INTRODUCTION TO MACHINE LEARNING

ASSIGNMENT-3

REPORT

Submitted By:

Komal Arya(2022EEZ8483)

POOJA (2022PHM2608)

RIMPI BORAH(2022EEZ8212)

Objective: Finding the best suit Polynomial which fits the dataset by applying the concept of Linear Regression for Polynomial Curve Fitting.

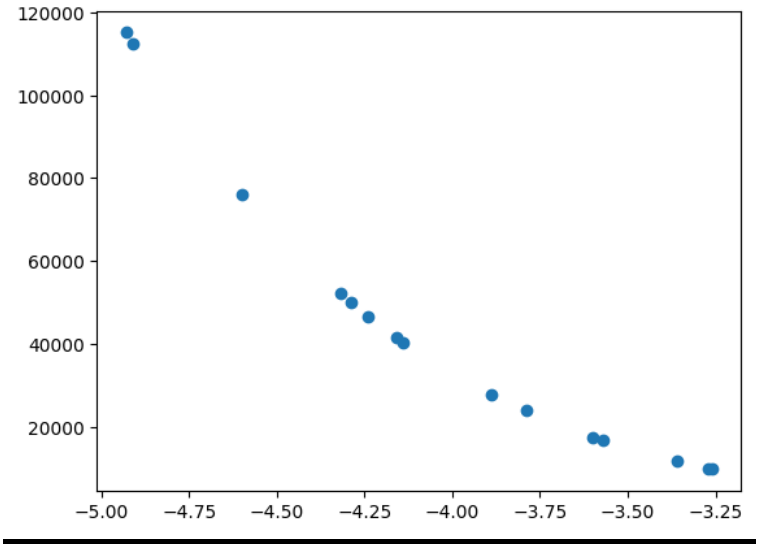
For this regression we have taken Huber Loss function which constitutes of both Mean squared Error and Mean Absolute Error depending on the value of error. It is basically quadratic (MSE) when the error is small else MAE. The function can be defined as below:

$$L_\delta(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2 & \text{for} |y - f(x)| \leq \delta, \\ \delta\,|y - f(x)| - \frac{1}{2}\delta^2 & \text{otherwise.} \end{cases}$$

Where δ =50(TAKEN) is called hyper parameter which is used as a threshold in order to distinguish between MSE and MAE. We find that when the error is less than delta, the error is quadratic else it is absolute. Below sample code shows the implementation of the Huber loss.

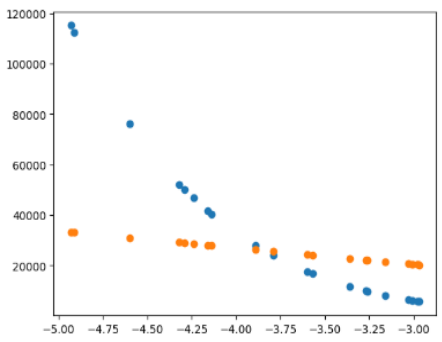
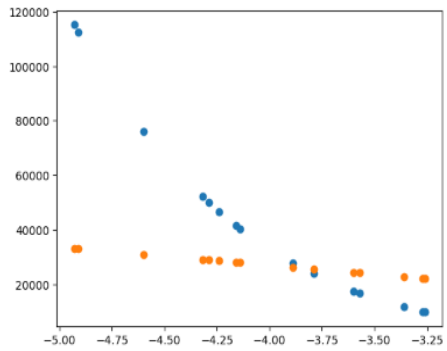
For First 20 Points

Data plot for 20 points:



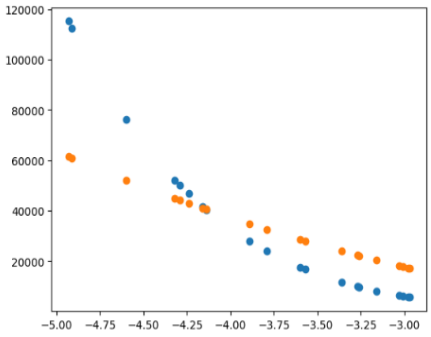
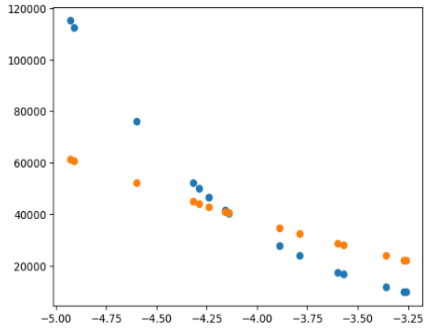
Degree = 1

alpha = 0.01, Noise variance = 2.9109834226201952e

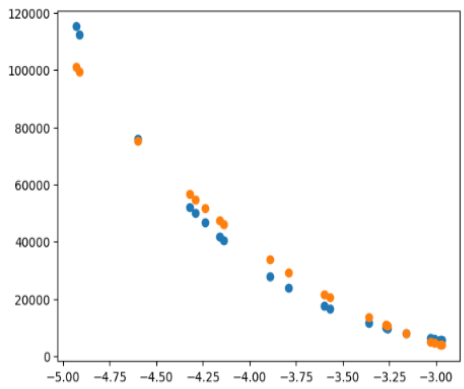
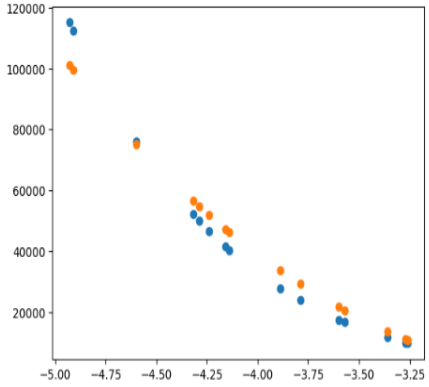
Normal	Testing
	

Degree = 2

alpha = 0.01, Noise variance = 2.0818018653658243e-09

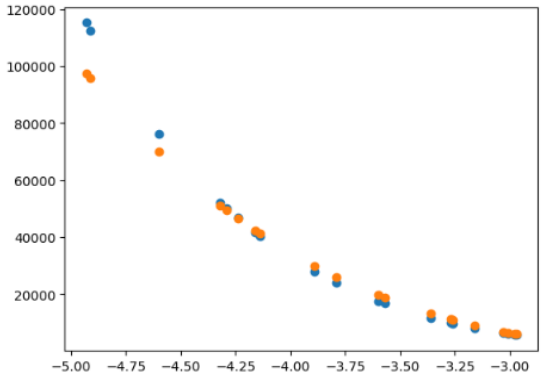
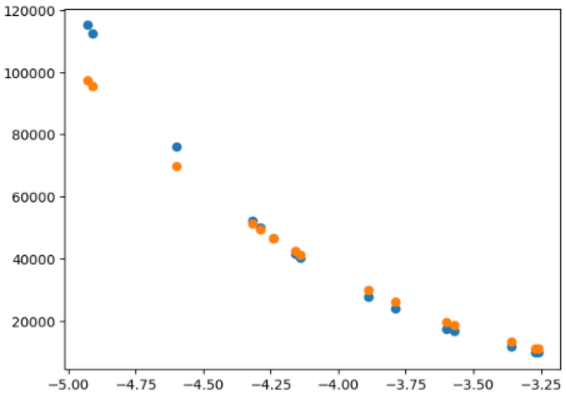
Normal	Testing
	

Degree = 3, alpha = 0.01, Noise variance = 4.027164668589359e-08

Normal	Testing
	

Degree = 4

alpha = 0.0001, Noise variance = 2.894996284099536e-08

Normal	Testing
	

- From the above results, noise variance is increasing so we can conclude that model is being overfitted.
- From the observations, we got to know that test error is being increased after 3rd degree polynomial, so we can conclude that after this degree model will start to overfit.

REGULARIZATION

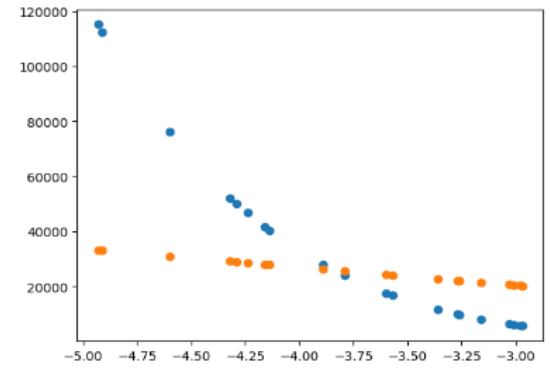
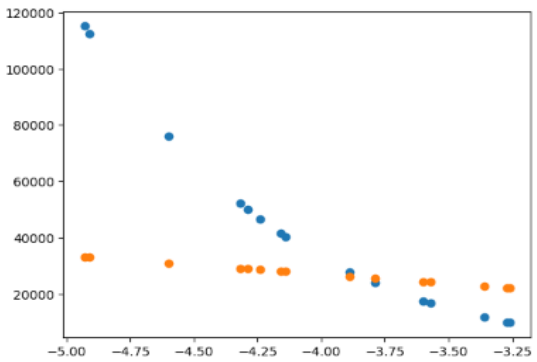
Machine learning model sometimes performs well with the training data but does not perform well with the test data. It means the model is not able to predict the output when deals with unseen data by introducing noise in the output, and hence the model is called over fitted. This problem can be deal with the help of a regularization technique.

Regularisation Parameter: 0.05

Regularization For 20 points

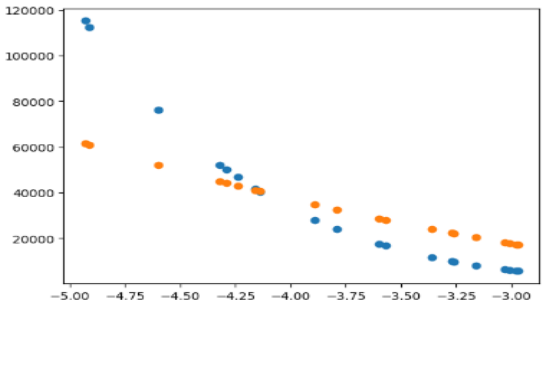
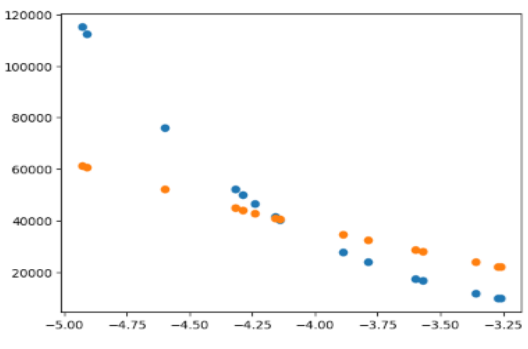
Degree = 1

alpha= 0.01, Noise variance = 2.9103088696396 77e-09

Normal	Testing
	

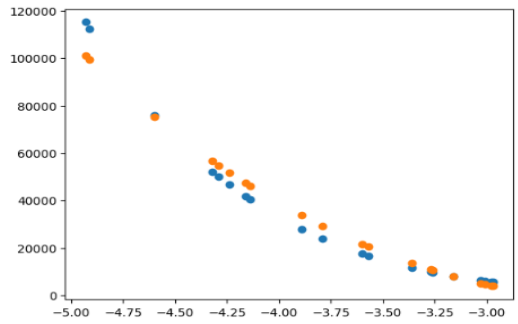
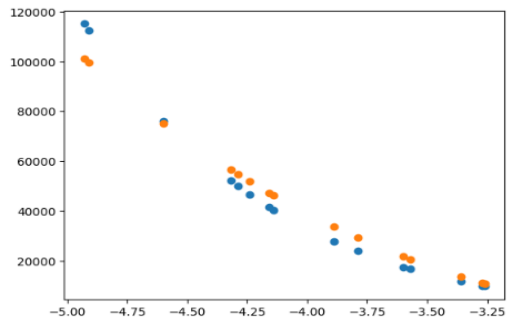
Degree = 2

alpha= 0.01, Noise variance = 3.2464592578899734e-09

Normal	Testing
	

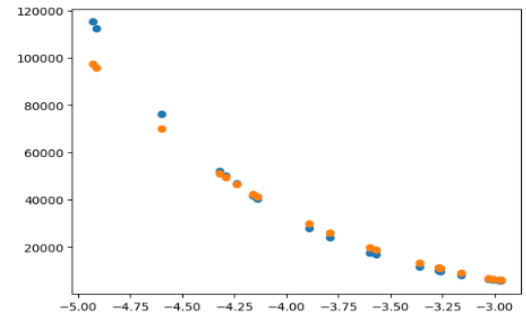
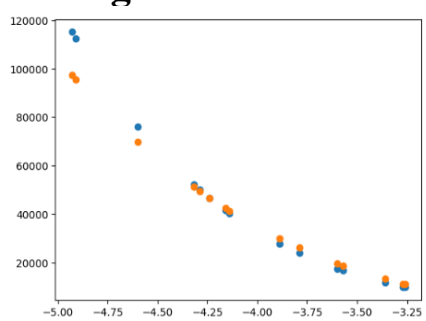
Degree = 3

alpha= 0.01, Noise variance = 4.024770127209025e-08

Normal	Testing
	

Degree = 4

alpha= 0.0001, Noise variance = 2.8948116212850866e-08

Normal	Testing
	

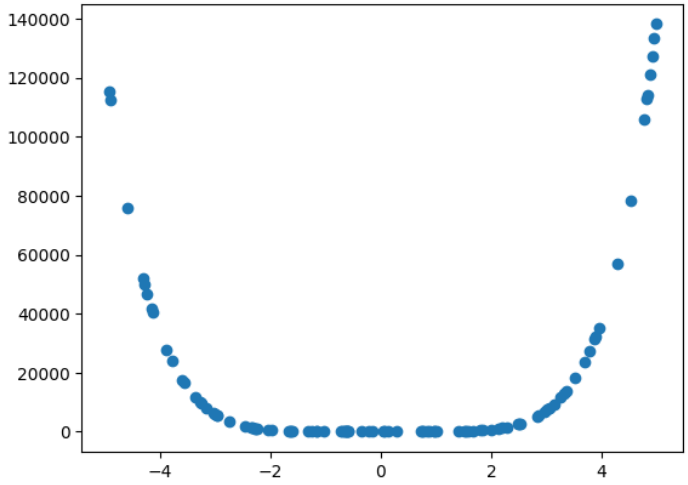
The noise variance which we obtained without regularisation is high compare to with regularisation so we confirm that regularisation term helps the model to not become overfitted.

Results for 20 Points:

- Therefore 3rd degree polynomial is best fit.
- Underfit model is 1st degree
- Overfit model has degree greater than and equal 4.

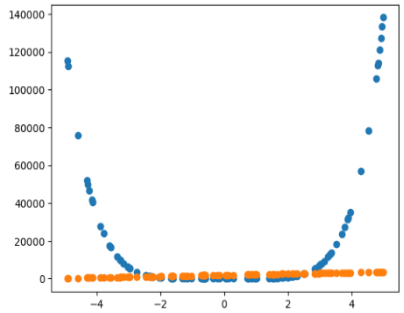
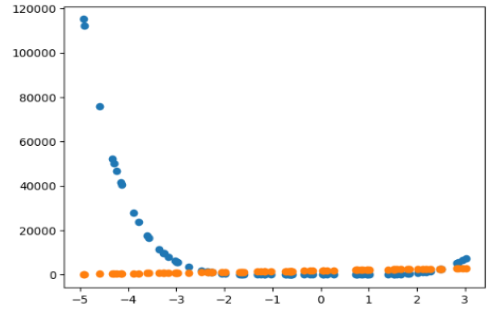
For 100 data Points

Data plot for 100 points:



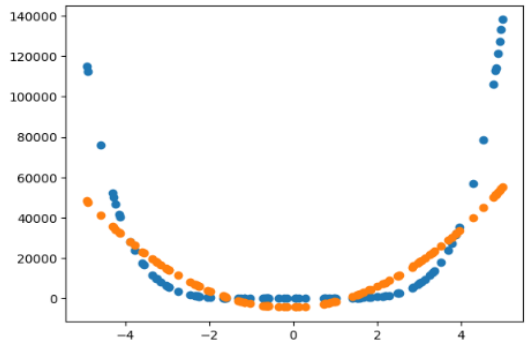
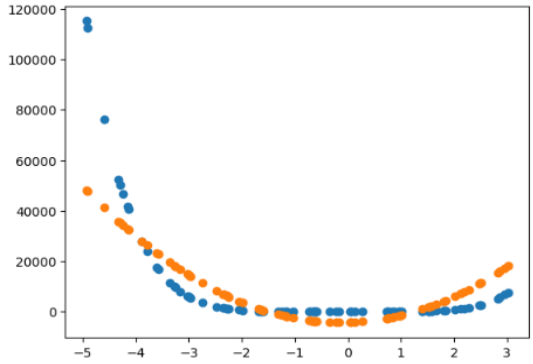
Degree = 1

alpha= 0.01, Noise variance = 1.7649088607961262e-12

Normal	Testing
	

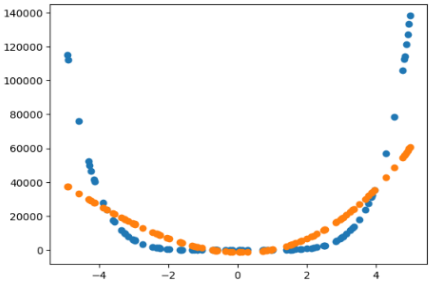
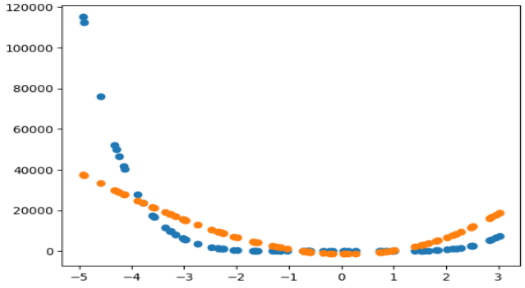
Degree = 2

alpha= 0.01, Noise variance = 2.2944816308180542e-10

Normal	Testing
	

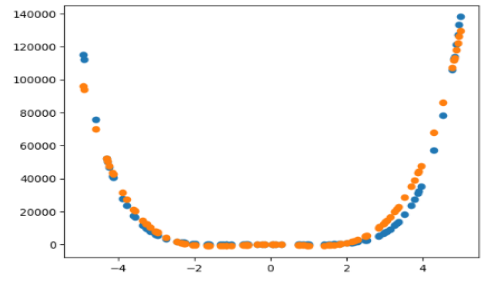
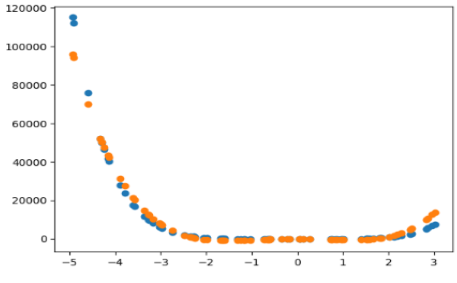
Degree = 3

alpha= 0.01, Noise variance = 1.4265209370579218e-09

Normal	Testing
	

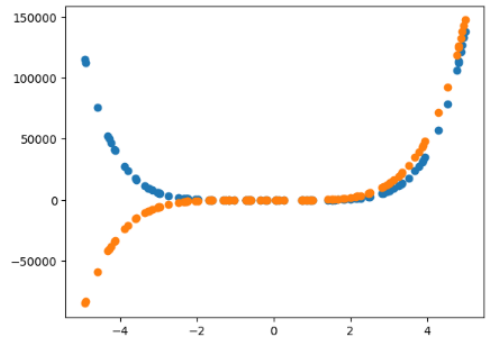
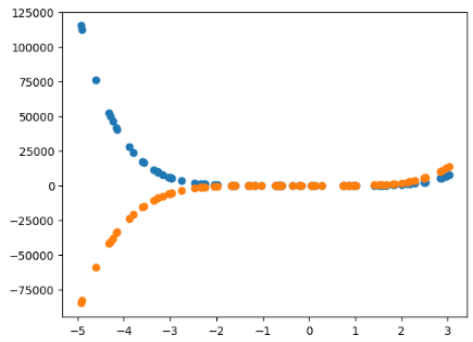
Degree = 4

alpha= 0.0001, Noise variance = 8.605422534985641e-09

Normal	Testing
	

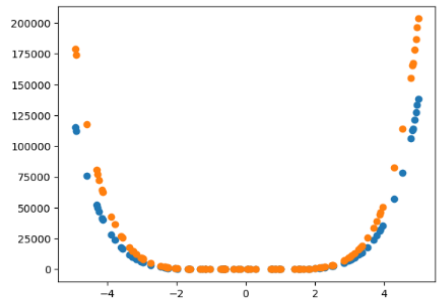
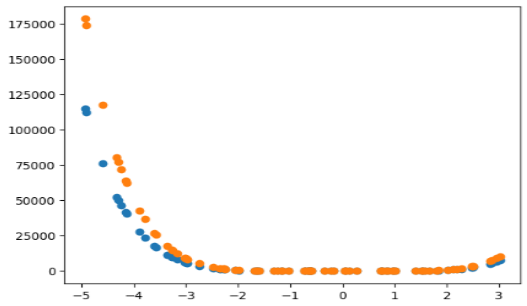
Degree = 5

alpha= 0.0001, Noise variance = 4.7038079185261 56e-08

Normal	Testing
	

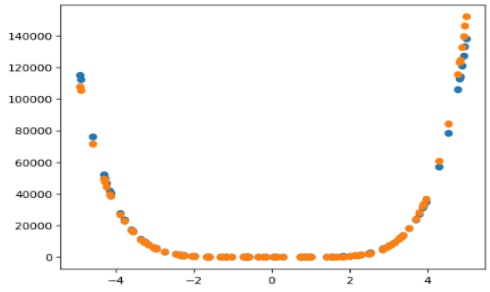
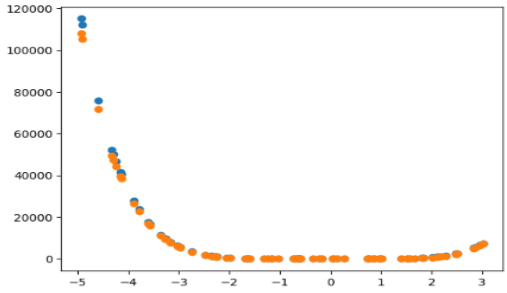
Degree = 6

alpha= 0.000001, Noise variance = 6.57392648583684848e-08

Normal	Testing
	

Degree = 7

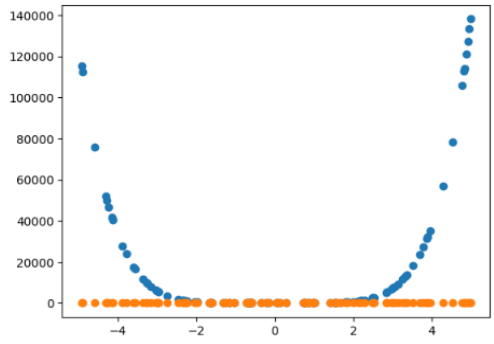
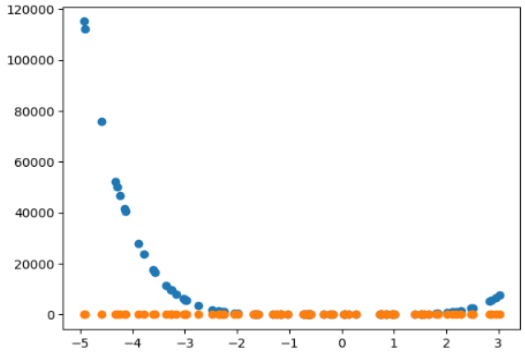
alpha= 0.00000001, Noise variance = 5.786154501947346e-07

Normal	Testing
	

Regularization(For 100 data points)

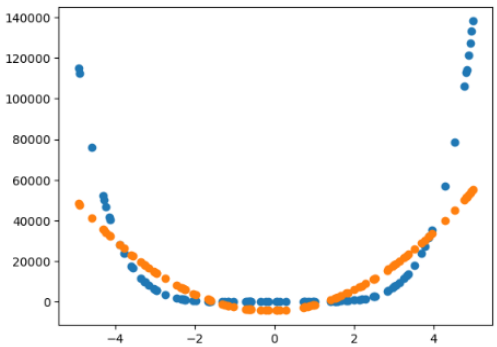
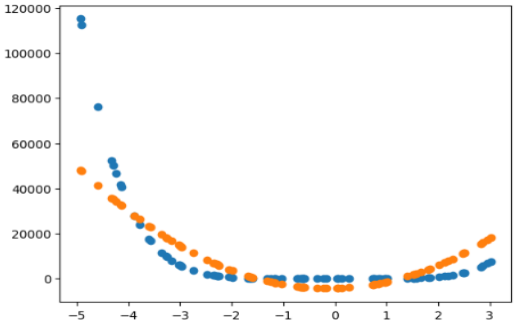
Degree = 1

alpha= 0.01, Noise variance = 2.8775788208742 37e-14

Normal	Testing
	

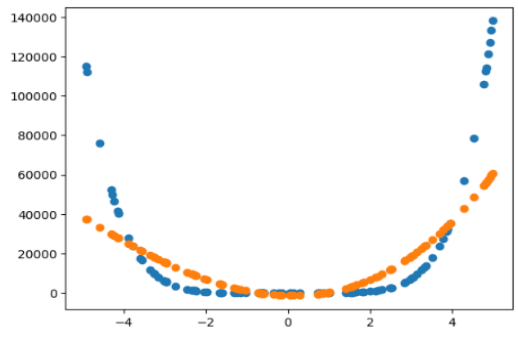
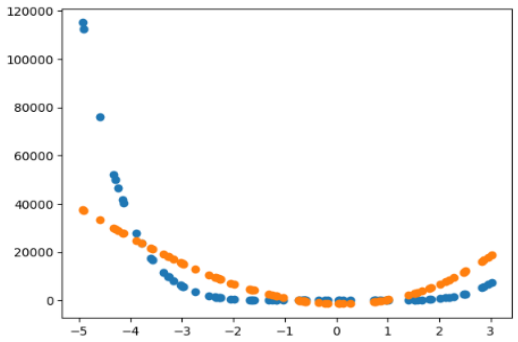
Degree = 2

alpha= 0.01, Noise variance = 2.2941463054667088e-10

Normal	Testing
	

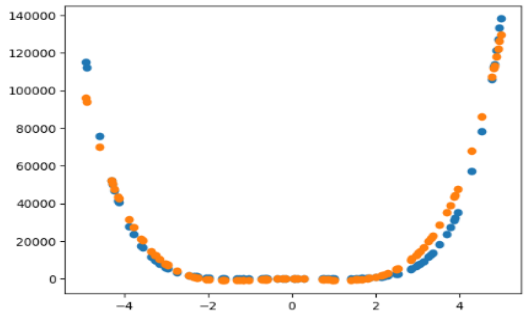
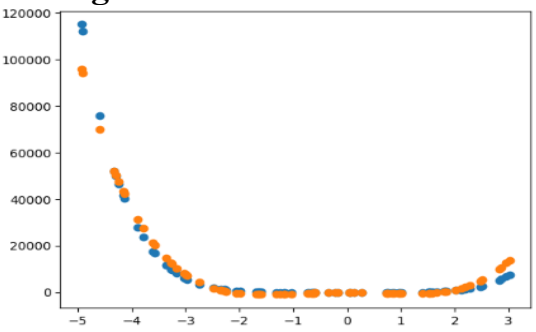
Degree = 3

alpha= 0.001, Noise variance = 1.4264524697878712e-09

Normal	Testing
	

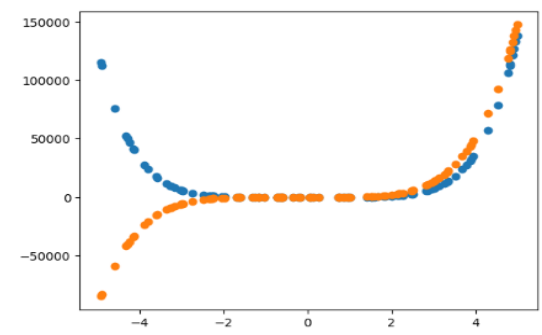
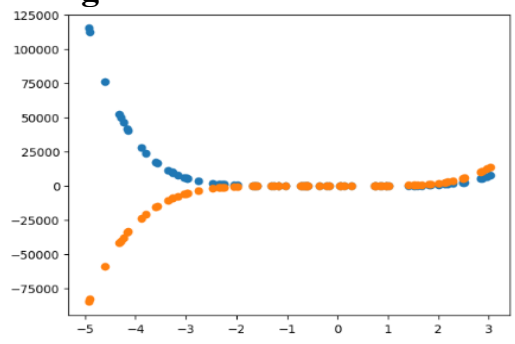
Degree = 4

alpha= 0.0001, Noise variance = 8.60517338753603e-09

Normal	Testing
	

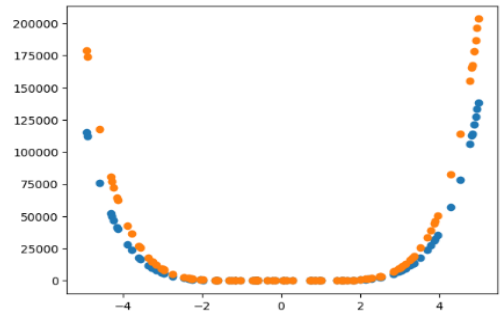
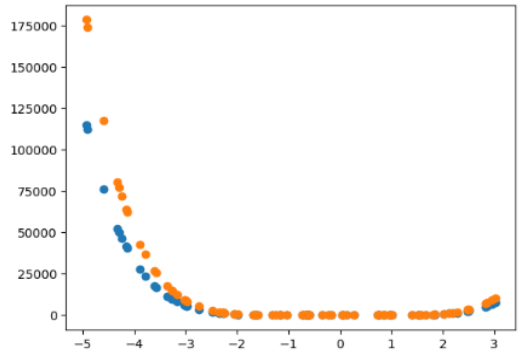
Degree = 5

alpha= 0.0001, Noise variance = 4.7038072122904895e-08

Normal	Testing
	

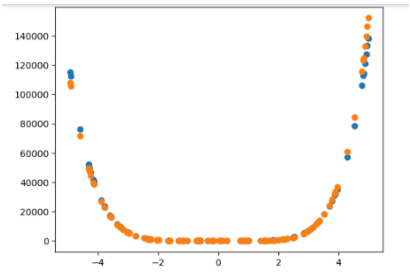
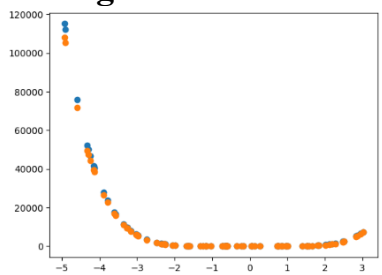
Degree = 6

alpha= 0.000001, Noise variance = 1.0814781025111575e-07

Normal	Testing
	

Degree = 7

alpha= 0.00000001, Noise variance = 5.786152827012 94e-07

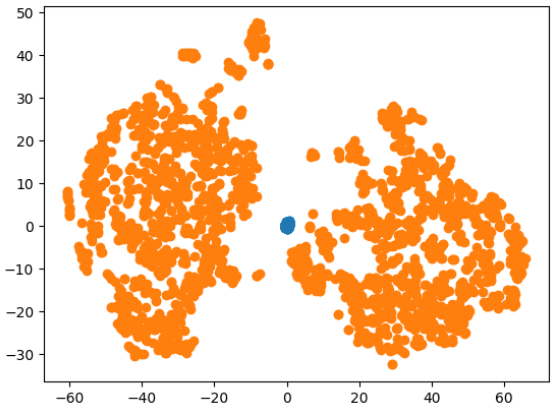
Normal	Testing
	

Results for all points:

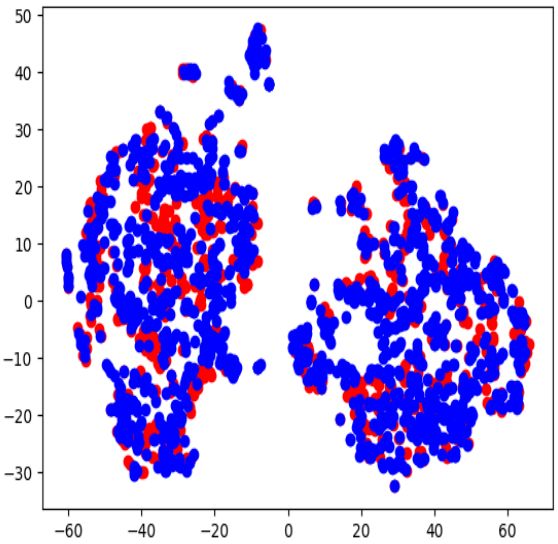
- Therefore Degree 1 polynomial is underfit
- Degree 6 model is best fit
- Degree 7 and above degree model is overfit

SVM Classifier

The data is distributed as by using PCA with no. of component as =2



We can see the data set is seems separable in 2D space but not clearly separable when projected into 2D space Class wise as shown below.



Our aim is to classify the dataset into two linearly separable classes. For which Kernel Function can be used which can transform our data set into higher dimensional space and from where it can be further classify the data set into two linearly separable classes.

Kernel Function can be defined as below: if we have two data points x1,x2 then Kernel can be defined as below:

- $k(\mathbf{x}_i, \mathbf{x}_j) \triangleq \boldsymbol{\phi}^T(\mathbf{x}_i)\boldsymbol{\phi}(\mathbf{x}_j) = \boldsymbol{\phi}^T(\mathbf{x}_j)\boldsymbol{\phi}(\mathbf{x}_i)$

Kernel Used here:

1. Linear Kernel
2. Gaussian Radial Basis Function(RBF)

1. Linear Kernel:

Using this Linear Kernel we will classify our model and find out the accuracy for different values of C.

```
C_2d_range = [1e-3,1e-2,1e-1,1,10,1e2]
```

The accuracy that we have got from different values of C are:

Accuracy array:

[[67.75]

[67.75]

[67.75]

[88.75]

[93.25]

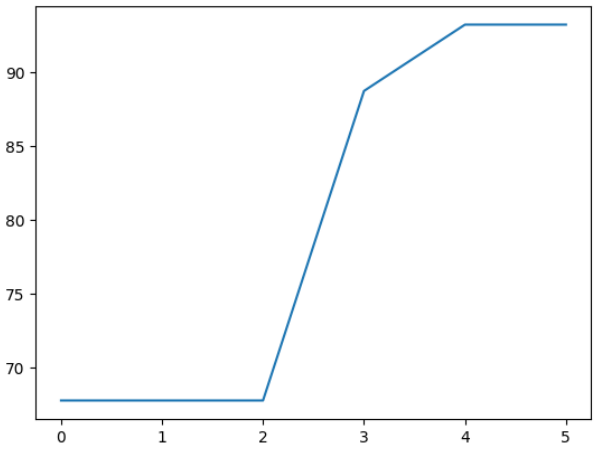
[93.25]]

The maximum accuracy we have obtained at the value of C =10 which is 93%

Therefore:

Maximum Accuracy: 0.9325

Optimal Value of C is : 10



Above figure shows the variation of accuracy vs the parameter C.

Corresponding Optimal Coefficients:

[[42.62255715 -4.240323 4.54597997 2.64347022 0.48345448 5.63709817

3.0556774 2.29229434]]

Intercept:

[-40.77879085]

2) **Gaussian Radial Basis Function (RBF):**

```
C_values_RBF = [1,10,50,100,150]

gamma_range_RBF = [1e-2,1e-1, 1, 1e1,1e+2]
```

Accuracy Matrix :

[[67.8 67.8 93.65 91.9 82.3]

[67.8 94.1 94.55 92.05 83.7]

[90.85 94.6 94.75 91.95 83.8]

[94. 94.65 94.4 91.35 83.95]

[94.2 94.7 94.35 91.05 84.]]

Maximum Accuracy: 94.75

Optimal Value of C is: 50

Optimal Value of Gamma: 1

Optimal Value of Sigma: 0.7071067811865476

Corresponding Optimal Intercept: [1.37997926]