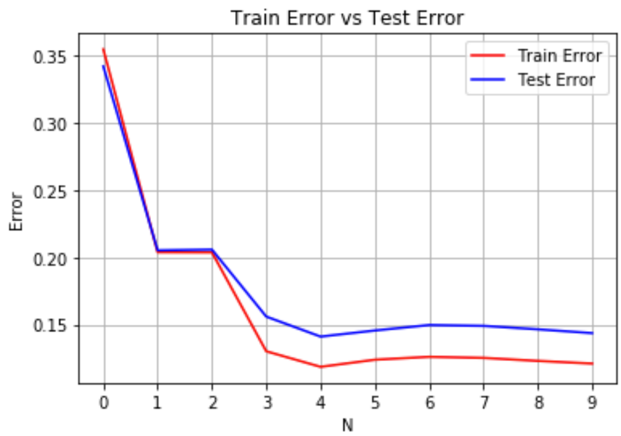


$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7 + \beta_8 * X^8 + \beta_9 * X^9$



Learned Value of the parameter i.e. beta matrix =
`array([ 0.15281478, 6.7259517 , -15.03001334, -3.87416105, 12.20769499])`

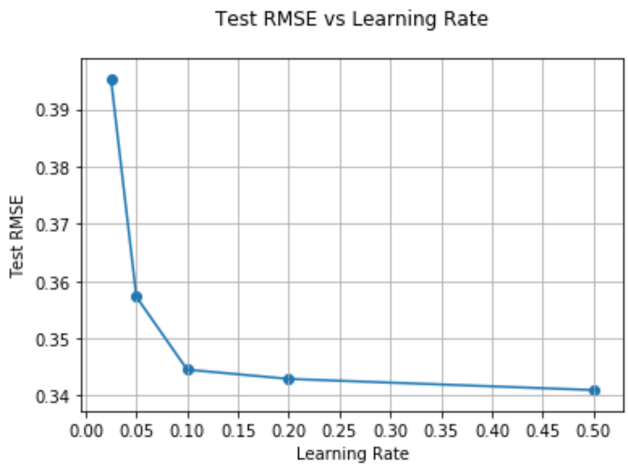
Train Error vs Test Error:



From the above plot we can conclude that N = 4 is the best fit because the test error is minimum. For N>4 there is overfitting and as the test error is increasing thus N=9 is ignored.

Therefore,
`Y = B0 + B1*X1 + B2*X2 + B3*X3 + B4*X4`

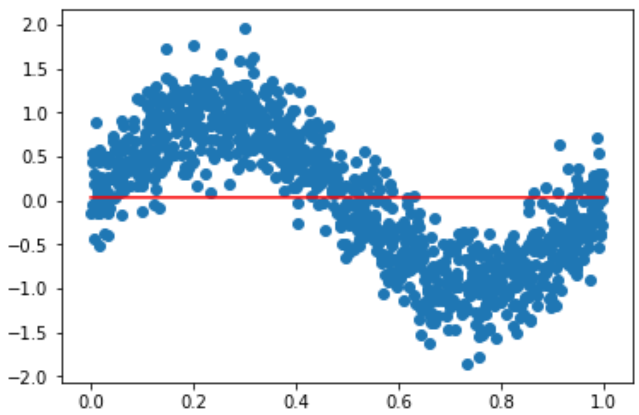
Test RMSE vs Learning Rate:



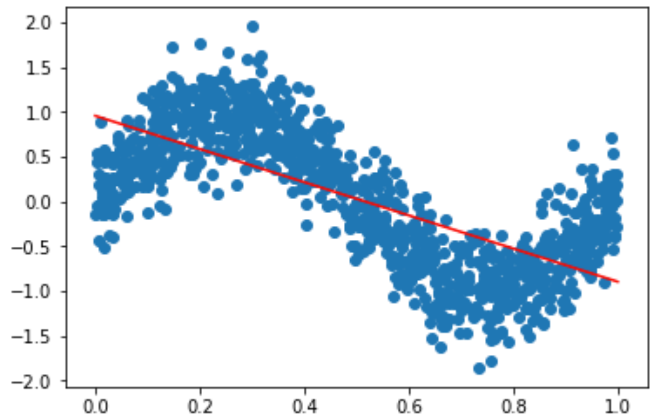
Lower values of RMSE indicate better fit. Thus from the plot we can conclude that α = 0.5 is the best fit because the test RMSE is minimum.

III. Number of data points = 1000

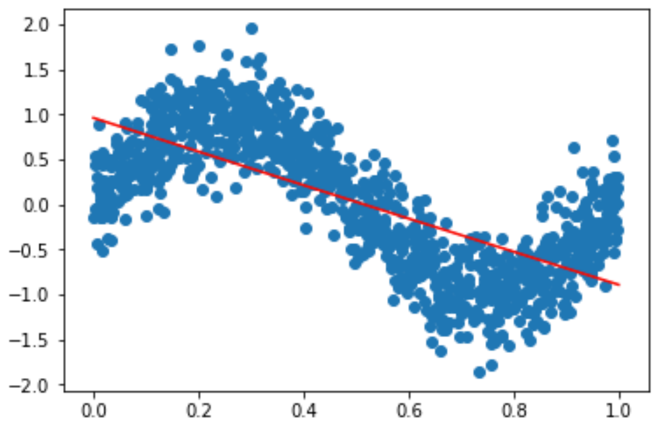
$Y = \beta_0$



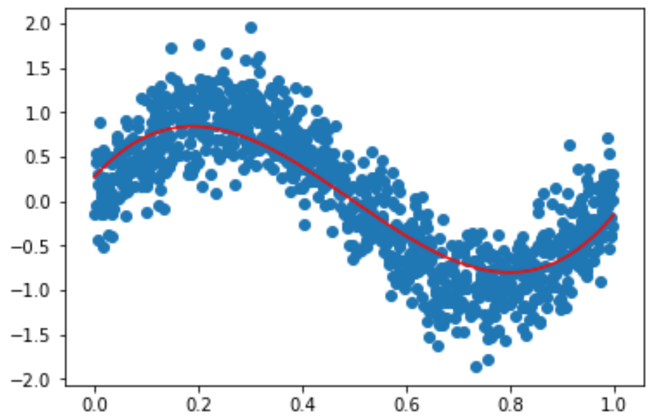
$Y = \beta_0 + \beta_1 * X$



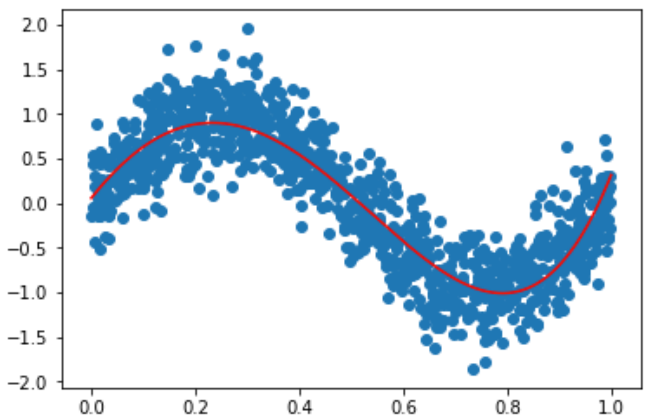
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2$



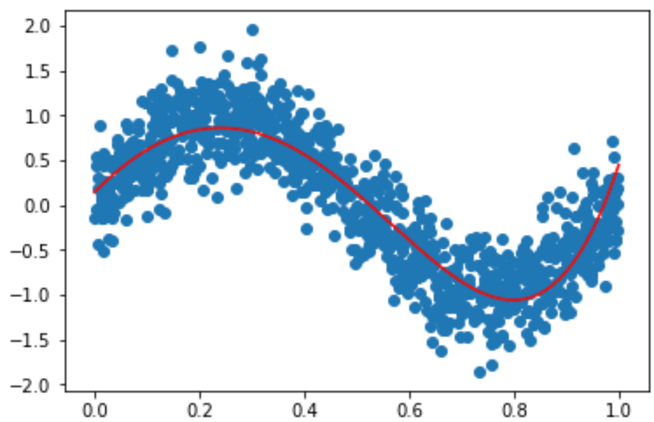
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3$



$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4$

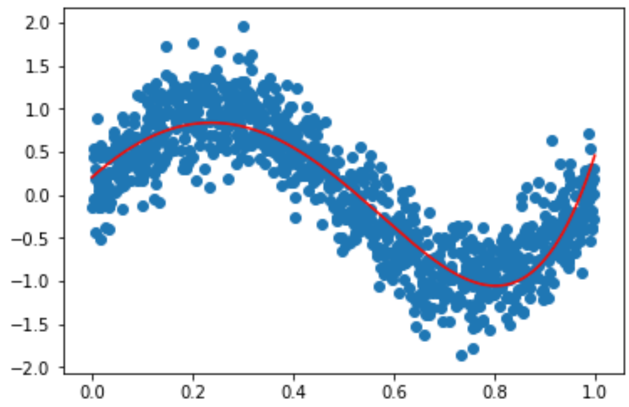


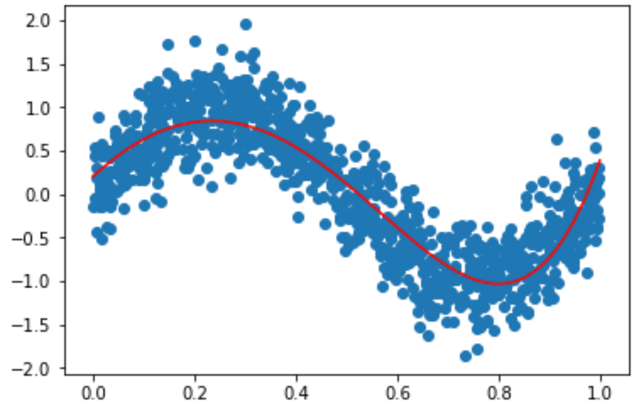
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5$



$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6$

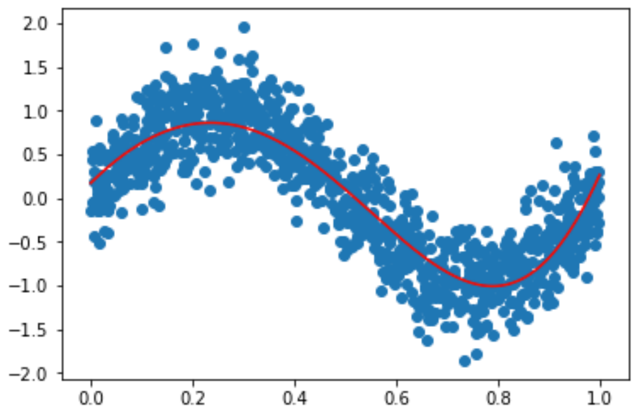
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7$

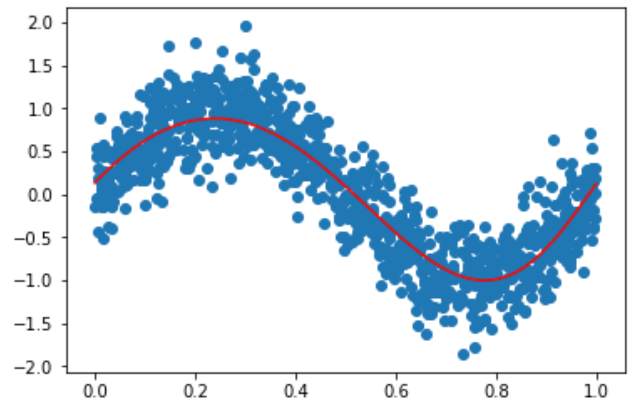




$Y = \beta_0 + \beta_1{*}X + \beta_2{*}X^2 + \beta_3{*}X^3 + \beta_4{*}X^4 + \beta_5{*}X^5 + \beta_6{*}X^6 + \beta_7{*}X^7 + \beta_8{*}X^8$

$Y = \beta_0 + \beta_1{*}X + \beta_2{*}X^2 + \beta_3{*}X^3 + \beta_4{*}X^4 + \beta_5{*}X^5 + \beta_6{*}X^6 + \beta_7{*}X^7 + \beta_8{*}X^8 + \beta_9{*}X^9$

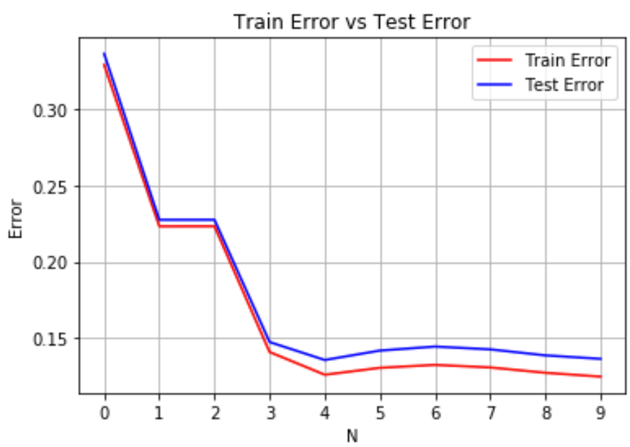




Learned Value of the parameter i.e. beta matrix =

```
array([ 0.05323679, 7.35433672, -15.7254371 , -4.07592419, 12.70868067])
```

Train Error vs Test Error:

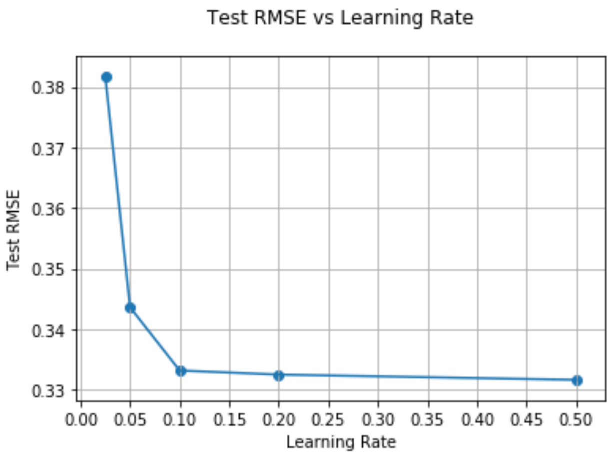


From the above plot we can conclude that N = 4 is the best fit because the test error is minimum. For N>4 there is overfitting and as the test error is increasing thus N=9 is ignored.

Therefore,

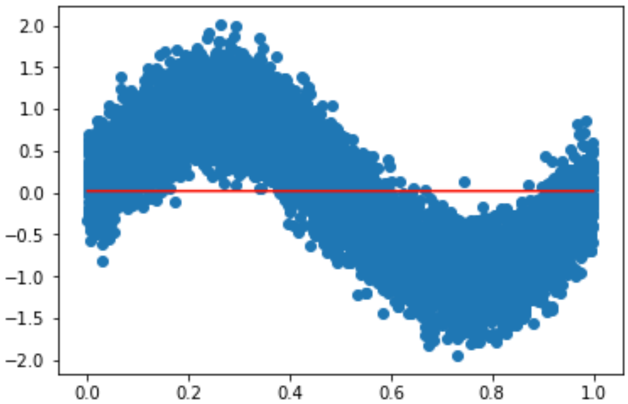
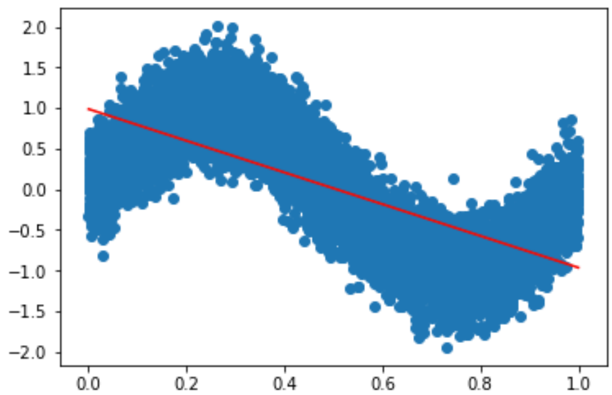
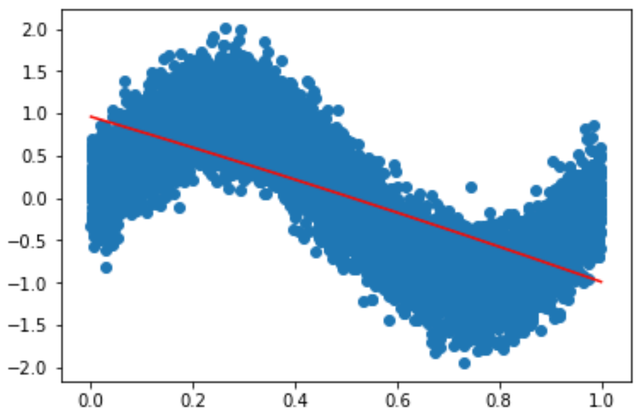
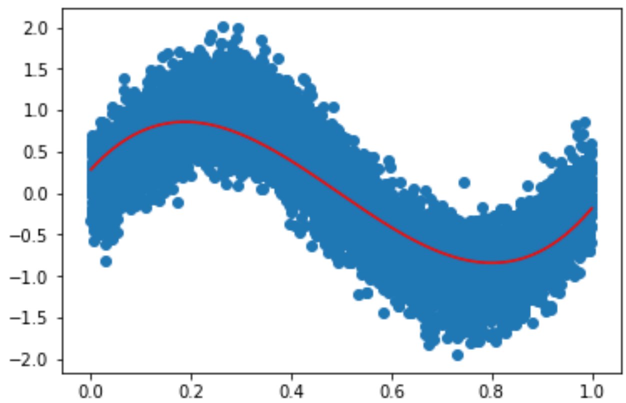
```
Y = B0 + B1*X1 + B2*X2 + B3*X3 + B4*X4
```

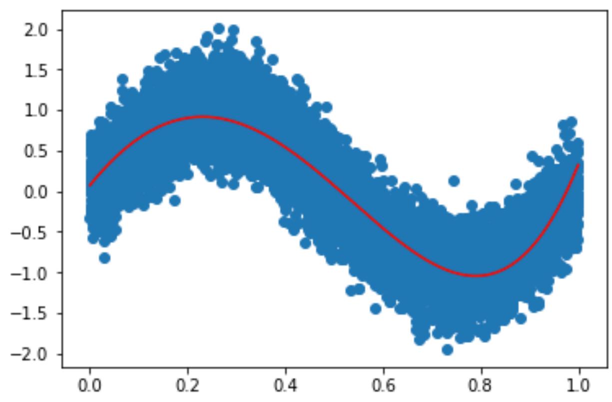
Test RMSE vs Learning Rate:



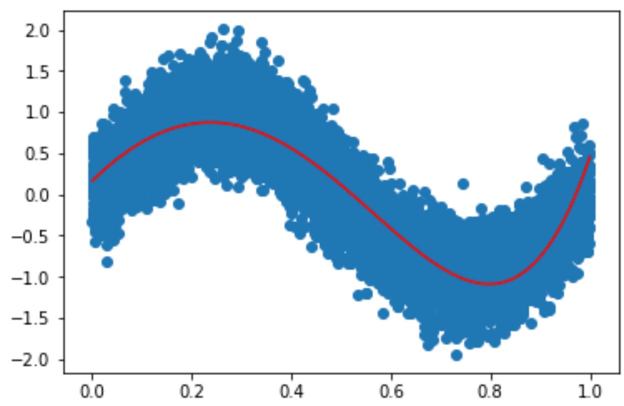
Lower values of RMSE indicate better fit.
Thus from the plot we can conclude that α = 0.5 is the best fit because the test RMSE is minimum.

IV. Number of data points = 10000

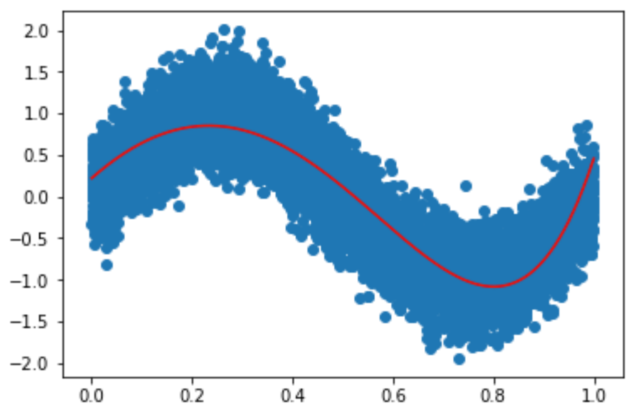
$Y = \beta_0$	$Y = \beta_0 + \beta_1 {*} X$
	
$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2$	$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2 + \beta_3 {*} X^3$
	
$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2 + \beta_3 {*} X^3 + \beta_4 {*} X^4$	$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2 + \beta_3 {*} X^3 + \beta_4 {*} X^4 + \beta_5 {*} X^5$



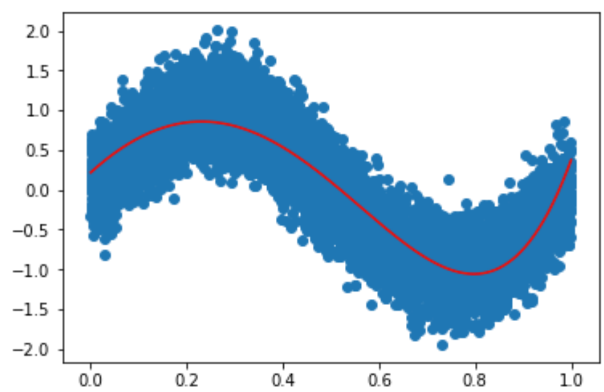
$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2 + \beta_3 {*} X^3 + \beta_4 {*} X^4 + \beta_5 {*} X^5 + \beta_6 {*} X^6$



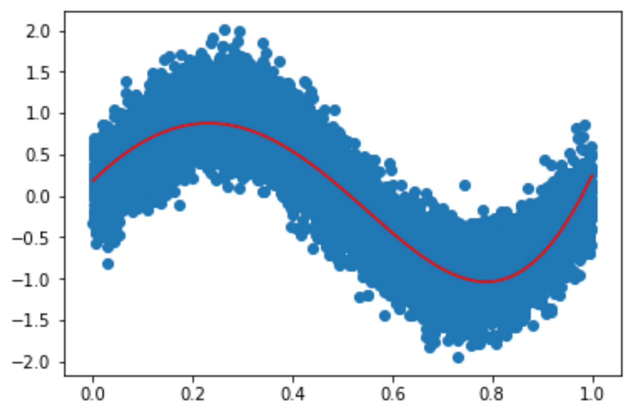
$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2 + \beta_3 {*} X^3 + \beta_4 {*} X^4 + \beta_5 {*} X^5 + \beta_6 {*} X^6 + \beta_7 {*} X^7$

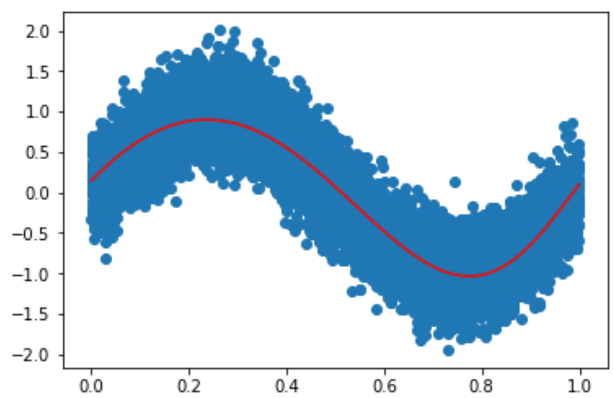


$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2 + \beta_3 {*} X^3 + \beta_4 {*} X^4 + \beta_5 {*} X^5 + \beta_6 {*} X^6 + \beta_7 {*} X^7 + \beta_8 {*} X^8$



$Y = \beta_0 + \beta_1 {*} X + \beta_2 {*} X^2 + \beta_3 {*} X^3 + \beta_4 {*} X^4 + \beta_5 {*} X^5 + \beta_6 {*} X^6 + \beta_7 {*} X^7 + \beta_8 {*} X^8 + \beta_9 {*} X^9$

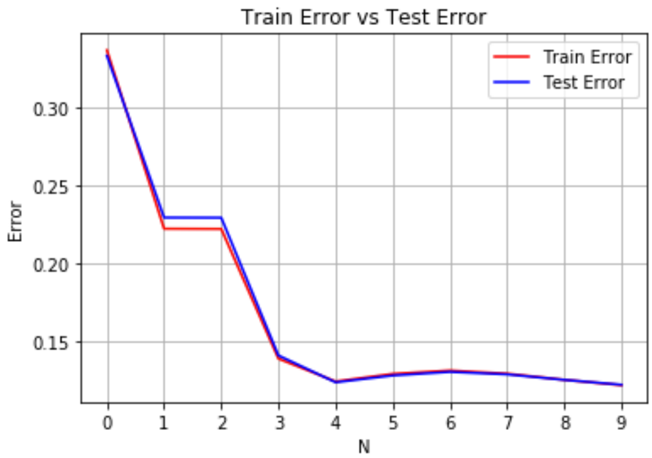




Learned Value of the parameter i.e. beta matrix =

```
array([ 0.07294845, 7.42442875, -16.10020293, -3.97589704, 12.90236626])
```

Train Error vs Test Error:

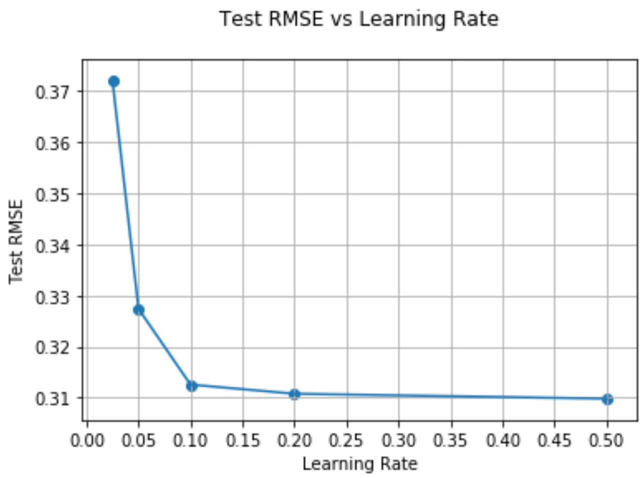


From the above plot we can conclude that N = 4 is the best fit because the test error is minimum. For N>4 there is overfitting and as the test error is increasing thus N=9 is ignored.

Therefore,

`Y = B0 + B1*X1 + B2*X2 + B3*X3 + B4*X4`

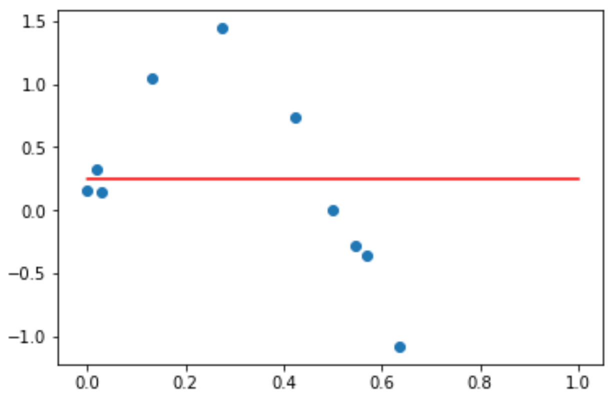
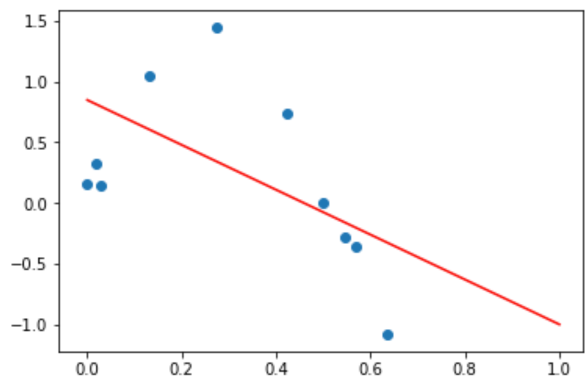
Test RMSE vs Learning Rate:

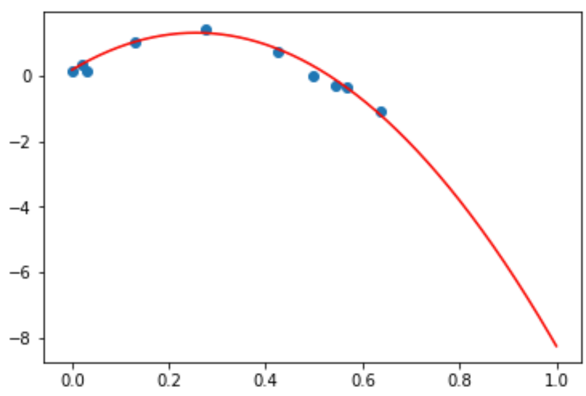


Lower values of RMSE indicate better fit. Thus from the plot we can conclude that α = 0.5 is the best fit because the test RMSE is minimum.

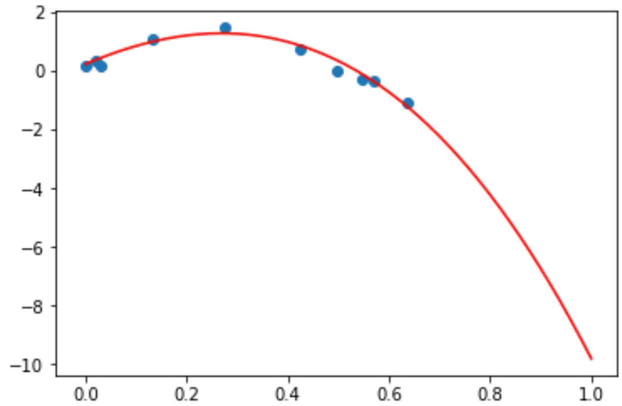
- **<u>Cost Function 3:</u> Fourth Power Error**

I. <u>Number of data points = 10</u>

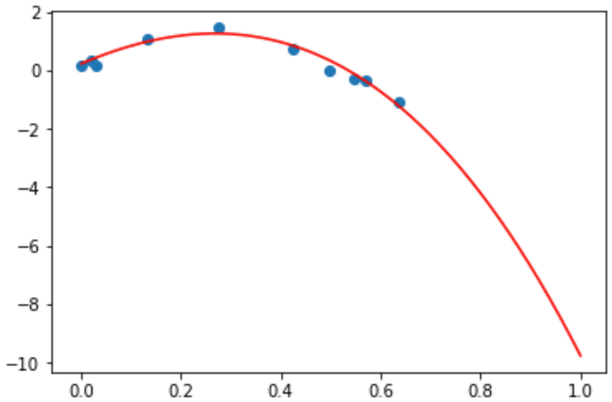
$Y = \beta_0$	$Y = \beta_0 + \beta_1{*}X$
	
$Y = \beta_0 + \beta_1{*}X + \beta_2{*}X^2$	$Y = \beta_0 + \beta_1{*}X + \beta_2{*}X^2 + \beta_3{*}X^3$



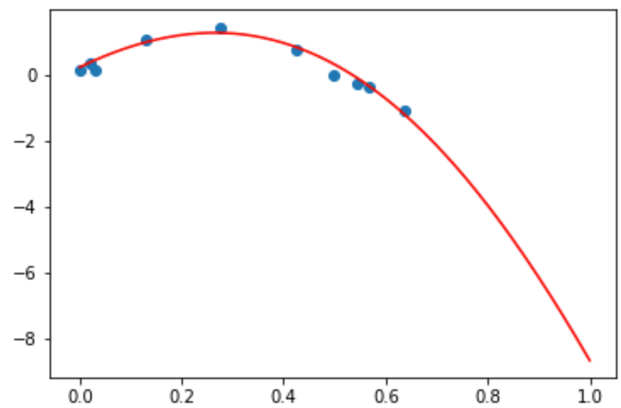
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4$



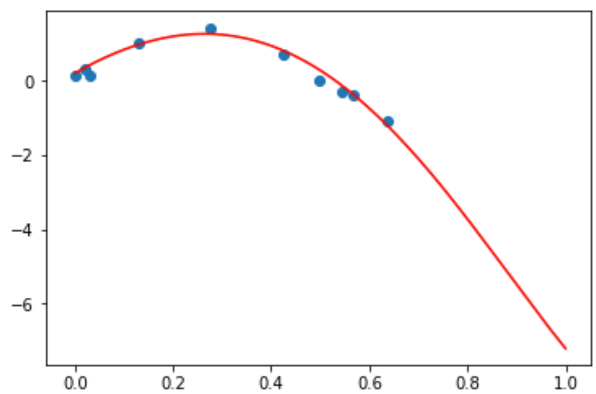
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5$



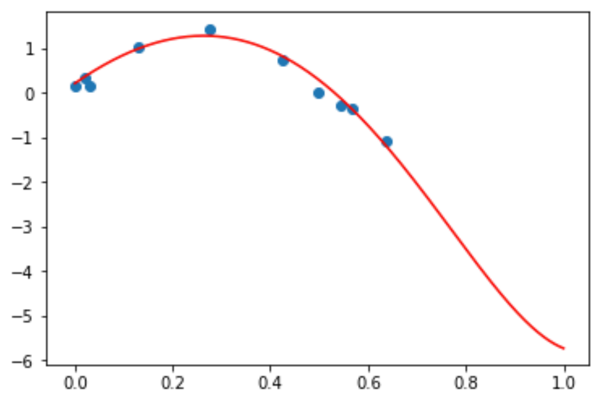
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6$



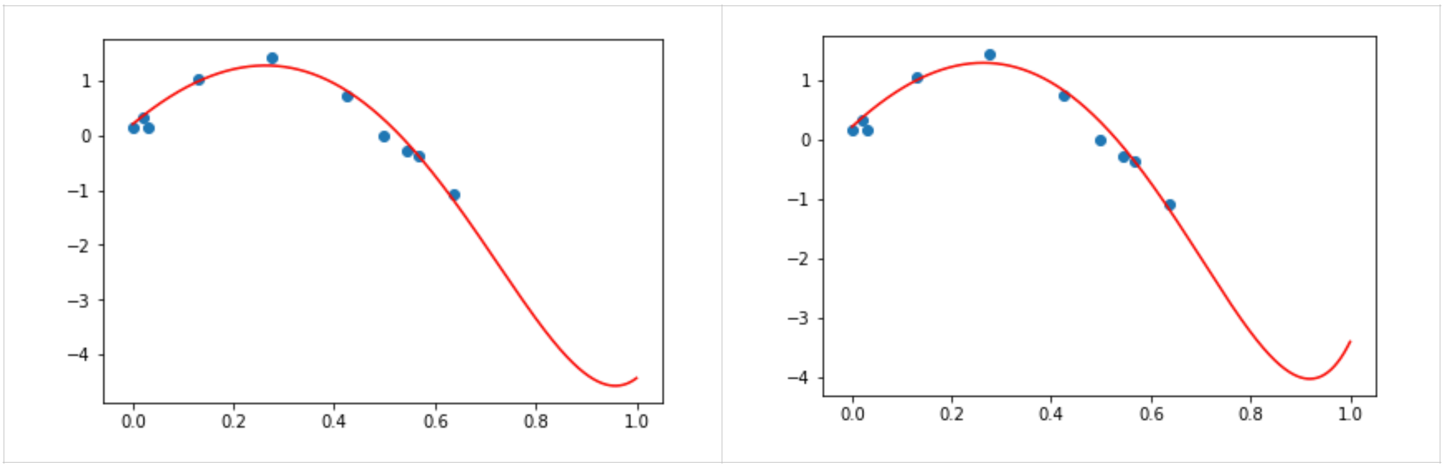
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7$



$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7 + \beta_8 * X^8$

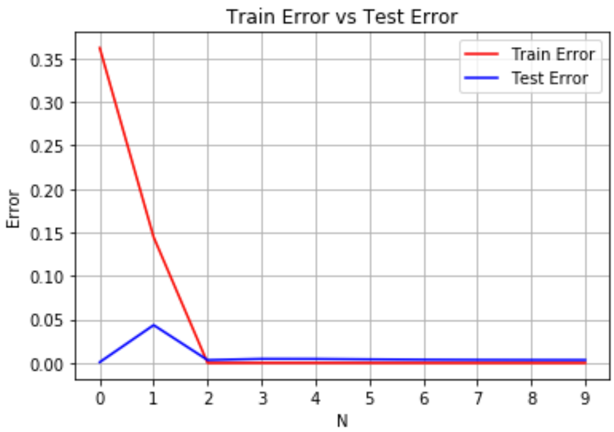


$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7 + \beta_8 * X^8 + \beta_9 * X^9$



Learned Value of the parameter i.e. beta matrix =
`array([ 0.17723536, 8.84116349, -17.27306056])`

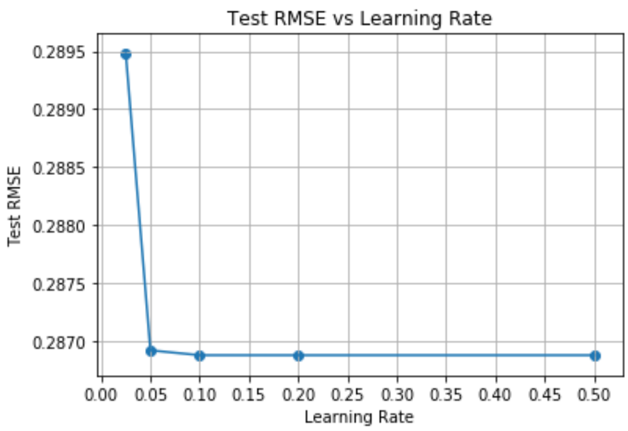
Train Error vs Test Error:



From the above plot we can conclude that N = 2 is the best fit because the test error is minimum, and for N>2 there is a slight increase in test error which can result into overfitting

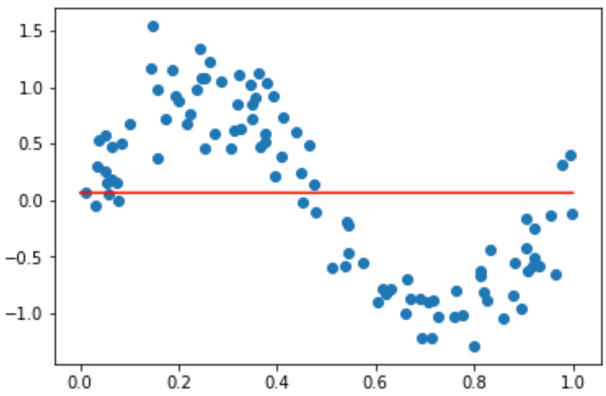
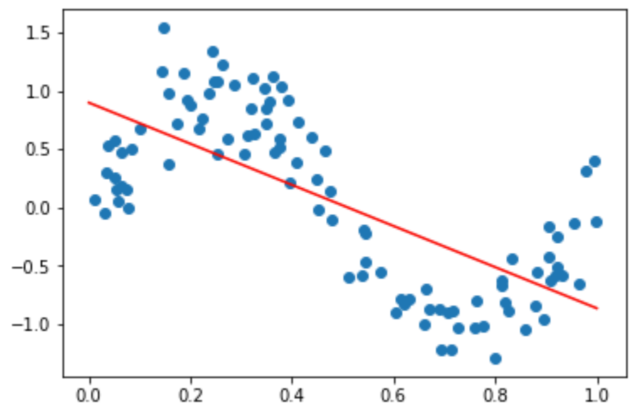
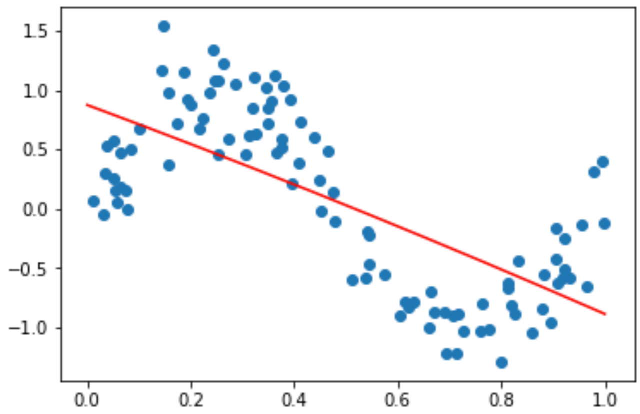
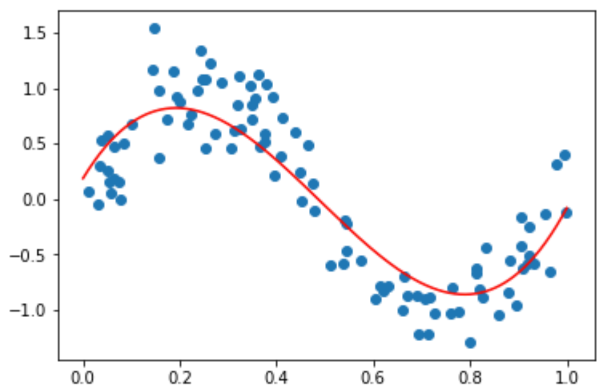
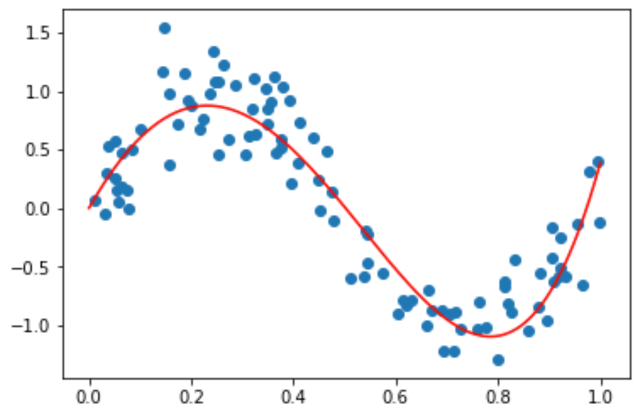
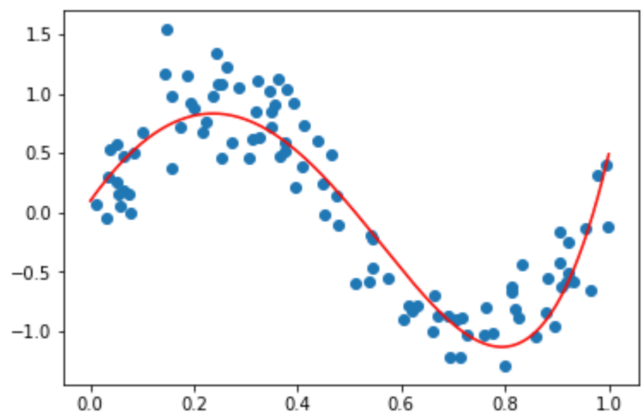
Therefore,
`Y = B0 + B1*X1 + B2*X2`

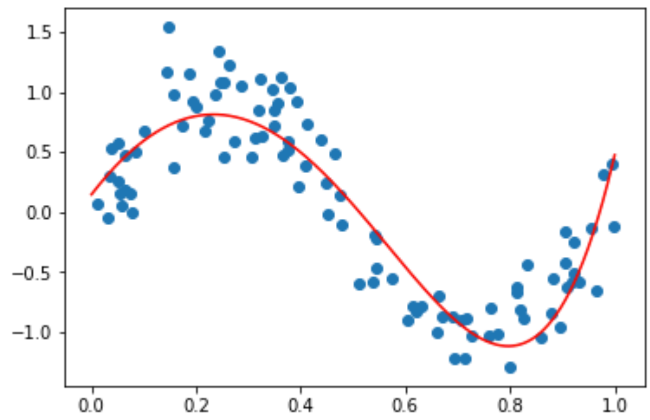
Test RMSE vs Learning Rate:



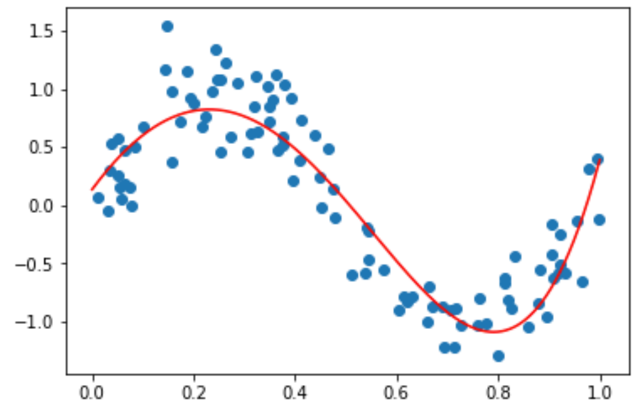
Lower values of RMSE indicate better fit.
Thus from the plot we can conclude that α >= 0.1 is the best fit because the test RMSE is minimum.

II. Number of data points = 100

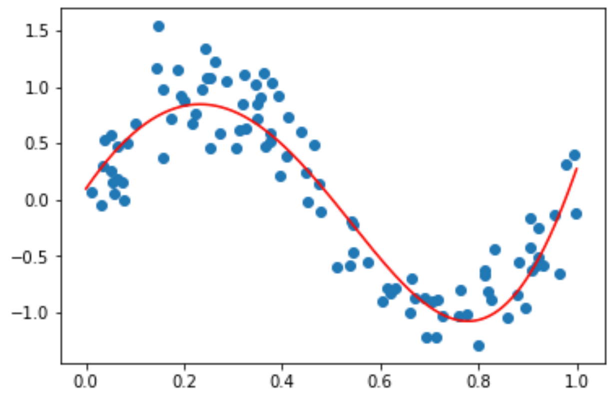
$Y = \beta_0$	$Y = \beta_0 + \beta_1{*}X$
	
$Y = \beta_0 + \beta_1{*}X + \beta_2{*}X^2$	$Y = \beta_0 + \beta_1{*}X + \beta_2{*}X^2 + \beta_3{*}X^3$
	
$Y = \beta_0 + \beta_1{*}X + \beta_2{*}X^2 + \beta_3{*}X^3 + \beta_4{*}X^4$	$Y = \beta_0 + \beta_1{*}X + \beta_2{*}X^2 + \beta_3{*}X^3 + \beta_4{*}X^4 + \beta_5{*}X^5$
	
$Y = \beta_0 + \beta_1{*}X + \beta_2{*}X^2 + \beta_3{*}X^3 + \beta_4{*}X^4 + \beta_5{*}X^5 + \beta_6{*}X^6$	$Y = \beta_0 + \beta_1{*}X + \beta_2{*}X^2 + \beta_3{*}X^3 + \beta_4{*}X^4 + \beta_5{*}X^5 + \beta_6{*}X^6 + \beta_7{*}X^7$

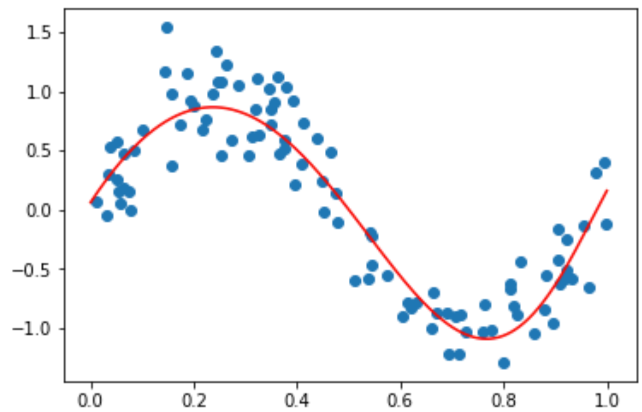


$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_4 X^4 + \beta_5 X^5 + \beta_6 X^6 + \beta_7 X^7 + \beta_8 X^8$$



$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_4 X^4 + \beta_5 X^5 + \beta_6 X^6 + \beta_7 X^7 + \beta_8 X^8 + \beta_9 X^9$$





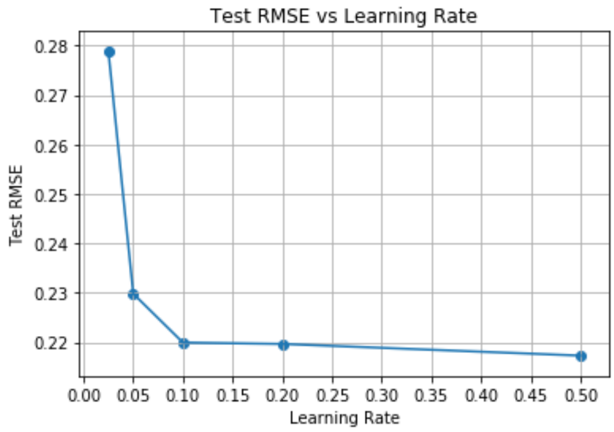
Learned Value of the parameter i.e. beta matrix =
```
array([-1.72626294e-03, 7.71462139e+00, -1.67980087e+01, -3.81813158e+00,
1.32889974e+01])
```

Train Error vs Test Error:

From the above plot we can conclude that N = 4 is the best fit because the test error is minimum. For N>4 there is overfitting and as the test error is increasing thus N=9 is ignored.

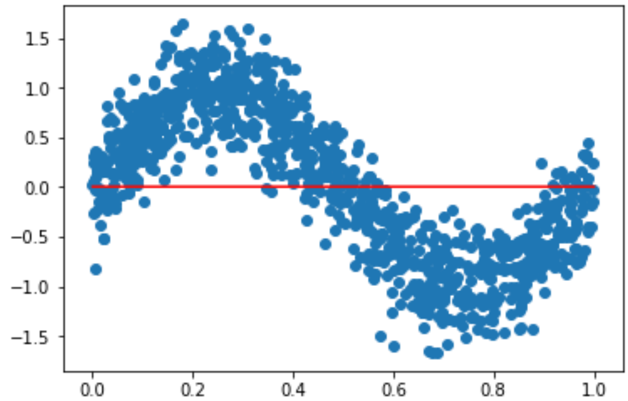
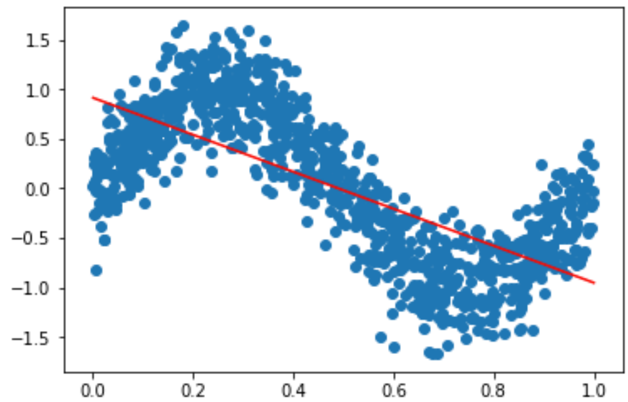
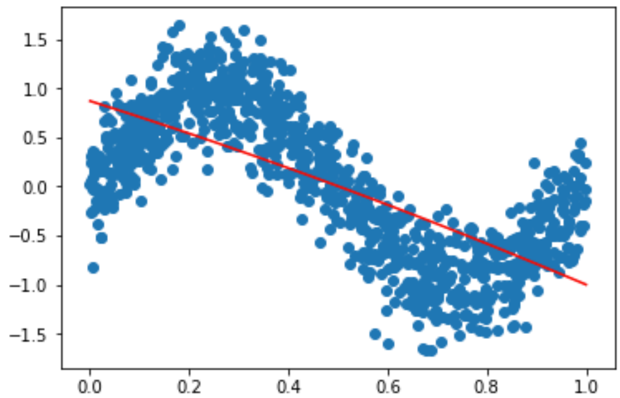
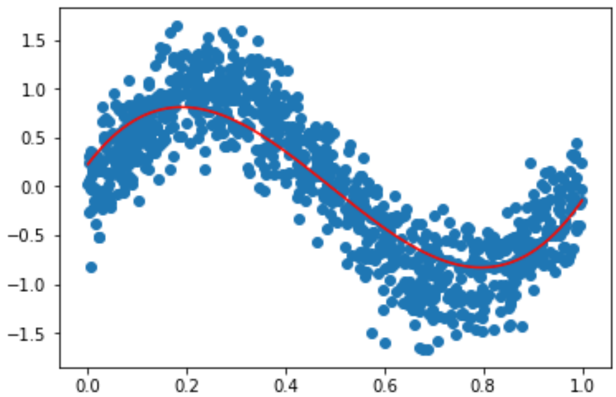
Therefore,
```
Y = B0 + B1*X1 + B2*X2 + B3*X3 + B4*X4
```

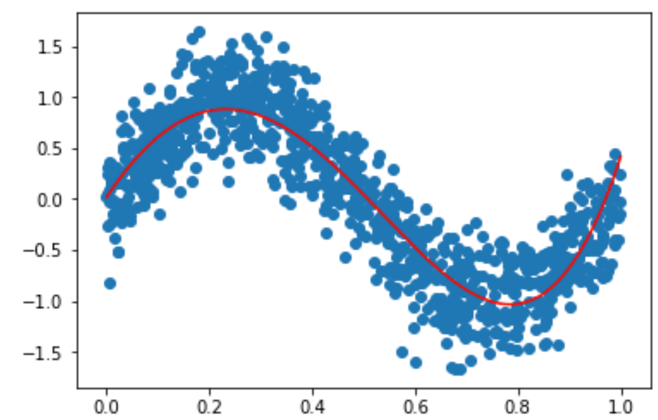
Test RMSE vs Learning Rate:

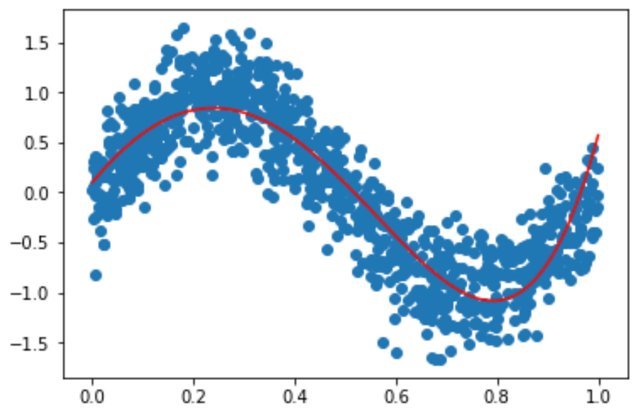


Lower values of RMSE indicate better fit.
Thus from the plot we can conclude that α = 0.5 is the best fit because the test RMSE is minimum.

III. Number of data points = 1000

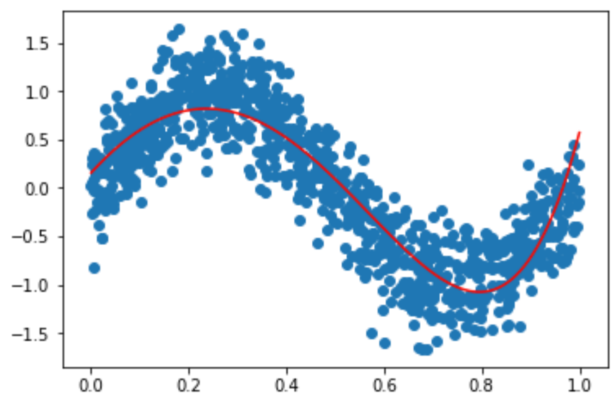
$Y = \beta_0$	$Y = \beta_0 + \beta_1*X$
	
$Y = \beta_0 + \beta_1*X + \beta_2*X^2$	$Y = \beta_0 + \beta_1*X + \beta_2*X^2 + \beta_3*X^3$
	
$Y = \beta_0 + \beta_1*X + \beta_2*X^2 + \beta_3*X^3 + \beta_4*X^4$	$Y = \beta_0 + \beta_1*X + \beta_2*X^2 + \beta_3*X^3 + \beta_4*X^4 + \beta_5*X^5$

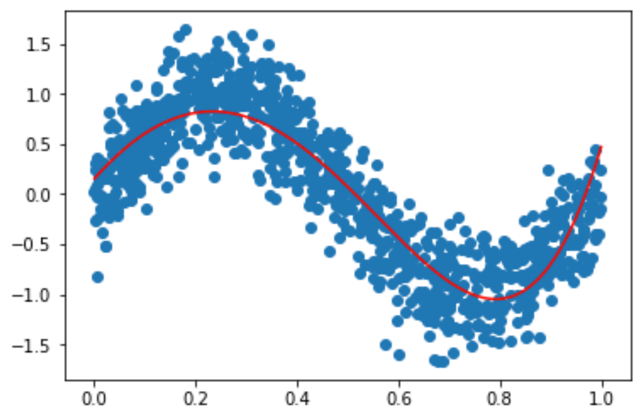




$$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6$$

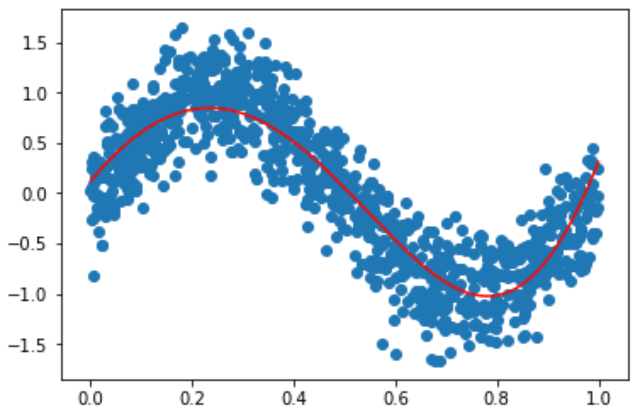
$$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7$$

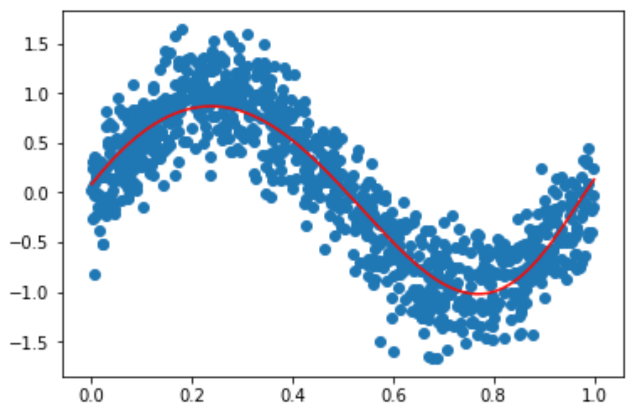




$$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7 + \beta_8 * X^8$$

$$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7 + \beta_8 * X^8 + \beta_9 * X^9$$

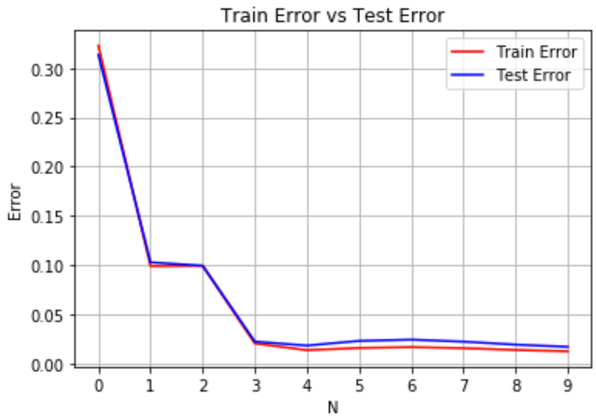




Learned Value of the parameter i.e. beta matrix =

```
array([ 0.02099019, 7.54683013, -16.36958272, -3.75137485, 12.96764109])
```

Train Error vs Test Error:

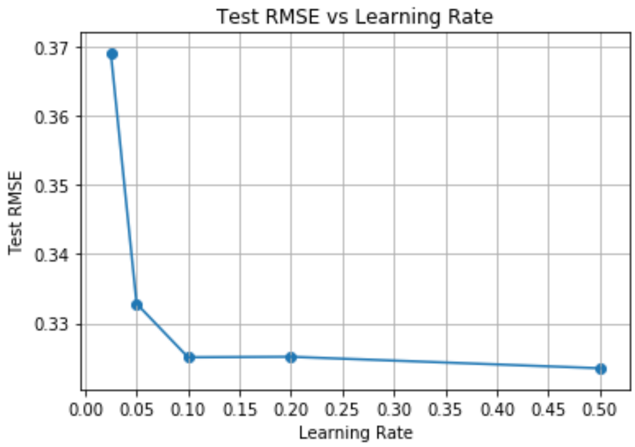


From the above plot we can conclude that N = 4 is the best fit because the test error is minimum. For N>4 there is overfitting and as the test error is increasing thus N=9 is ignored.

Therefore,

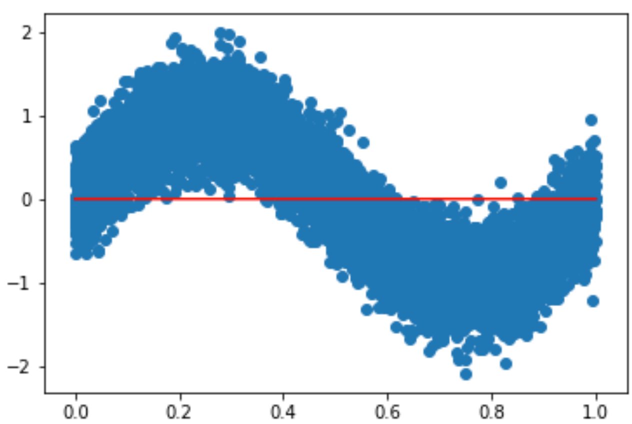
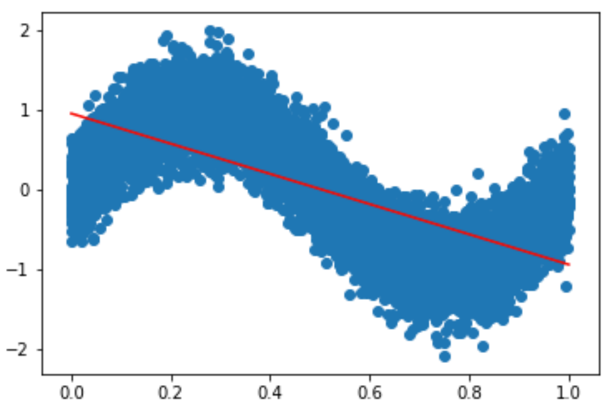
`Y = B0 + B1*X1 + B2*X2 + B3*X3 + B4*X4`

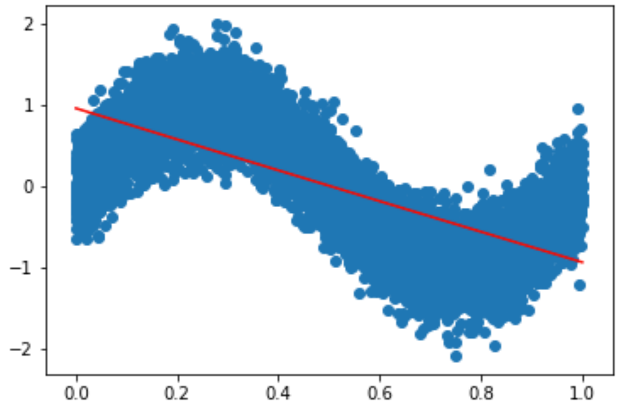
Test RMSE vs Learning Rate:



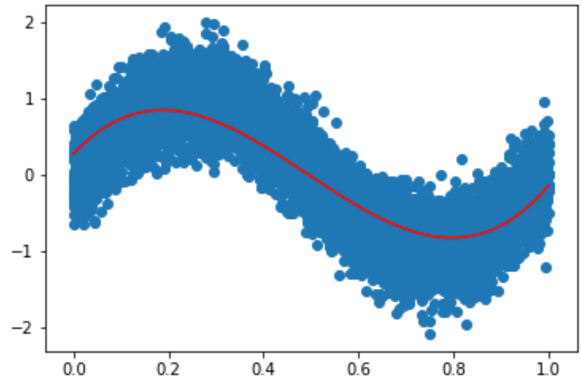
Lower values of RMSE indicate better fit. Thus from the plot we can conclude that α = 0.5 is the best fit because the test RMSE is minimum.

IV. Number of data points = 10000

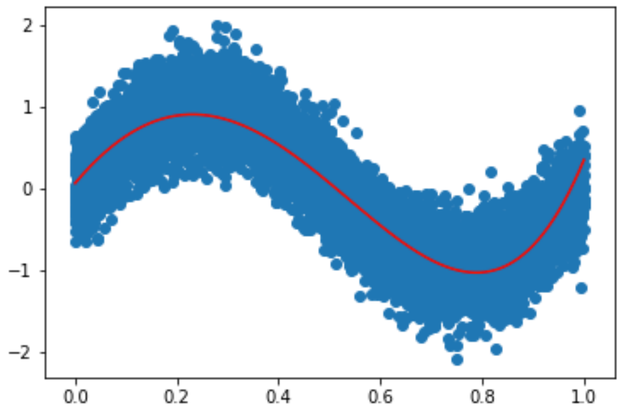
$Y = \beta_0$	$Y = \beta_0 + \beta_1{*}X$
	
$Y = \beta_0 + \beta_1{*}X + \beta_2{*}X^2$	$Y = \beta_0 + \beta_1{*}X + \beta_2{*}X^2 + \beta_3{*}X^3$



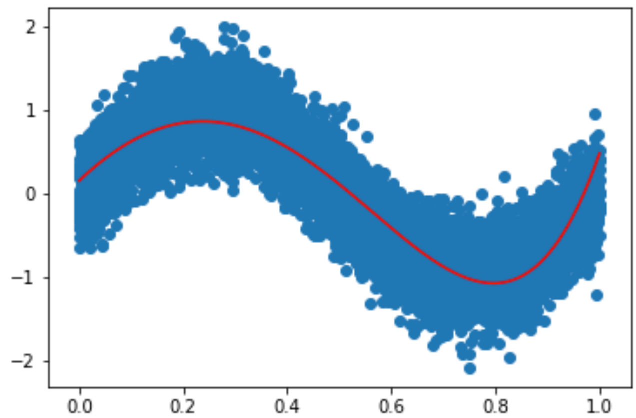
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4$



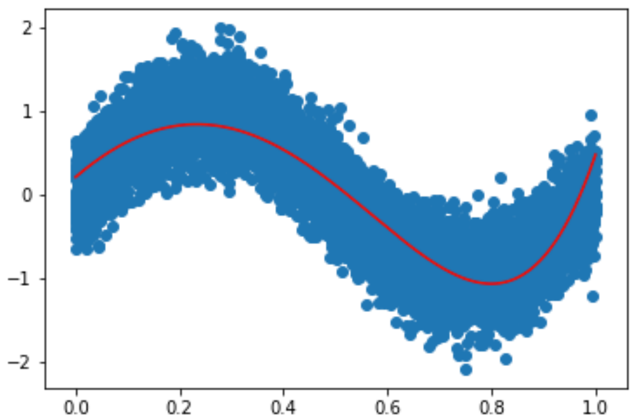
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5$



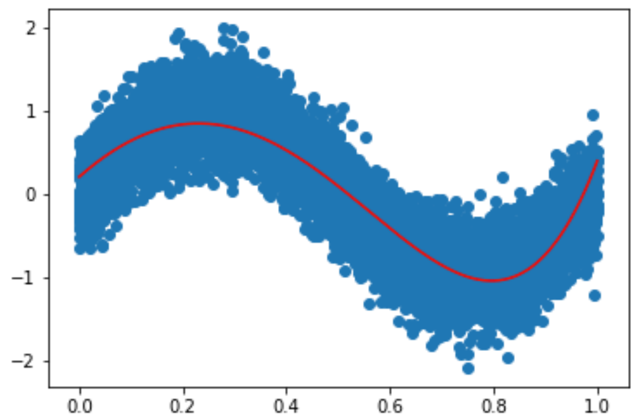
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6$



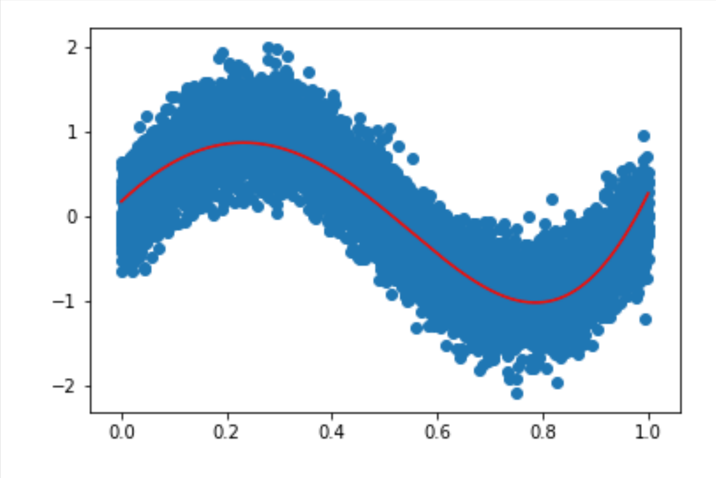
$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7$

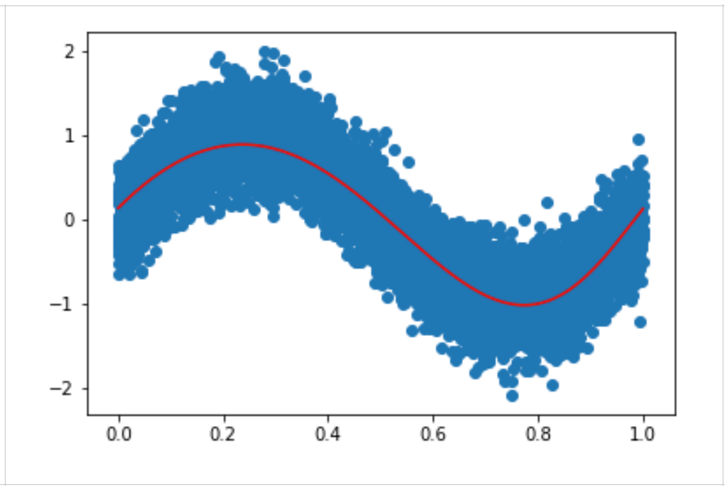


$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7 + \beta_8 * X^8$



$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 + \beta_5 * X^5 + \beta_6 * X^6 + \beta_7 * X^7 + \beta_8 * X^8 + \beta_9 * X^9$

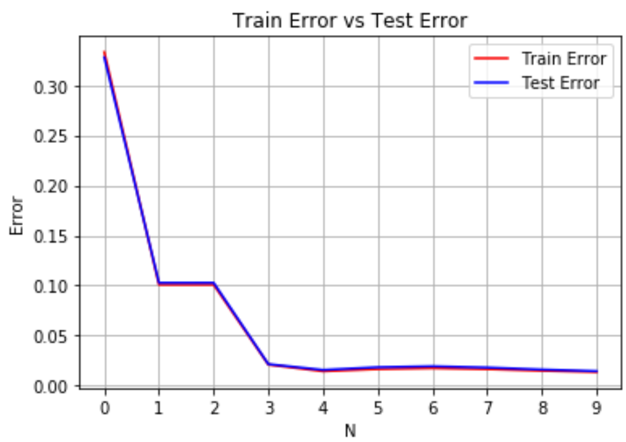




Learned Value of the parameter i.e. beta matrix =

`array([ 0.05738343, 7.40843018, -16.08139841, -3.84373336, 12.80301762])`

Train Error vs Test Error:

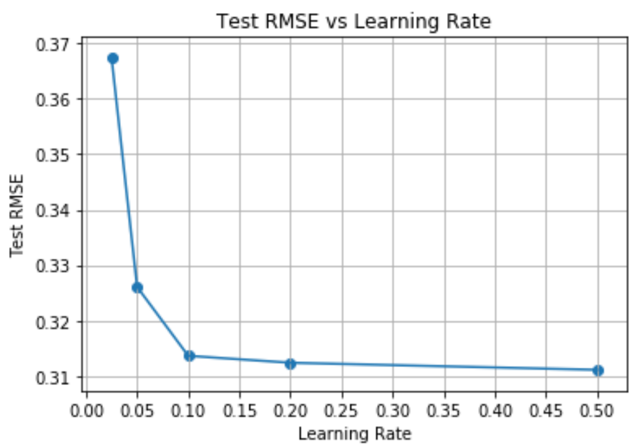


From the above plot we can conclude that N = 4 is the best fit because the test error is minimum. For N>4 there is overfitting and as the test error is increasing thus N=9 is ignored.

Therefore,

`Y = B0 + B1*X1 + B2*X2 + B3*X3 + B4*X4`

Test RMSE vs Learning Rate:

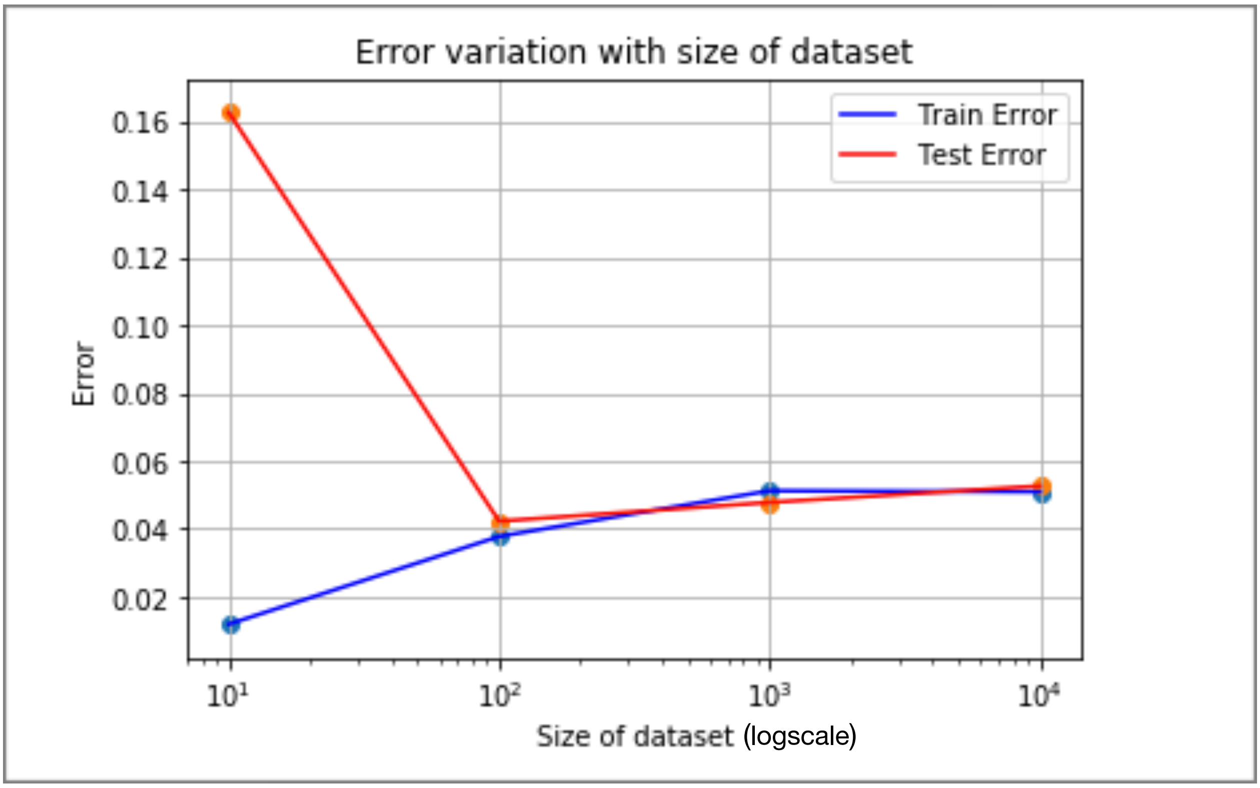


Lower values of RMSE indicate better fit. Thus from the plot we can conclude that α = 0.5 is the best fit because the test RMSE is minimum.

The learning curve of how train and test error varies with increase in size of datasets:

Considering n=4 for all the four datasets, viz. size = 10, 100, 1000, 10000

Thus, the variation of training error and test error is as follows:



From the curve it can be seen that, as the size of data set is increasing the test error is decreasing. Thus, we can note that having a large dataset is beneficial for a model.