Pattern Recognition
Final Report

Assignment - 4
Support Vector Machines

SUBMITTED BY

Yash Agrawal

B16120

B.Tech. (Computer Science)

2016 - 2020 Batch, IIT Mandi

Table of Contents

Introduction & Background

In general classifiers are used to classify a set of data-points to their corresponding class. There are many types of classifiers used in pattern recognition field in which pattern may be in the form of speech, image data, voice data and more than one dimensional data. These datas needs to be classified to their particular class from which the data came from. In context to the course CS669 pattern recognition mainly bayes classifier is used to distinguish data of a class from other classes while other classifiers are also used as perceptron based classifier (for linearly separable data) and SVM-Based classifier.
The classification and building of model depends on the nature of data and how the data is extracted from the its original representation which is explained further.

The classification of data is done using following five basic steps :

1) Data Input
2) Feature Extraction
3) Feature Representation
4) Pattern Analysis Models
5) Output

Data Input → Extracting Features→ Representation ($\bar{x} \in \mathbb{R}^d$) → Build Model → Output

Points to keep in mind while building a pattern classification model :

1) Limited set of data per class
2) Intra class variability
3) Inter class similarity
4) Noise / Uncertainty in Data

Datasets Given

Datasets :

Dataset 1 : 2-Dimensional Artificial Data :

(A) Linearly Separable Dataset used in Assignment 1

(B) Non-Linearly Separable Dataset used in Assignment 1

Dataset 2 : 3 Class Scene Image Dataset :

32-Dimensional BOVW Representation from Assignment 2

Parameters to be observed

1. Classification accuracy
2. Precision for every class and mean precision
3. Recall for every class and mean recall
4. F-Measure for every class and mean F-Measure.
5. Confusion Matrix based on the performance for test data.
6. Observation on the nature of decision surface obtained for Dataset-1 for SVM.
7. Plot of 1-D and 2-D reduced dimensional representations using FDA.
8. Comparison with all the classifiers fo each datasets.

Objective :

Classifiers to be built :

(a) Build Bayes Classifier using Gaussian Mixture Model (GMM) with 1,2,4 and 8 mixtures on the reduced dimensional representations of dataset-2 obtained using PCA (Principal Component Analysis).

(b) Apply Fisher Discriminant Analysis (FDA) on Dataset-1 and Dataset-2. Use Bayes Classifier using both unimodal Gaussian and Gaussian Mixture Model with mixtures as 1,2,4 and 8.

(c) Perceptron Based Classifier on Dataset-1(a).

(d) SVM-Based Classifier using :

(i) Linear Kernel
(ii) Polynomial Kernel
(iii) Gaussian/RBF Kernel

Which is to be applied on dataset-1 and dataset-2 as described in the above section.

In the above classifiers described, the classifier in part (a) that is bayes classifier using GMM on reduced dimensional representation of dataset-2 obtained using PCA is already done in the previous assignment (Assignment-3).

Procedure

FDA (Fisher Discriminant Analysis) on dataset-1 and dataset-2 :

1) Principal component analysis doesn't make sure to give always better discrimination because it only checks for the higher eigenvalues.
2) FDA is used to find a projection such that the separability in the projected data is maximum.
3) $a_n = \overline{w}^T \overline{x}_n$ will give a single value which represent the one dimensional representation of the data where $\overline{w}$ is the direction of projection (d-dimension) and $\overline{x}_n$ is the d-dimensional data.
4) For maximum separability, the mean of two classes should be as separate as possible i.e. distance between mean of two classes should be as large as possible.
5) Scatter of two classes should be as minimum as possible.

$$J = \frac{(m_+ - m_-)^2}{s_+^2 + s_-^2}$$

6) Maximize the value of J in the above equation which after differentiation gives :

$$\overline{w} = \lambda \zeta_w^{-1} (\overline{\mu_+} - \overline{\mu_-})$$

   Which gives the direction of projection.

7) The Procedure of FDA from step 1 to step 6 is repeated for all the pair of classes for a particular datapoint.

8) Using a decision logic (voting scene) a class label is assigned to that particular datapoint.

Perceptron Based Classifier on dataset-1(a) :

1) Perceptron is a discriminative learning technique which classify the data of two classes come from linear separable dataset.
2) Consider the data coming from two classes one as positive class and other as negative class.
3) D_m be the set of all misclassified examples.
4) Classification :
   a) If $\overline{a}^T z_n > 0$: Data Point belongs to positive class.
   b) If $\overline{a}^T z_n < 0$: Data Point belongs to negative class.

5) Checking misclassification :
   a) If $y_n \times (\overline{a}^T z_n) > 0$: Data Point is correctly classified.
   b) If $y_n \times (\overline{a}^T z_n) < 0$: Data Point is misclassified.

6) Initialize a(0) at k = 0th iteration.
7) For any (k+1)th iteration determine $D_m(K)$ for a(K).

$$\overline{a}(K+1) = \overline{a}(K) + \Delta\overline{a}(K)$$

$$\overline{a}(K+1) = \overline{a}(K) + \eta\,\Sigma_{x_n \leftarrow D_m(K)}\, y_n \overline{z_n}$$

8) Repeat step 2 till $D_m(K)$ is empty.
9) At starting η should be large and then after each iteration decrease the value of η by a constant factor as when η is large then for convergence there will be lot of oscillations.

Support Vector Machines on dataset-1 and dataset-2 :

1) It does not get trapped in local minima.
2) Gives the optimal solution given the dataset.
3) Only Method which gives the optimal value.
4) The objective function is quadratic surface which is independent of dimension of data.
5) Free from curse of dimensionality.
6) As this method does not depend on the dimension of data so the transformation of data to a higher dimensional space takes place.
7) The transformation to higher dimensional space for conversion of non-linear relations into linear relations which is known as **Cover's theorem.**

$$\bar{x} \rightarrow \phi(\bar{x})$$

$$\bar{x} \in R^d \qquad \phi(\bar{x}) \in R^D \qquad \text{where } D >> d.$$

8) Pairwise inner product are computed efficiently directly from original representation of data using kernel function.
9) Kernel Gram Matrix is formed which is symmetric and positive semi definite i.e. the eigen should be non-negative.
10) Different types of kernels are used for kernel gram matrix mainly :
    a) Linear Kernel :

$$K(\bar{x}_m, \bar{x}_n) = \bar{x}_m^T \bar{x}_n$$

    b) Polynomial Kernel :

$$K(\bar{x}_m, \bar{x}_n) = (a\,\bar{x}_m^T \bar{x}_n + b)^p$$

    c) Gaussian Kernel (Radial Base function) :

$$K(\bar{x}_m, \bar{x}_n) = e^{\frac{-\|\bar{x}_m - \bar{x}_n\|^2}{\sigma}}$$

Where a, b, σ, p are the parameters of SVM kernel should be chosen according to data.

Observations

Fisher Discriminant Analysis (FDA) :

(1) Dataset -1(a) : Linearly Separable Data :

(a) K = 1: Accuracy : 100.00 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	125	0	0
Class 2	0	125	0
Class 3	0	0	125

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	1.0	1.0	1.0	1.0
Precision	1.0	1.0	1.0	1.0
F-Measure	1.0	1.0	1.0	1.0

(b) K = 2: Accuracy : 100.00 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	125	0	0
Class 2	0	125	0
Class 3	0	0	125

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	1.0	1.0	1.0	1.0
Precision	1.0	1.0	1.0	1.0
F-Measure	1.0	1.0	1.0	1.0

(c) K = 4: Accuracy : 100.00 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	125	0	0
Class 2	0	125	0
Class 3	0	0	125

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	1.0	1.0	1.0	1.0
Precision	1.0	1.0	1.0	1.0
F-Measure	1.0	1.0	1.0	1.0

(d) K = 8: Accuracy : 99.733333 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	125	0	0
Class 2	0	125	0
Class 3	1	0	124

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	1.0	1.0	0.992	0.9973
Precision	0.992	1.0	1.0	0.9973
F-Measure	0.996	1.0	0.995	0.9973

(2) Dataset -1(b) : Non-Linearly Separable Data :

(a) K = 1: Accuracy : 79.00%

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	119	5	1
Class 2	10	115	0
Class 3	53	36	161

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.952	0.92	0.644	0.8386
Precision	0.6538	0.7371	0.9938	0.7949
F-Measure	0.7752	0.8185	0.7815	0.7917

(b) K = 2: Accuracy : 80.02%

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	121	4	0
Class 2	6	119	0
Class 3	54	35	161

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.968	0.92	0.644	0.8546
Precision	0.6685	0.7531	1.0	0.8072
F-Measure	0.7908	0.8409	0.7834	0.8050

(c) K = 4: Accuracy : 80.04%

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	118	7	0
Class 2	3	122	0
Class 3	52	36	162

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.944	0.976	0.648	0.856
Precision	0.682	0.739	1.0	0.807
F-Measure	0.791	0.8415	0.786	0.806

(d) K = 8: Accuracy : 79.08%

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	117	7	1
Class 2	3	122	0
Class 3	53	37	160

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.936	0.976	0.644	0.850
Precision	0.676	0.734	0.993	0.801
F-Measure	0.7852	0.8385	0.7785	0.800

(3) Dataset -2 : Class scene Image Dataset :

(a) K = 1: Accuracy : 55.33%

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	23	9	18
Class 2	7	39	4
Class 3	8	21	21

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.46	0.78	0.42	0.533
Precision	0.605	0.565	0.488	0.552
F-Measure	0.522	0.655	0.451	0.543

(b) K = 2: Accuracy : 58.66%

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	28	10	12
Class 2	8	39	3
Class 3	9	20	21

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.56	0.78	0.42	0.586
Precision	0.625	0.565	0.588	0.592
F-Measure	0.582	0.655	0.481	0.573

(c) K = 4: Accuracy : 50.00%

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	18	10	22
Class 2	10	35	5
Class 3	7	21	22

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.36	0.78	0.44	0.503
Precision	0.515	0.535	0.448	0.492
F-Measure	0.422	0.605	0.441	0.49

(d) K = 8: Accuracy : 33.33%

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	0	0	50
Class 2	0	0	50
Class 3	0	0	50

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0	0	1.0	0.33
Precision	0	0	0.33	0.11
F-Measure	0	0	0.5	0.16

Inferences :

1. For linearly separable dataset, the accuracy of bayes classifier using gaussian mixture model comes out to be nearly 100% which is expected.

2. The linearly separable data is already separated completely and when this data is projected to one dimension line then data coming from one class will be located totally different from other class.

3. Accuracy is 100% until number of clusters k = 4, after that for k=8 the accuracy is slightly decreased to 99.73% due to increase in number of clusters.

4. The increase in number of clusters leads to increase in chances of getting only one point in a cluster which can make covariance matrix zero.

5. For Non-linearly Separable Dataset, the data is already separated except some of the datapoints which are mixed together leading to misclassification of some test datapoint and decrease the accuracy of the model.

6. For class scene image dataset which is a real world dataset, datapoint are mixed up the thoroughly that even with FDA, discrimination of whole data of a class is not possible.

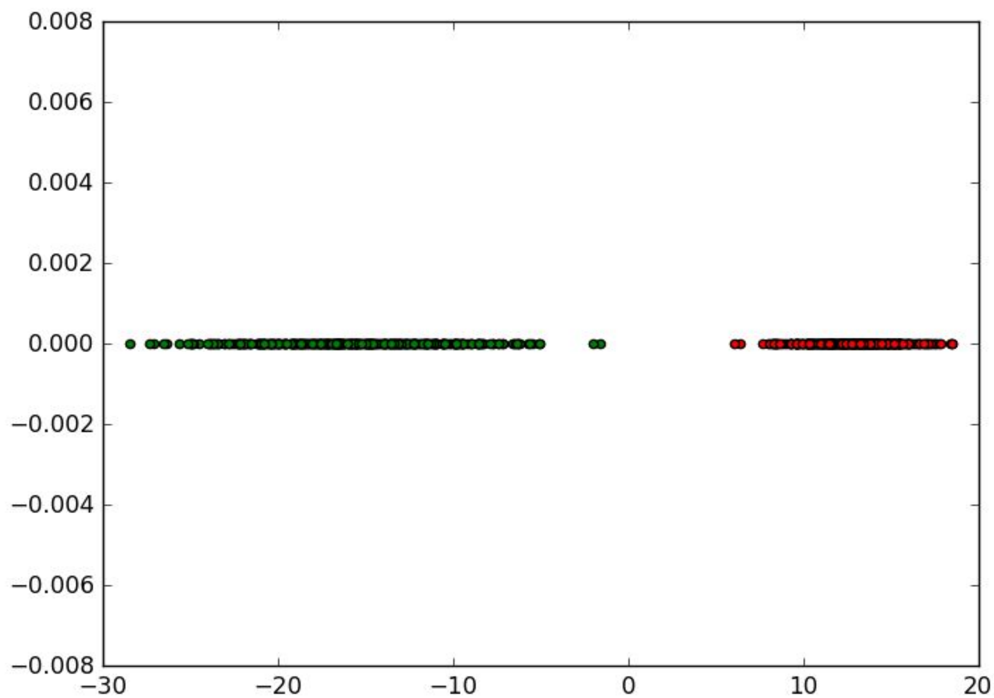
7. Due to which classififcation of test datpoint is poor for class image scene dataset.

Plot of 1-Dimensional reduced reprenstation using FDA :

(1) Linearly Separable Dataset :

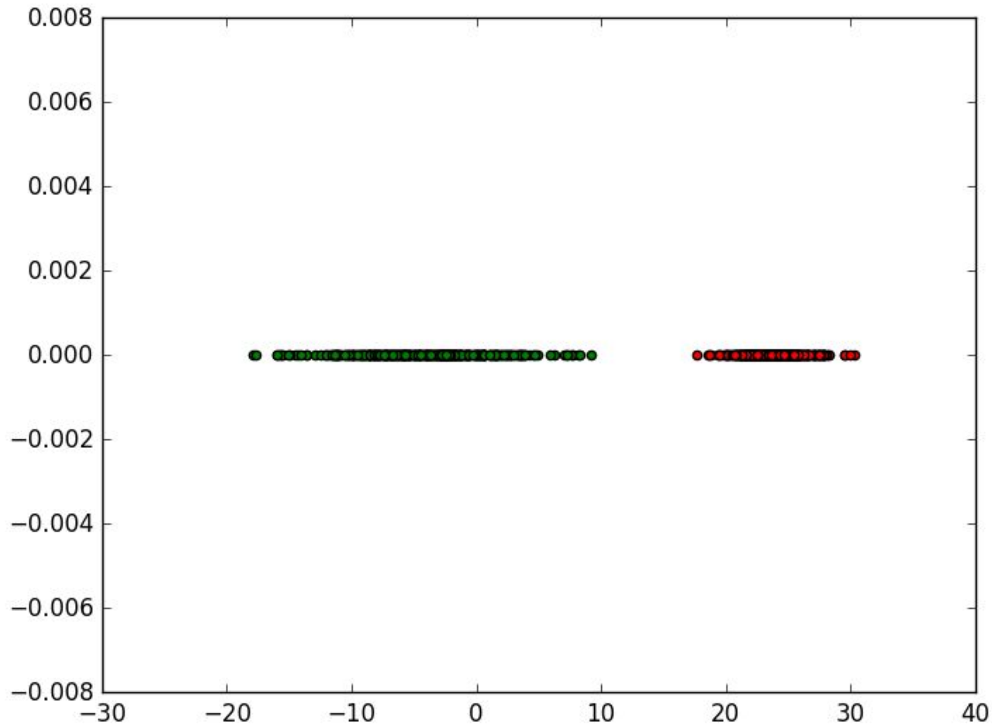
In all the plots x-axis represent the x-coordinate of datapoint and y-axis represent the y-coordinate of datapoint.

(a) Plot between class1 and class2 :



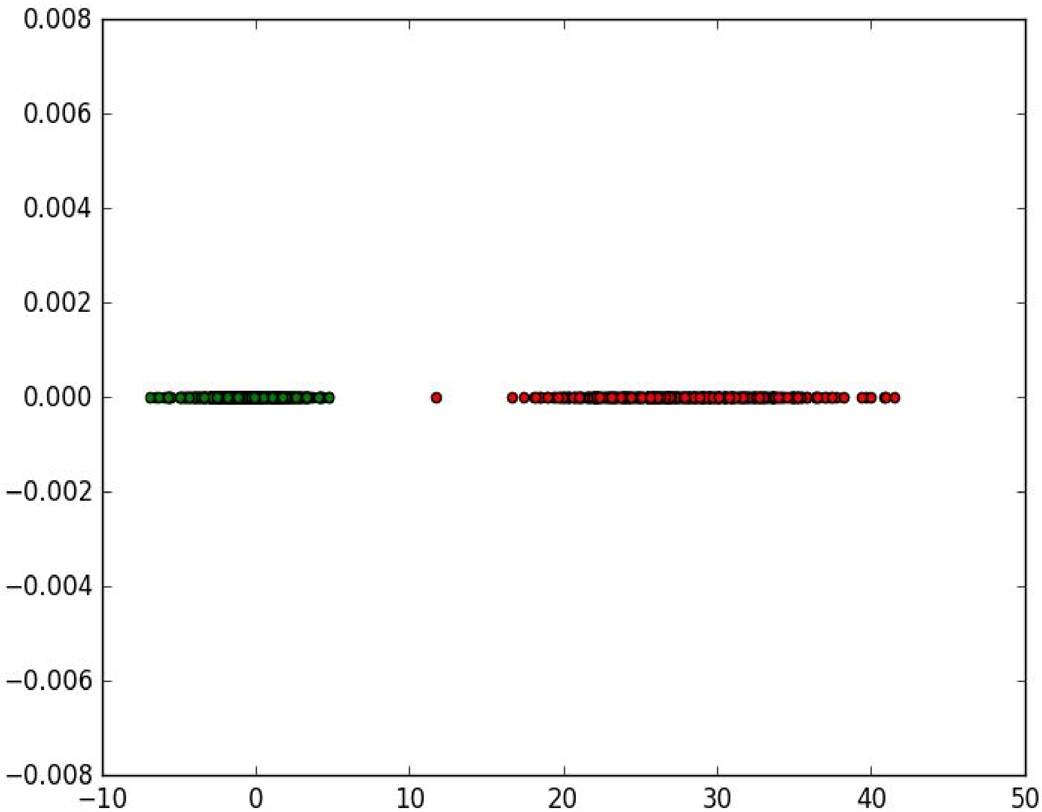
Plot for Fischer Discriminant Analysis for dataset-1a between class1 and class2

(b) Plot between class2 and class3 :



Plot for Fischer Discriminant Analysis for dataset-1a between class3 and class2

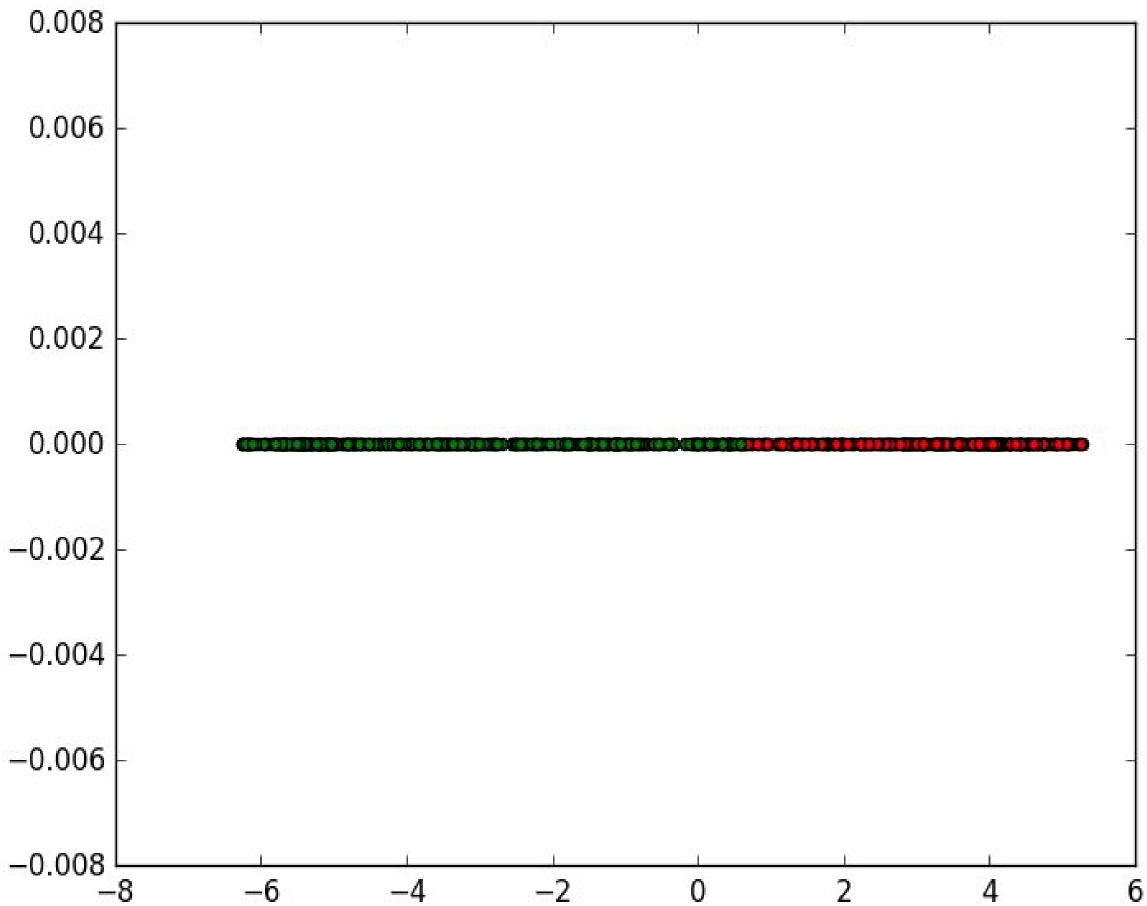
(c) **Plot between class1 and class3 :**



Plot for Fischer Discriminant Analysis for dataset-1a between class1 and class3

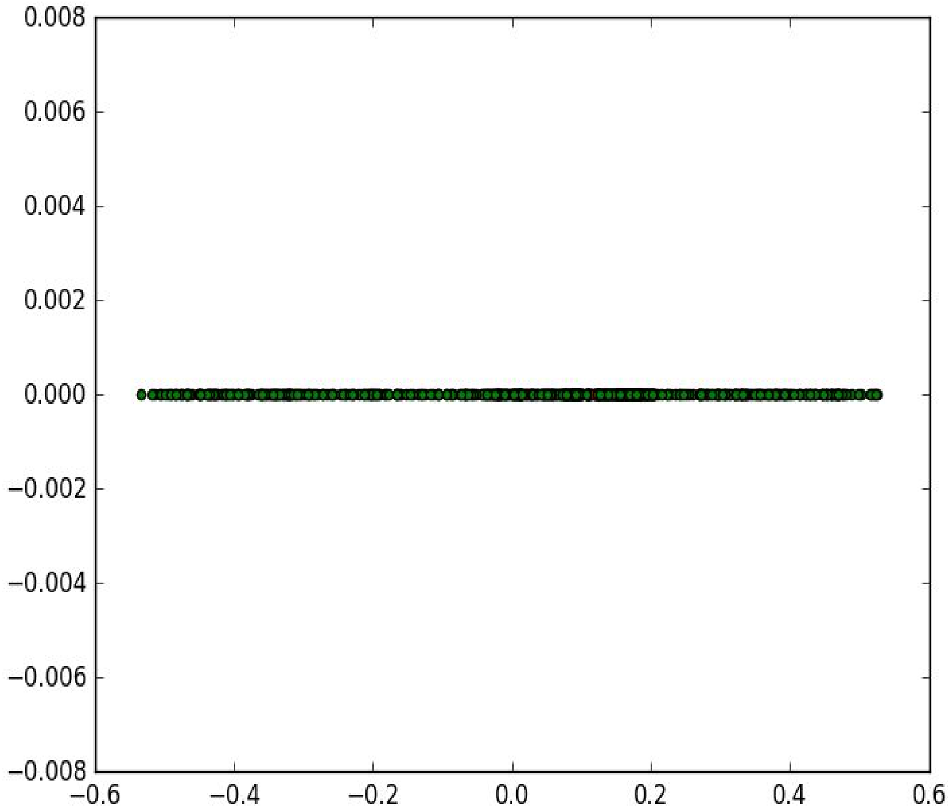
(2) Non-Linearly Separable Dataset :

(a) Plot between class1 and class2 :



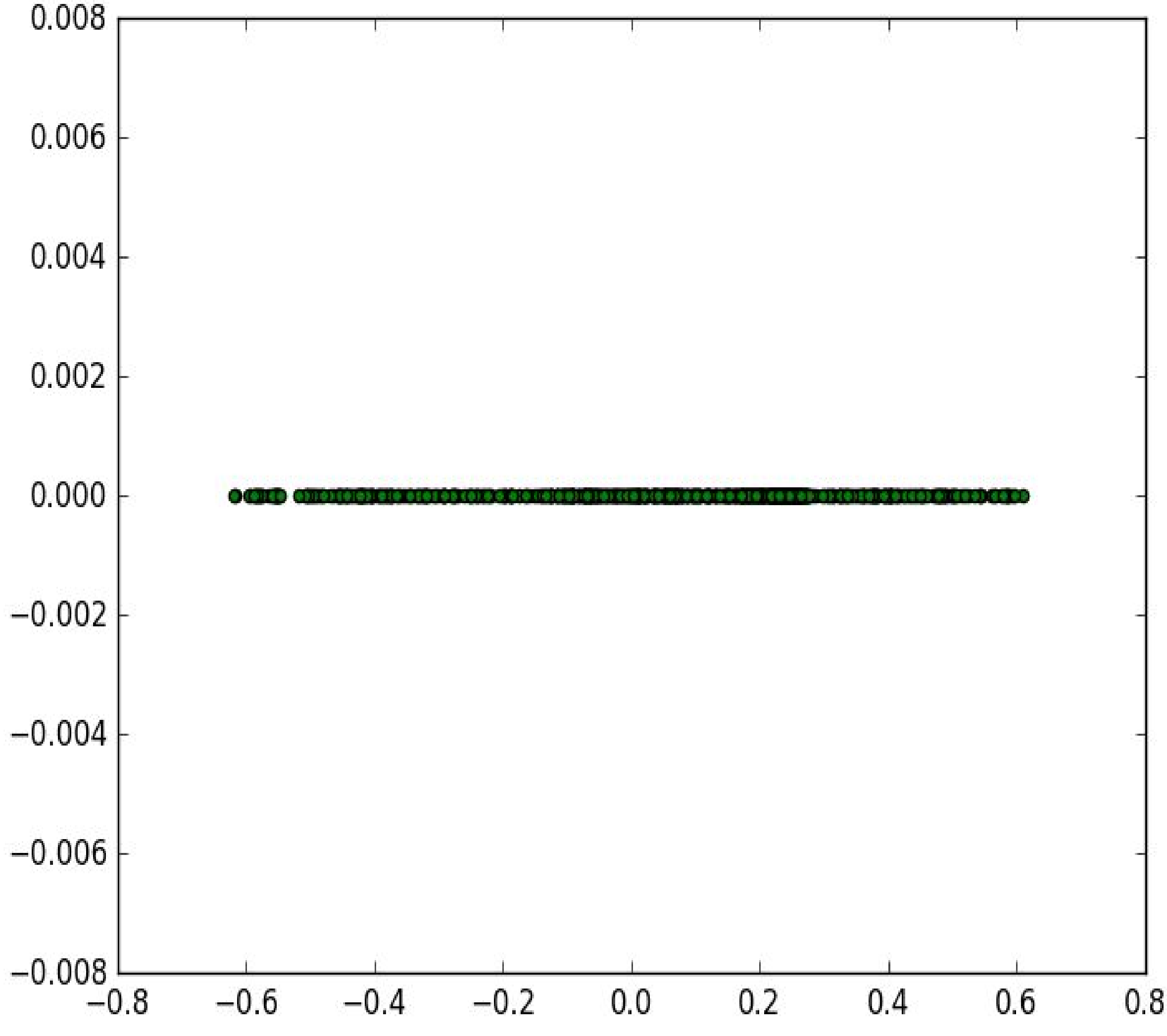
Plot for Fischer Discriminant Analysis for dataset-1b
between class1 and class2

(b) **Plot between class2 and class3 :**



Plot for Fischer Discriminant Analysis for dataset-1b
between class3 and class2

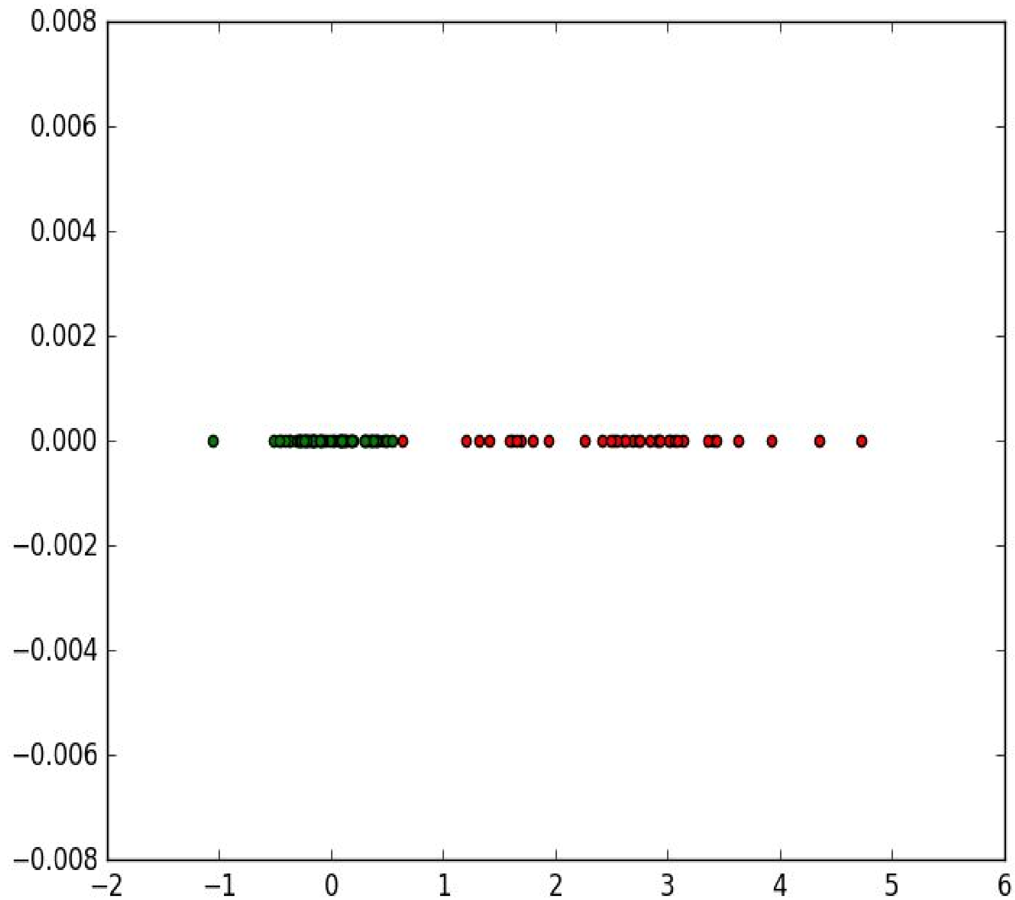
(c) **Plot between class1 and class3 :**



Plot for Fischer Discriminant Analysis for
dataset-1b between class1 and class3

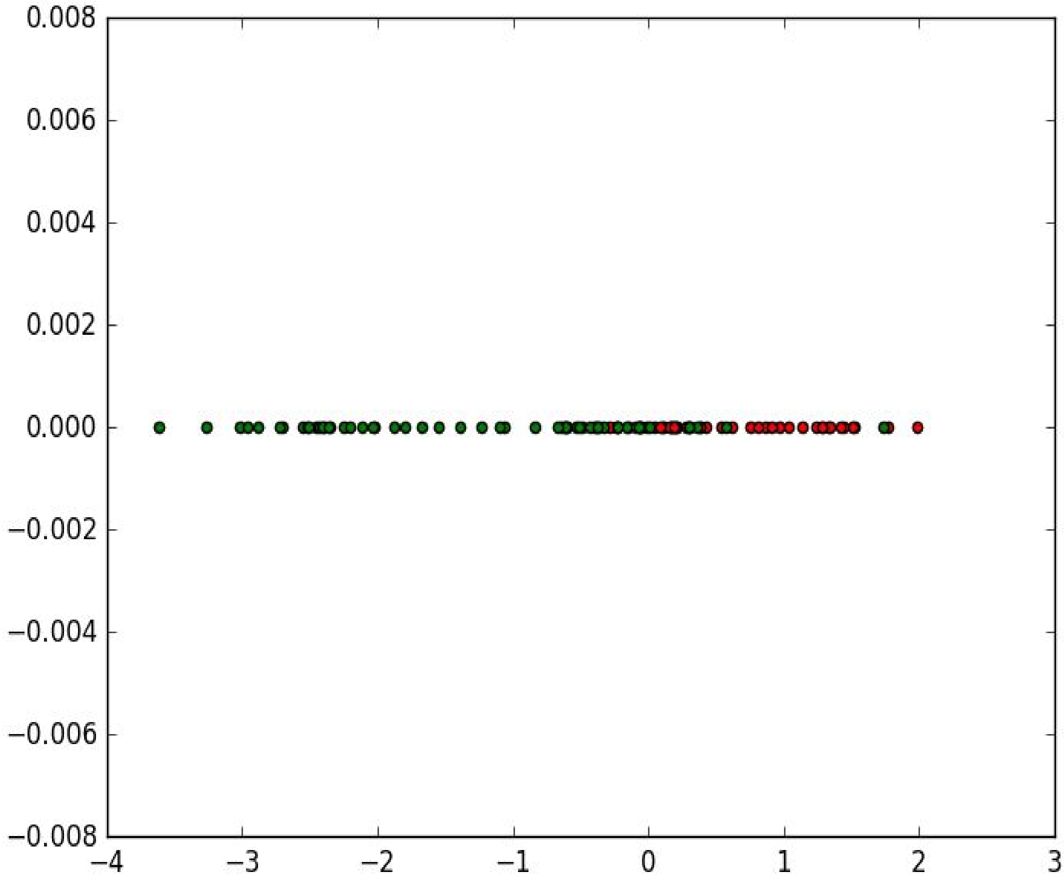
(3) Class Scene Image Dataset (BOVW representation) :

(a) **Plot between class1 and class2 :**



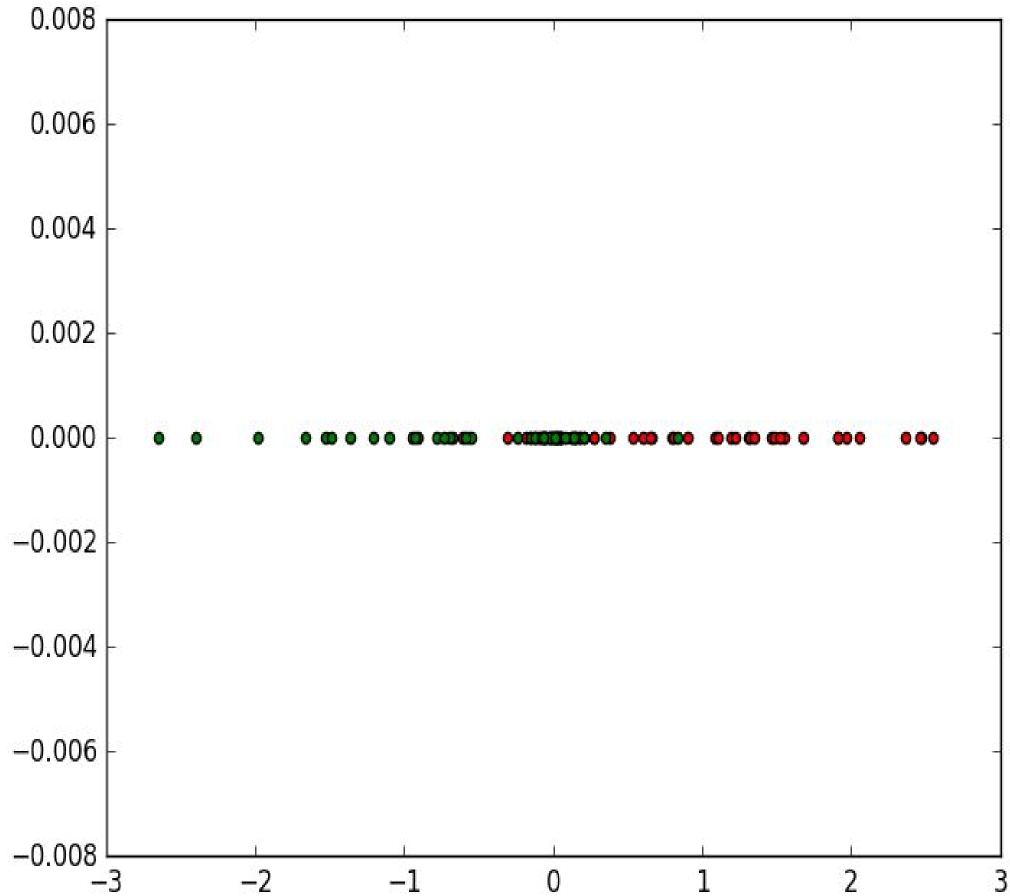
Plot for Fischer Discriminant Analysis for dataset-2 between class1 and class2

(b) **Plot between class2 and class3 :**



Plot for Fischer Discriminant Analysis for dataset-2
between class3 and class2

(c) **Plot between class1 and class3 :**



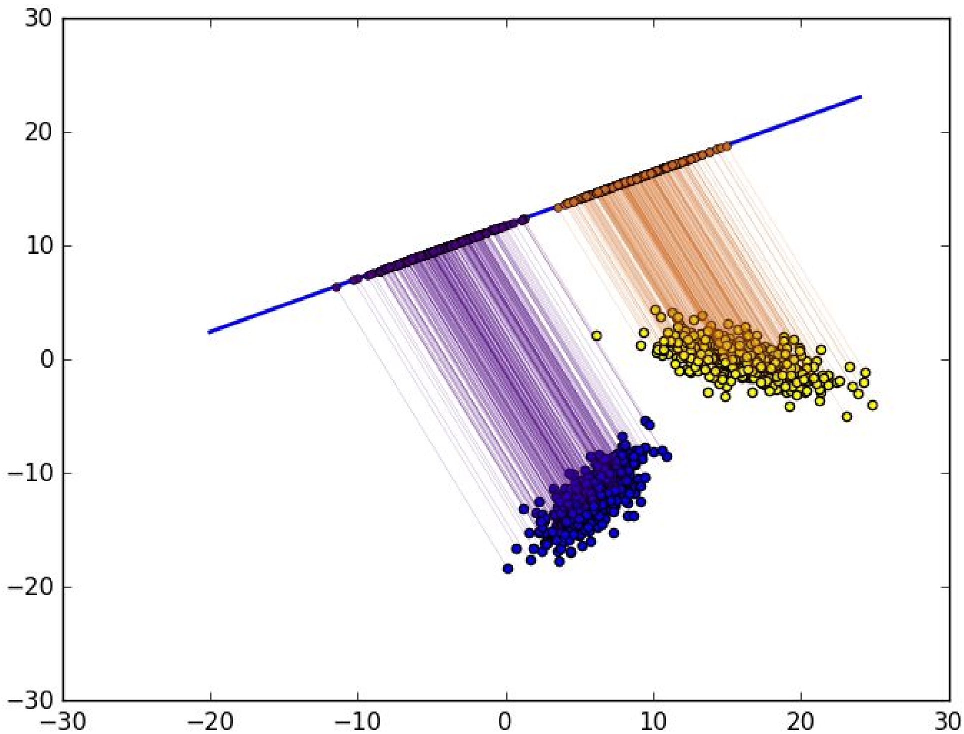
Plot for Fischer Discriminant Analysis for dataset-2
between class1 and class3

Plot of 2-Dimensional reduced reprenstation using FDA :

(1) Linearly Separable Dataset :

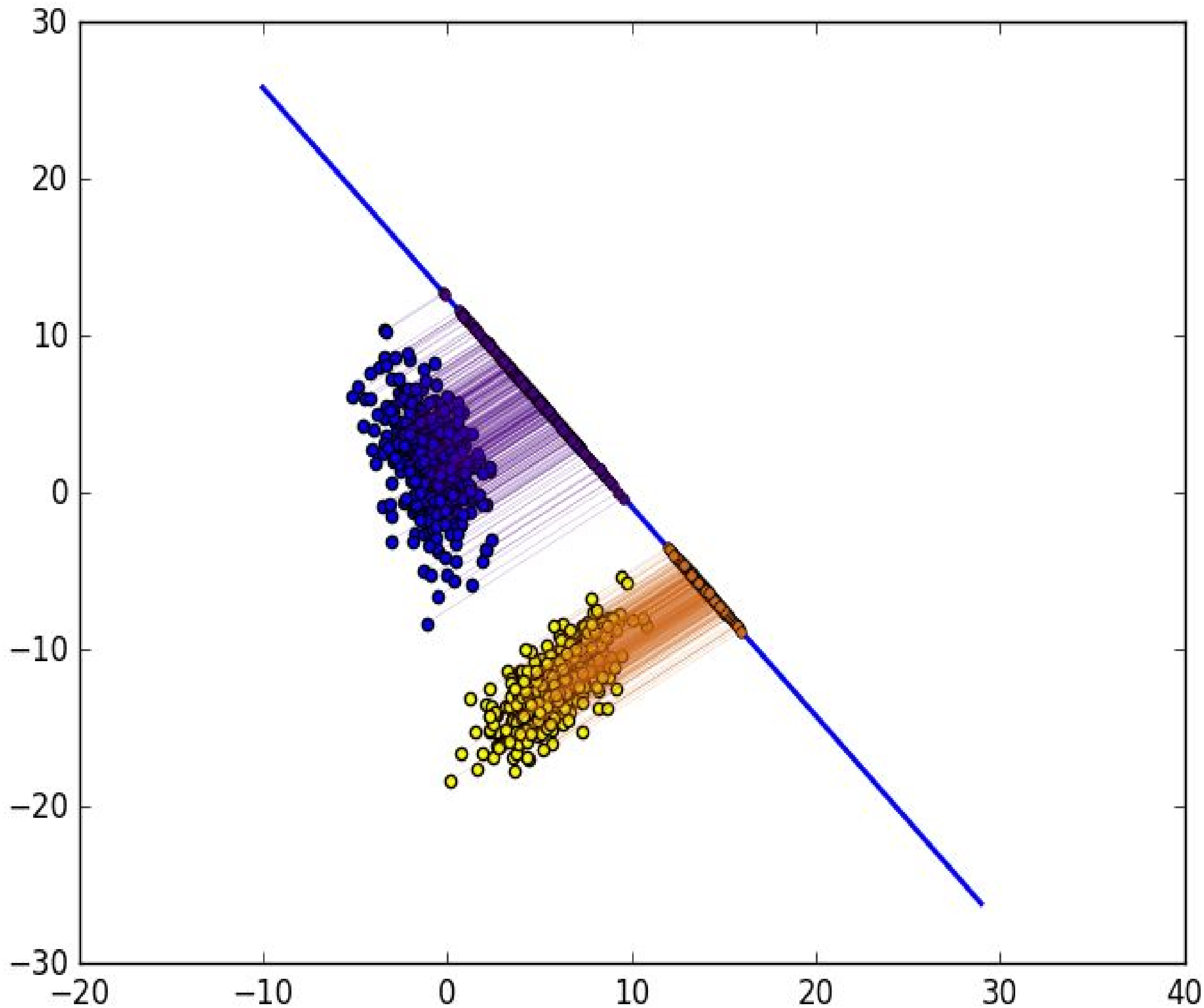
In all the plots x-axis represent the x-coordinate of datapoint and y-axis represent the y-coordinate of datapoint.

(A) Plot between class1 and class2 :



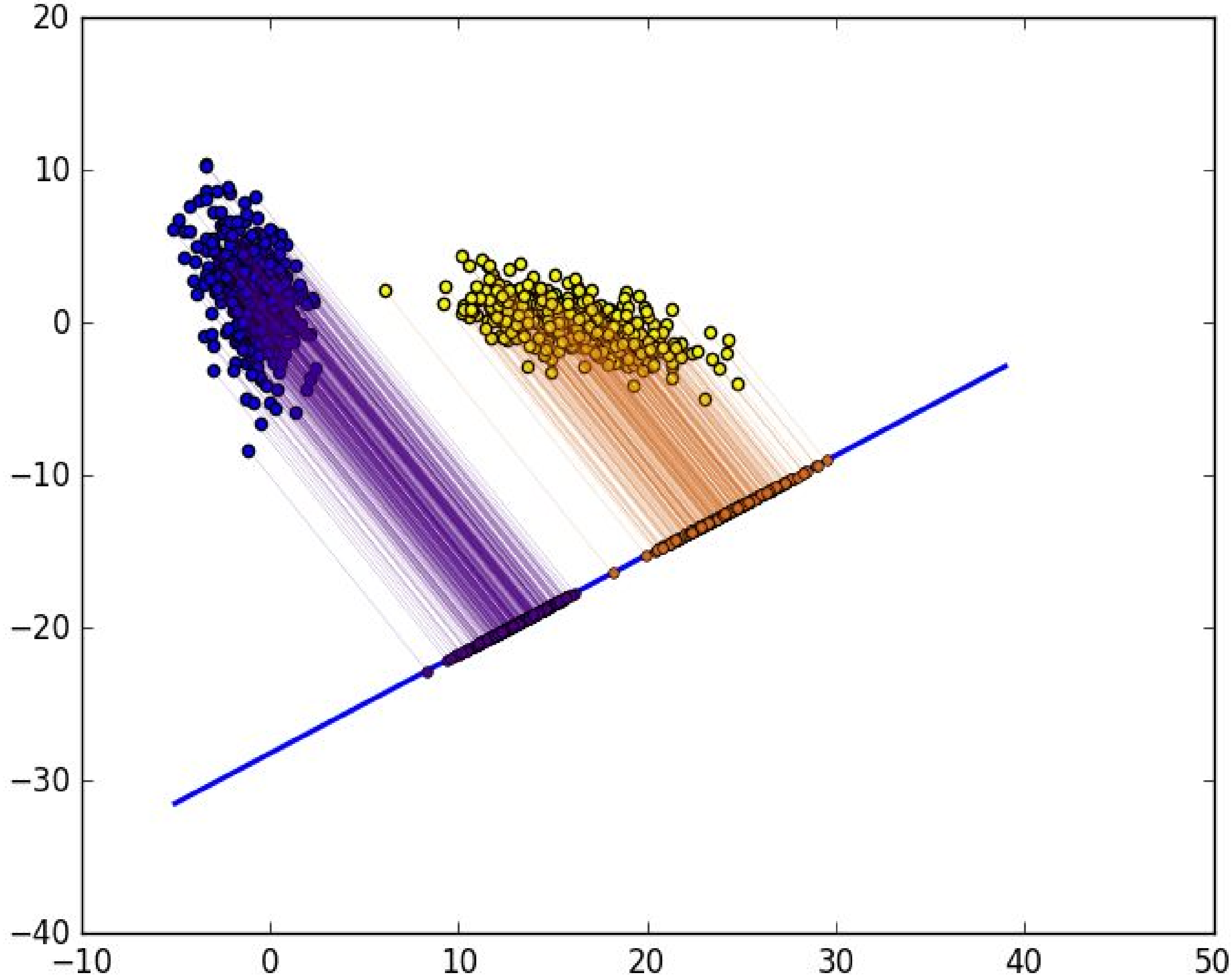
Plot for Fischer Discriminant Analysis for dataset-1a between class1 and class2

(B) Plot between class2 and class3 :



Plot for Fischer Discriminant Analysis for
dataset-1a between class3 and class2

(C) Plot between class1 and class3 :

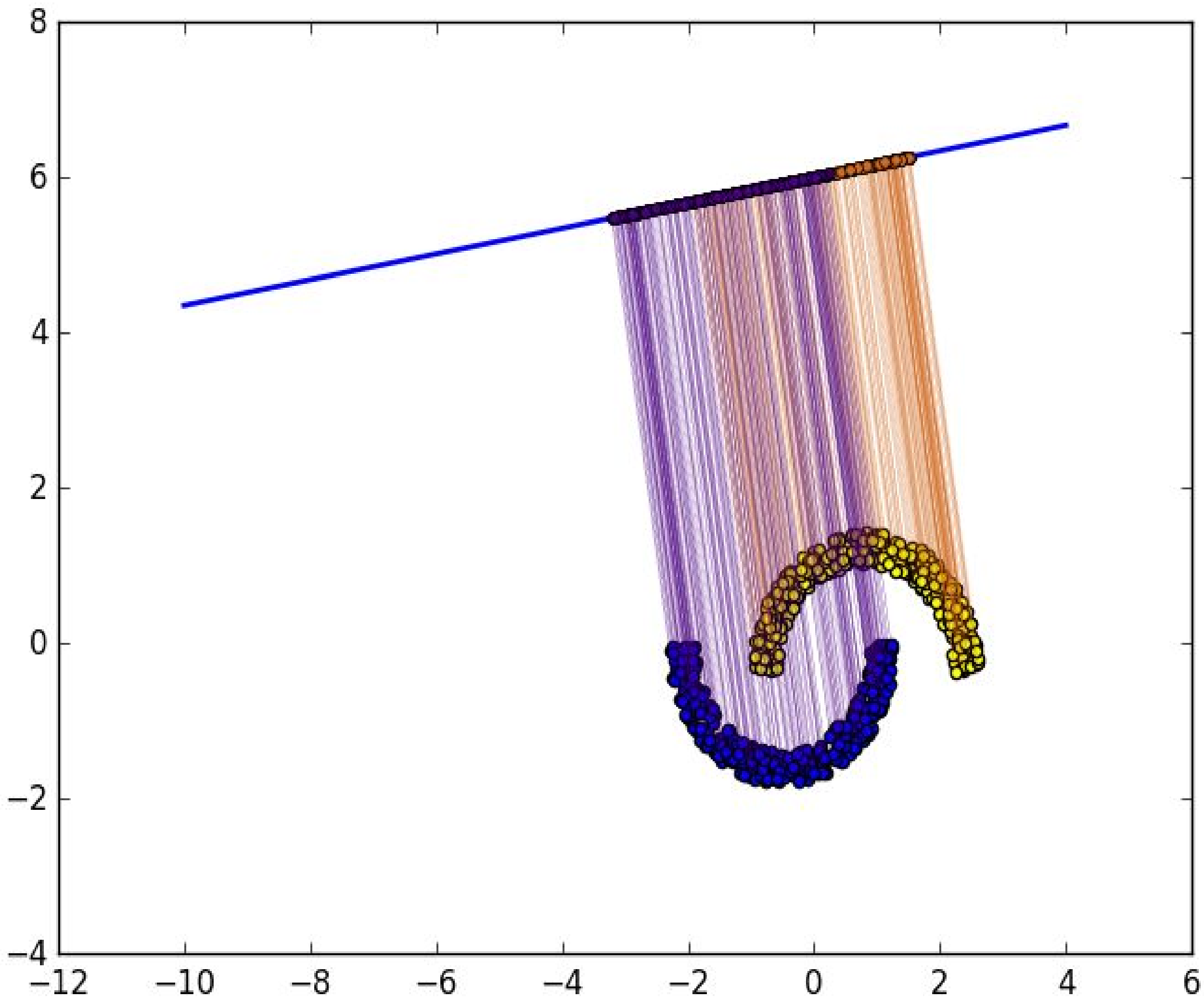


Plot for Fischer Discriminant Analysis for
dataset-1a between class1 and class3

(2) Non-Linearly Separable Dataset :

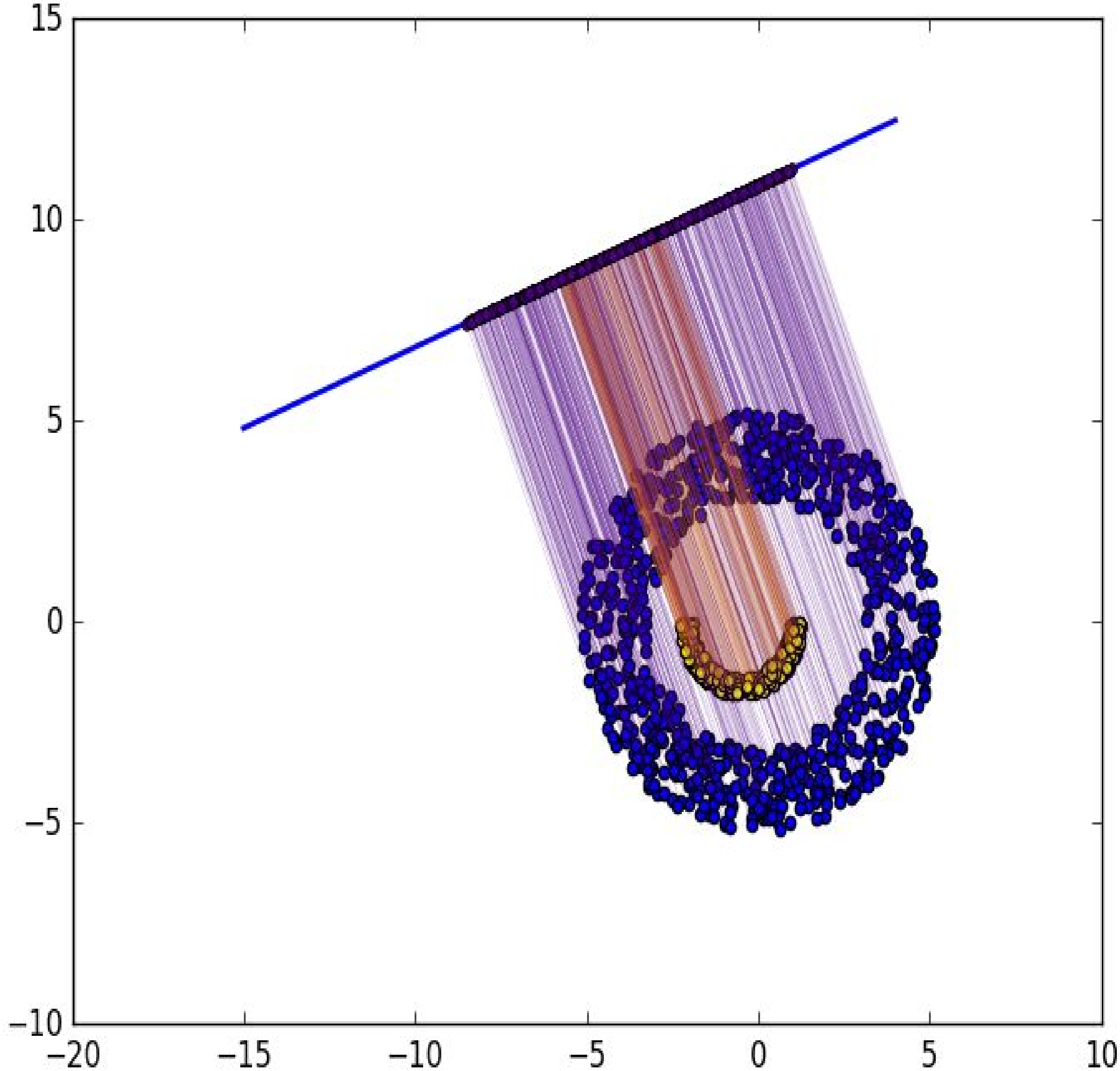
In all the plots x-axis represent the x-coordinate of datapoint and y-axis represent the y-coordinate of datapoint.

(A) Plot between class1 and class2 :



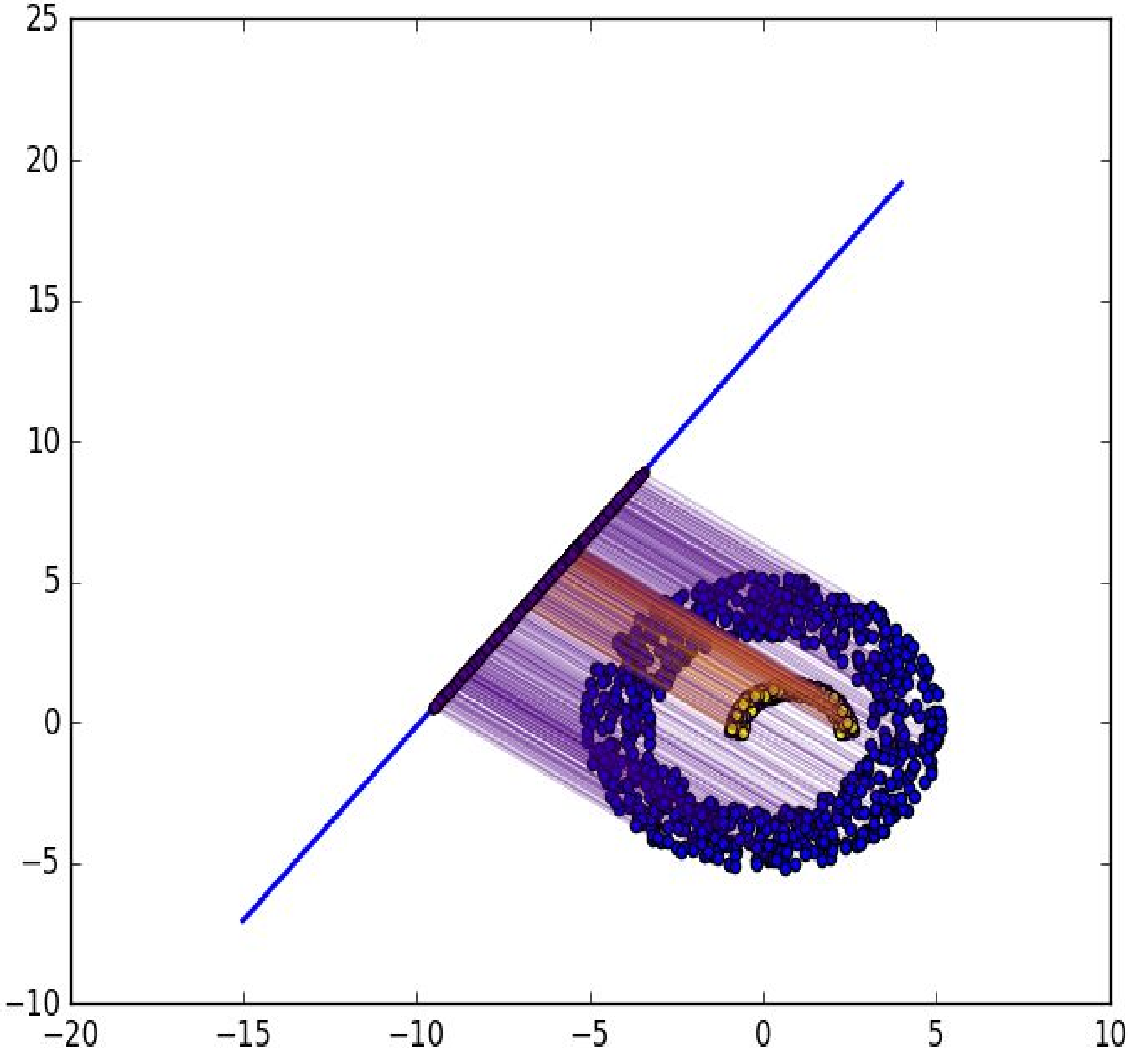
Plot for Fischer Discriminant Analysis for dataset-1b between class1 and class2

(B) Plot between class2 and class3 :



Plot for Fischer Discriminant Analysis for dataset-1b between class3 and class2

(C) Plot between class1 and class3 :



Plot for Fischer Discriminant Analysis for dataset-1b between class1 and class3

Perceptron Based Classifier on Dataset-1(a)

Results of studies for perceptron based classifier

Accuracy : 99.7333 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	125	0	0
Class 2	0	125	0
Class 3	0	1	124

Analysis

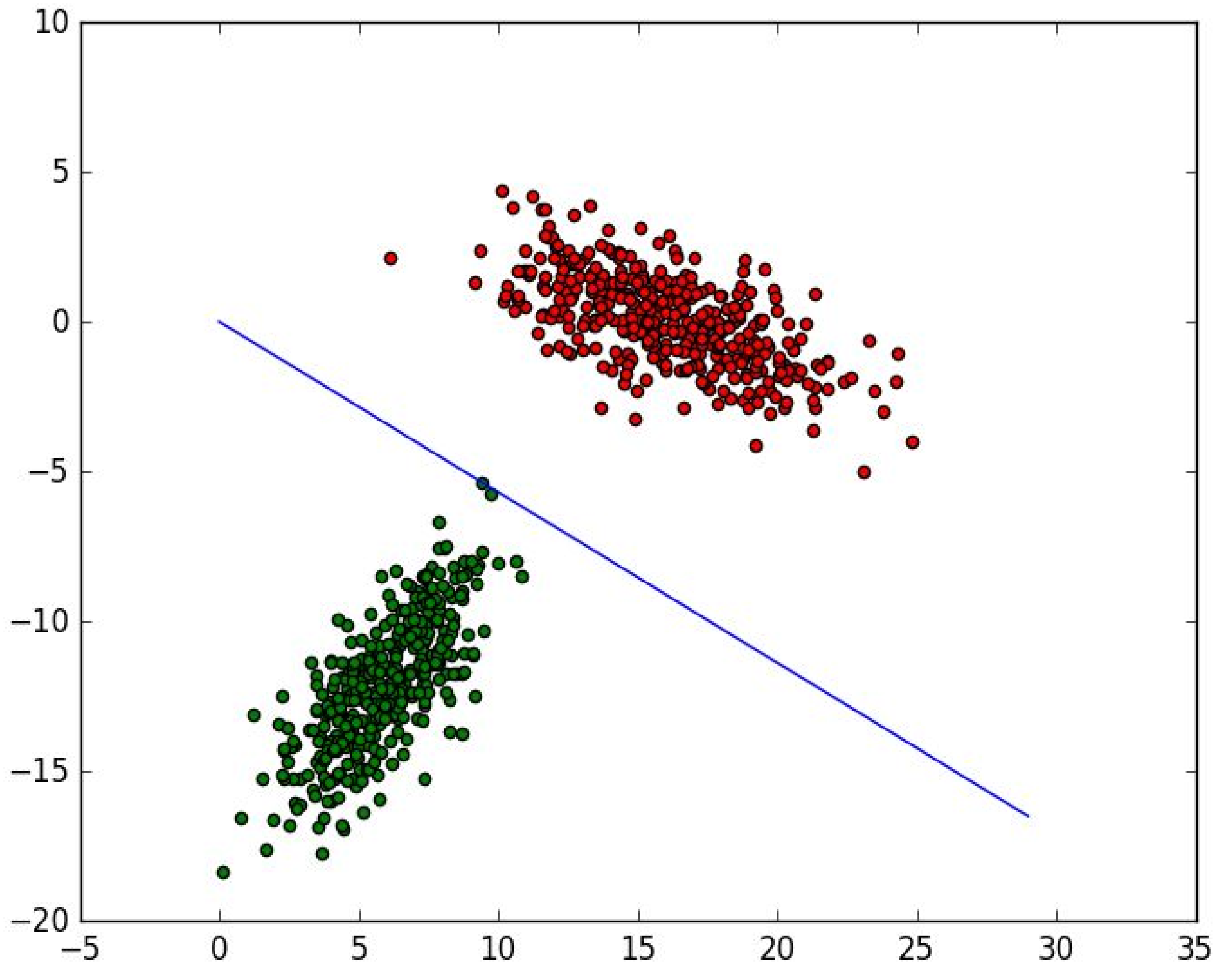
	Class 1	Class 2	Class 3	Mean
Recall	1.0	1.0	0.992	0.997
Precision	1.0	0.992	1.0	0.997
F-Measure	1.0	0.996	0.995	0.997

Inferences :

1. Perceptron Bases classifier is build on linearly separable data.
2. The accuracy of classifier quite significant as the data is already completely separated itself.
3. 75% of data is taken as training data for perceptron and 25% data is used for test as we know testing data is already linearly separated so the accuracy of the model should be near 100% which can be seen from the table shown above.
4. Perceptron takes some time to converge so sue to which we decrease the value of ETA (i.e. Proportionality constant) by large factor which makes the algorithm converge quicker than before.
5. ETA is taken as 1 at start and the value of ETA should be between 0 and 1 which will be decreasing after every iteration.

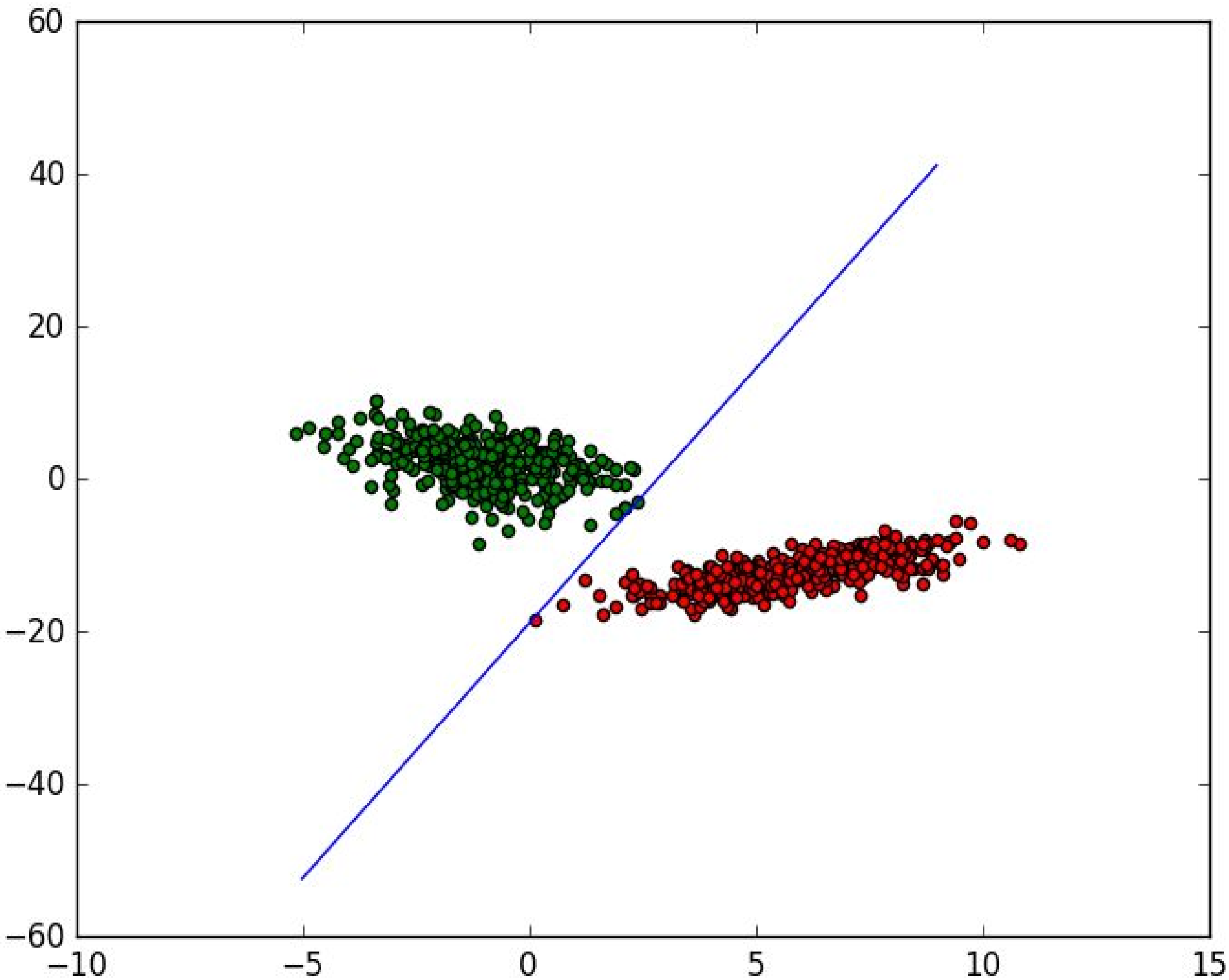
Plot of Perceptron Based classifier for different pair of classes :

(A) Plot of perceptron based classifier for class1 and class2 :



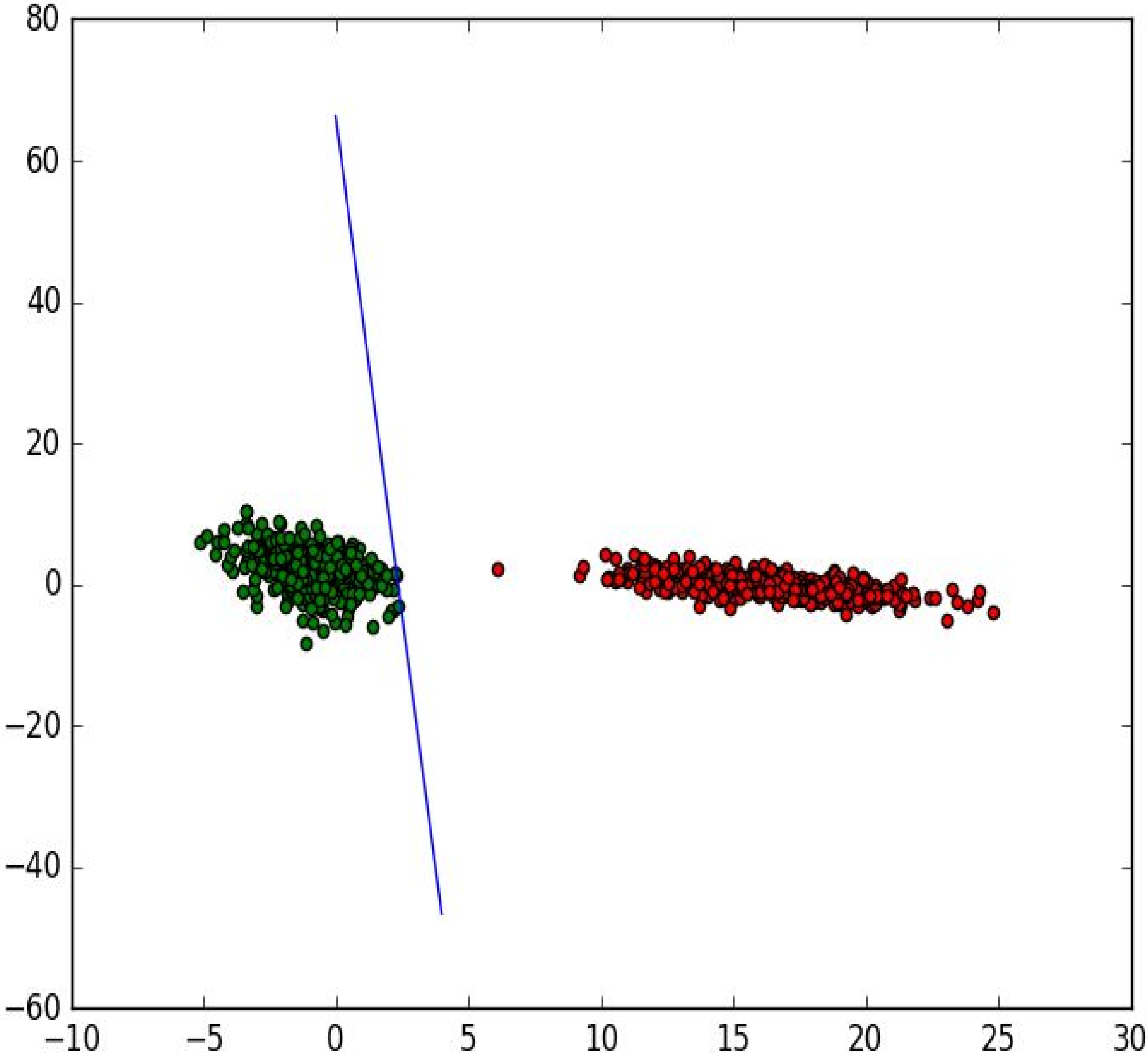
Plot for perceptron based classifier for class1 and class2 implying on dataset-1a

(B) Plot of perceptron based classifier for class2 and class3 :



Plot for perceptron based classifier for class2 and class3 implying on dataset-1a

(C) Plot of perceptron based classifier for class1 and class3 :



Plot for perceptron based classifier for class1 and class3 implying on dataset-1a

SVM Based Classifier on dataset-1 and dataset-2

For every dataset support vector machine based classifier is build for three types of kernel i.e. linear kernel, polynomial kernel, gaussian kernel (rbf). Kernel cooefficient gamma is taken as 1 for linearly separable dataset.

(A) Dataset -1(a) : Linearly Separable Data :

(i) Linear Kernel :

Accuracy : 100.00 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	125	0	0
Class 2	0	125	0
Class 3	0	0	125

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	1.0	1.0	1.0	1.0
Precision	1.0	1.0	1.0	1.0
F-Measure	1.0	1.0	1.0	1.0

(ii) Polynomial Kernel :

Accuracy : 100.00 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	125	0	0
Class 2	0	125	0
Class 3	0	0	125

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	1.0	1.0	1.0	1.0
Precision	1.0	1.0	1.0	1.0
F-Measure	1.0	1.0	1.0	1.0

(iii) Gaussian Kernel (RBF) :

Accuracy : 100.00 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	125	0	0
Class 2	0	125	0
Class 3	0	0	125

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	1.0	1.0	1.0	1.0
Precision	1.0	1.0	1.0	1.0
F-Measure	1.0	1.0	1.0	1.0

(B) Dataset -1(b) : Non-Linearly Separable Data : gamma = 0.01

(i) Linear Kernel :

Accuracy : 50.00 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	0	0	125
Class 2	0	0	125
Class 3	0	0	250

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.0	0.0	1.0	0.33
Precision	0.0	0.0	0.0	0.0
F-Measure	0.0	0.0	0.0	0.0

(ii) Polynomial Kernel :

Accuracy : 50.00 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	0	0	125
Class 2	0	0	125
Class 3	0	0	250

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.0	0.0	1.0	0.33
Precision	0.0	0.0	0.0	0.0
F-Measure	0.0	0.0	0.0	0.0

(iii) Gaussian Kernel (RBF) :

Accuracy : 96.00 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	118	7	0
Class 2	8	117	0
Class 3	4	1	245

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.94	0.93	0.98	0.95
Precision	0.90	0.93	1.0	0.94
F-Measure	0.92	0.93	0.98	0.95

(C) Dataset -2 : Class Scene Image Dataset : gamma = 1

(i) **Linear Kernel :**

Accuracy : 64.00 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	30	10	10
Class 2	10	36	4
Class 3	14	6	30

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.6	0.72	0.6	0.64
Precision	0.55	0.69	0.68	0.64
F-Measure	0.57	0.70	0.63	0.64

(ii) Polynomial Kernel :

Accuracy : 64.00 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	26	9	15
Class 2	4	43	3
Class 3	11	12	27

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.52	0.86	0.54	0.64
Precision	0.63	0.67	0.60	0.63
F-Measure	0.57	0.75	0.56	0.63

(iii) Gaussian Kernel (RBF) :

Accuracy : 54.00 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	4	4	42
Class 2	2	42	6
Class 3	5	10	35

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.08	0.84	0.7	0.54
Precision	0.36	0.75	0.42	0.51
F-Measure	0.13	0.79	0.52	0.48

Inferences :

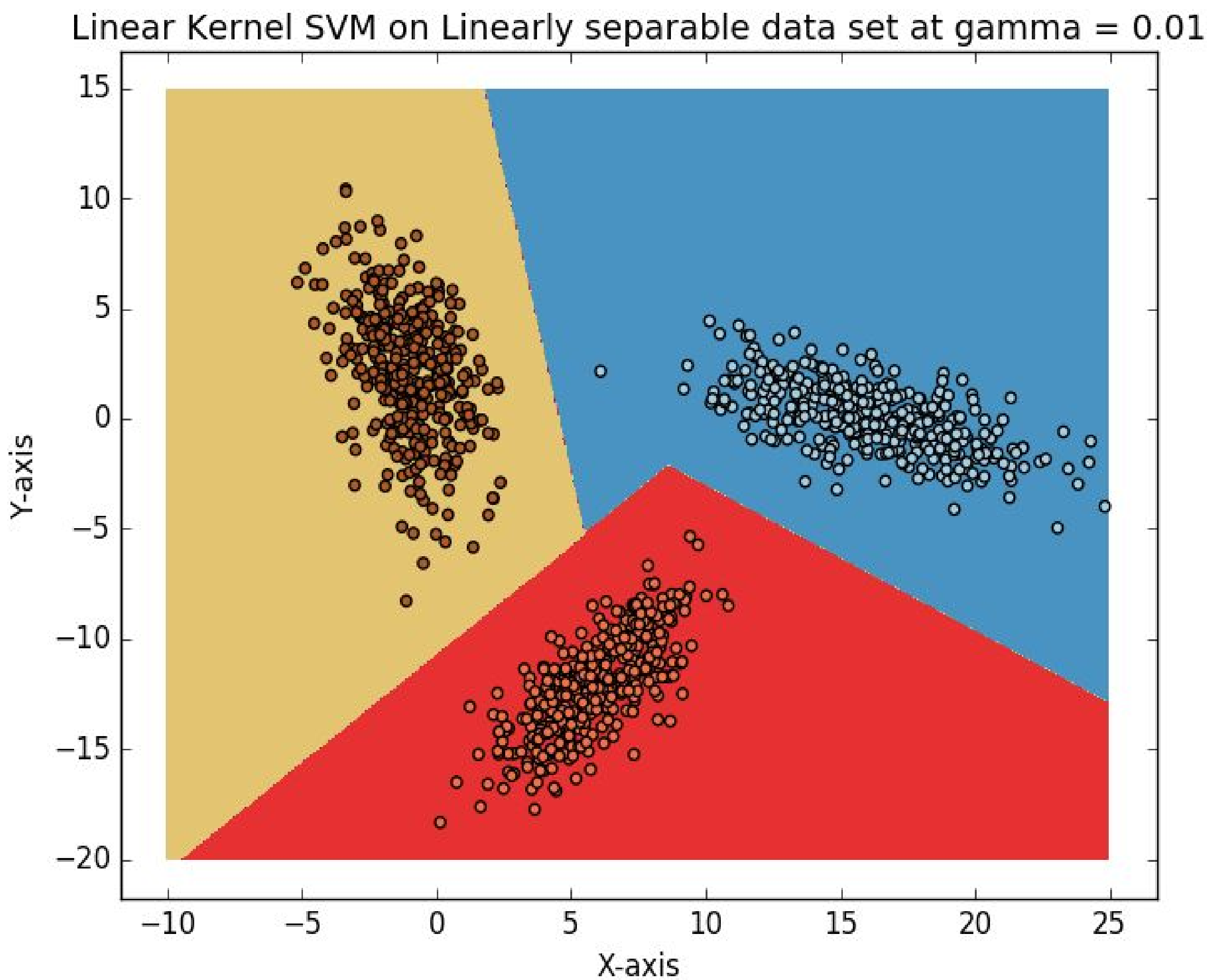
1. Support vector machine based classifier does not depend on the dimension of the input data which helps to classify non-linear data by making it a linearly separable data in higher dimension.
2. The expansion of dimension helps to classify the data more accurately as in higher dimension data can beassumed to be separated linearly.
3. SVM-Based classifier use three types of kernel i.e. linear kernel, polynomial kernel, and gaussian kernel (RBF).
4. For linearly separable data accuracy of the model comes out to be 100% for all the kernel as the data is already separated linearly.
5. For non-linearly separated data the performance of linear kernel and polynomial kernel is not satisfying as both these kernels classify data into 3rd class only but for gaussian kernel accuracy is 96% which is means the data classified using gaussian kernel is nearly correct.
6. For Class scene image data ,the accuracy of linear kernel comes out to be better than gaussian kernel which is 64%, which is good enough to classify the data.
7. The Kernel coeffiecient gamma is taken as 0.01,1 and 100 and from the observation it can be seen that increasing the value of gamma give better accuracy for all datasets and all types of kernel.
8. But for large value of gamma the accuracy decrease again so gamma should be chosen appropriate giving better accuracy than others.

Plots of SVM based classifier :

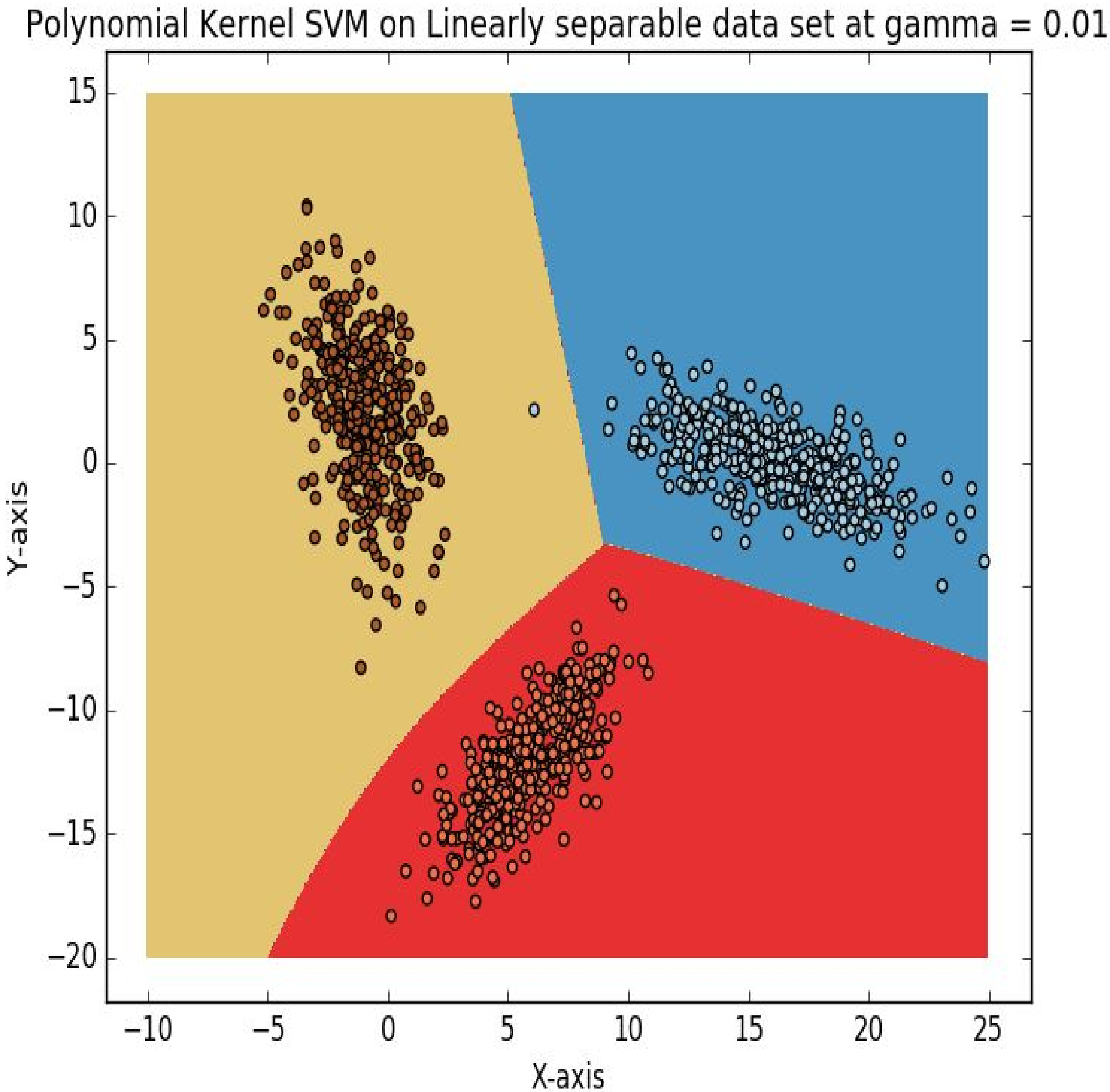
1. **Dataset -1(a) : Linearly Separable Data :**

   a. **Gamma (Kernel Coefficient) = 0.01**

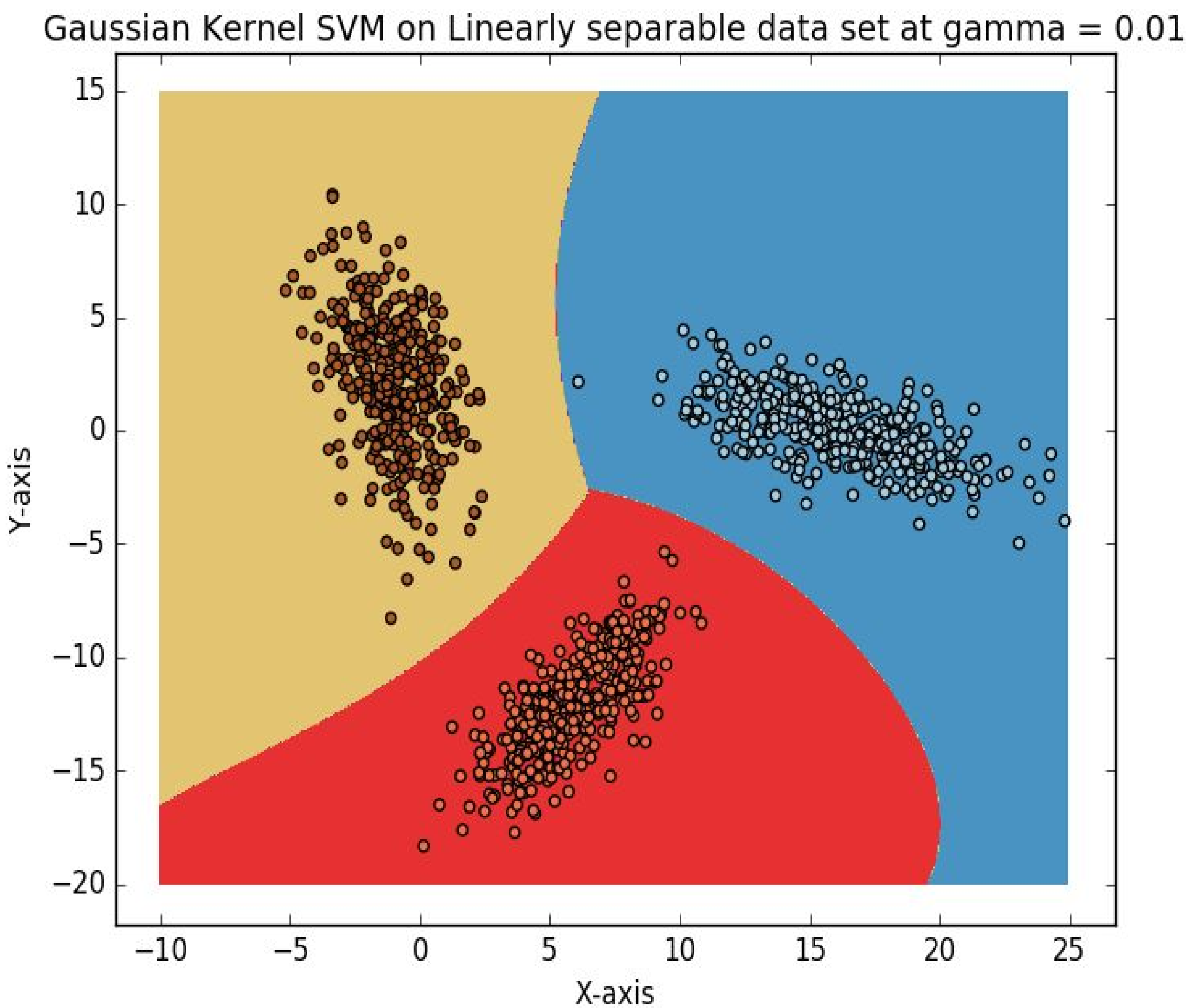
      i. **Linear Kernel**



ii. **Polynomial Kernel**

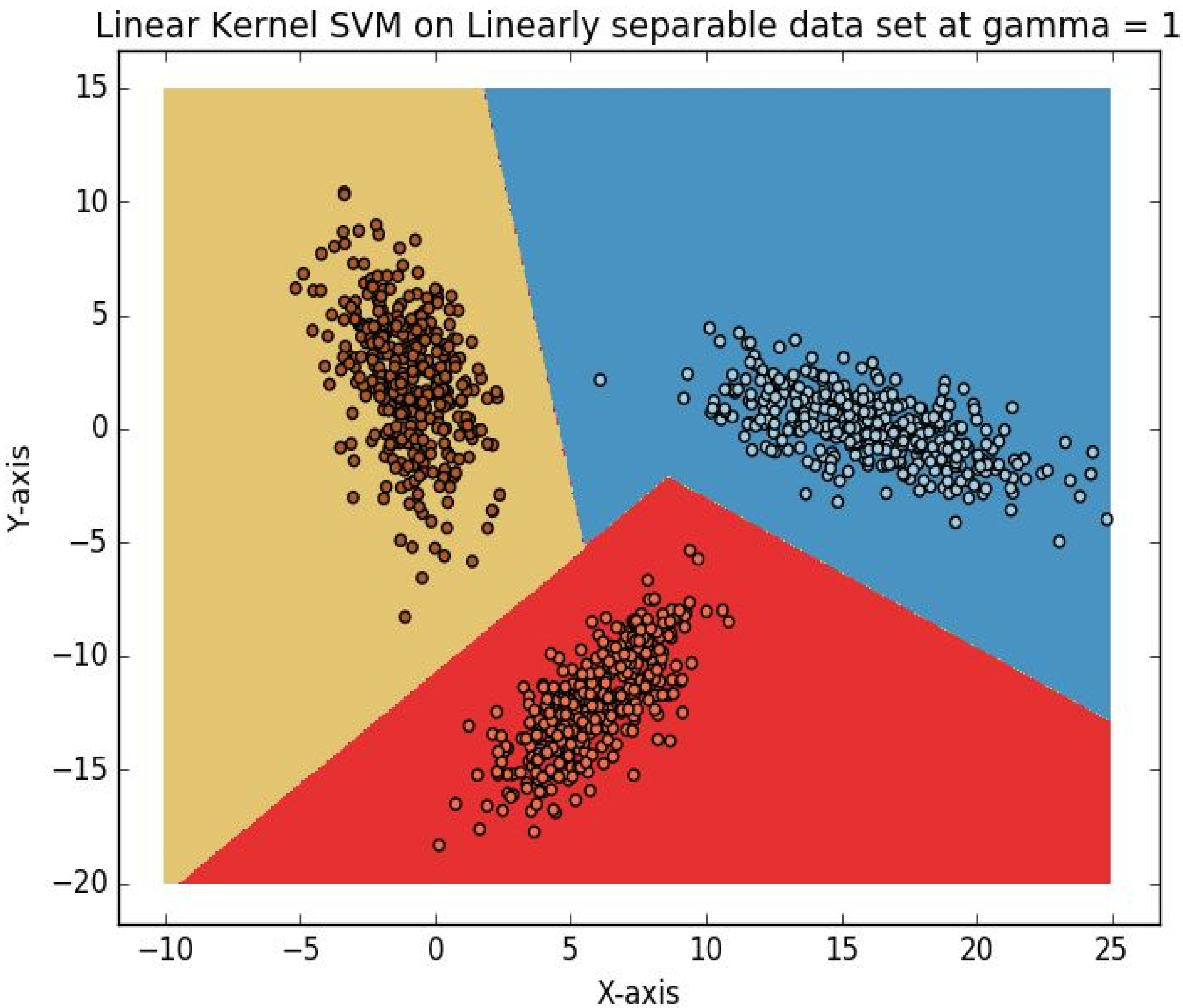


iii. Gaussian Kernel (RBF)

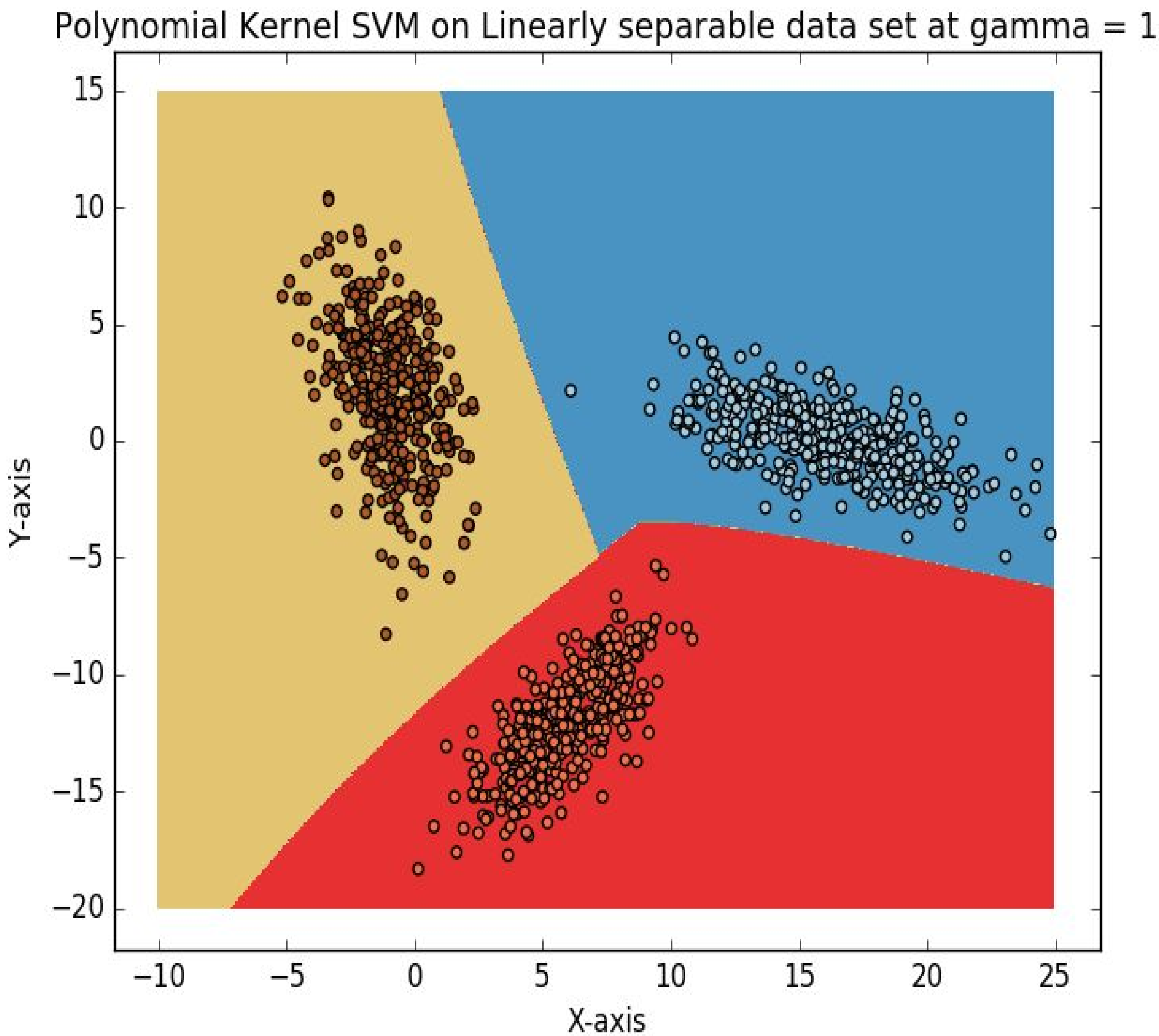


b. Gamma (Kernel Coefficient) = 1

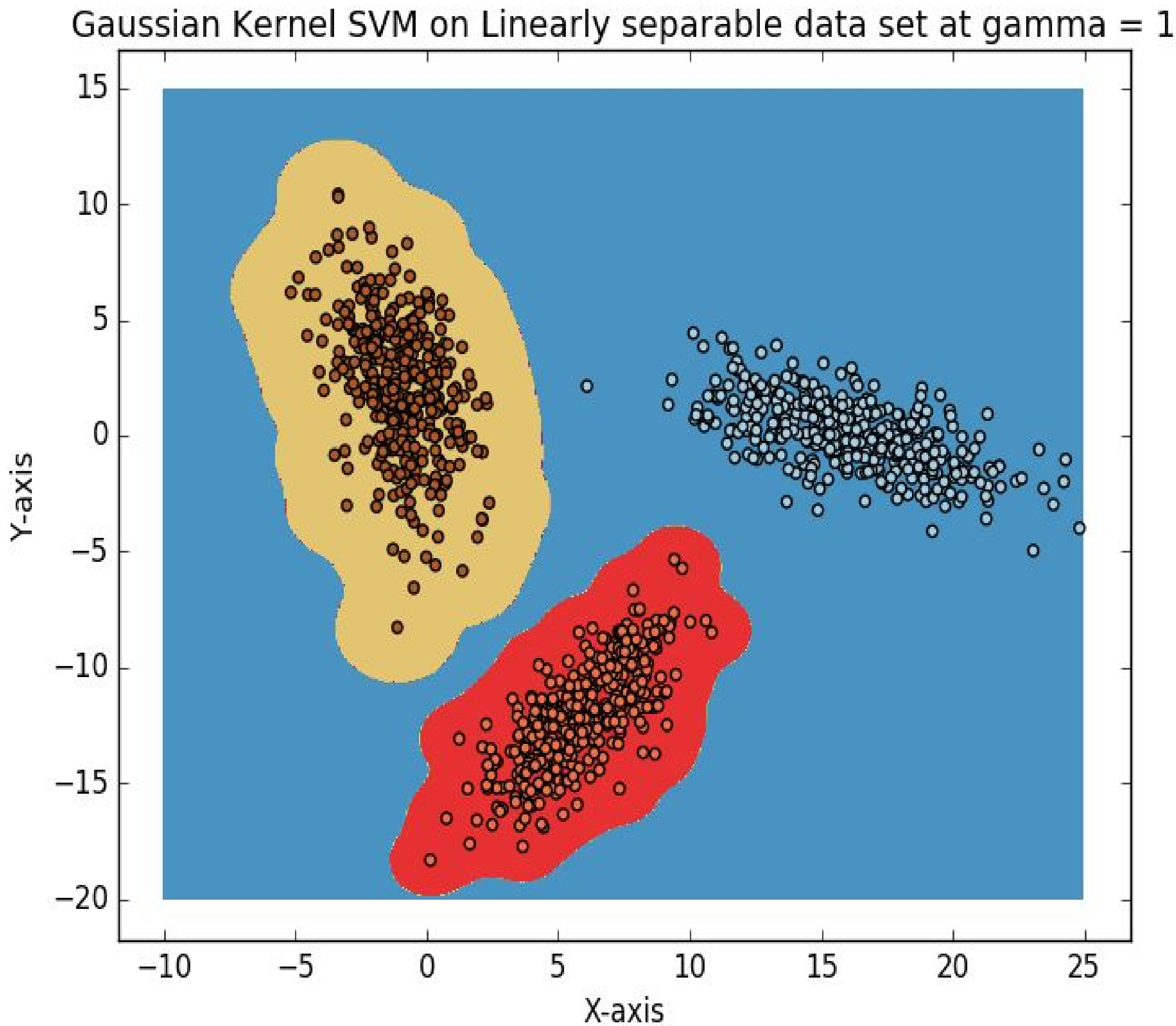
i. Linear Kernel



ii. Polynomial Kernel

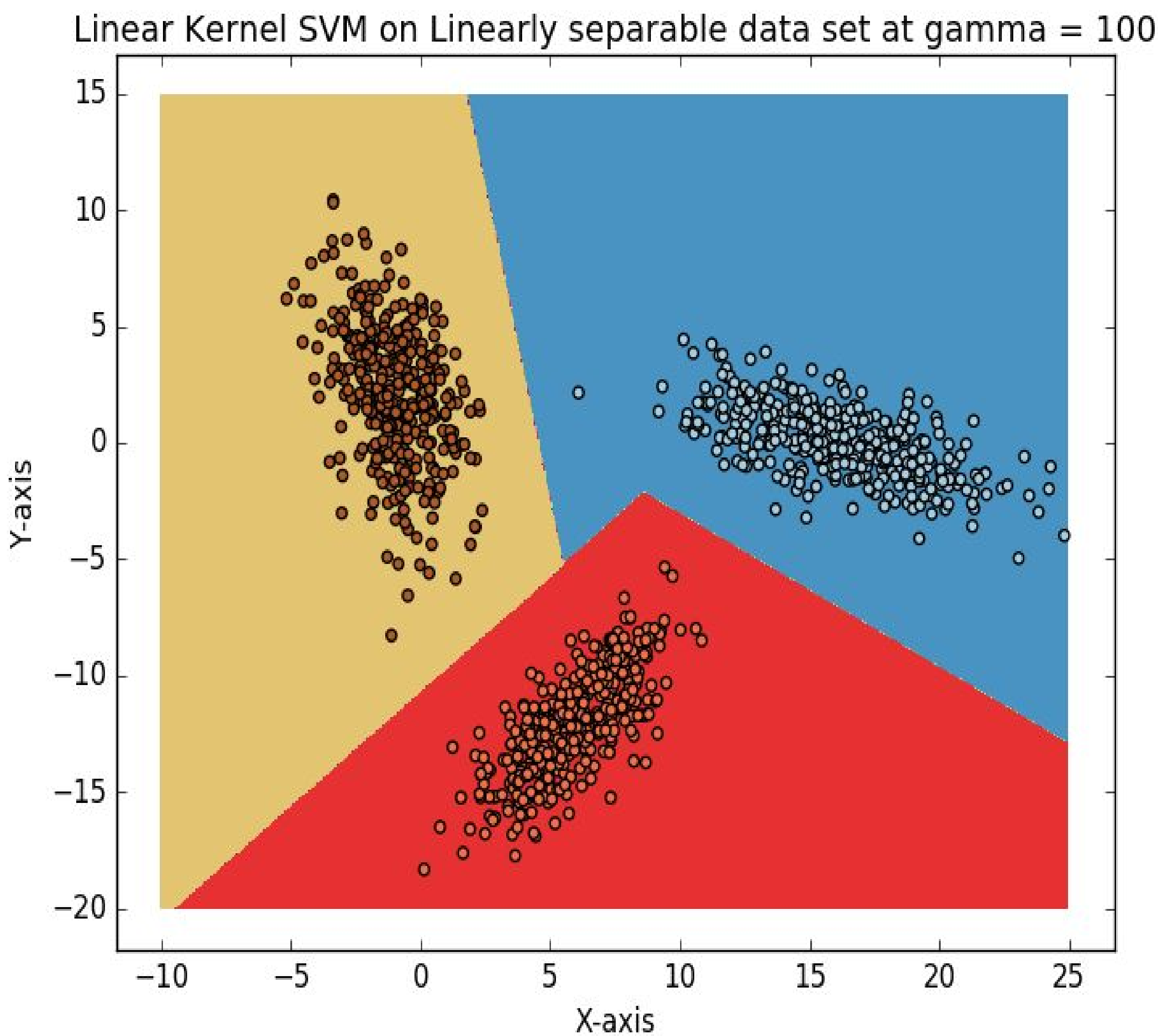


iii. Gaussian Kernel

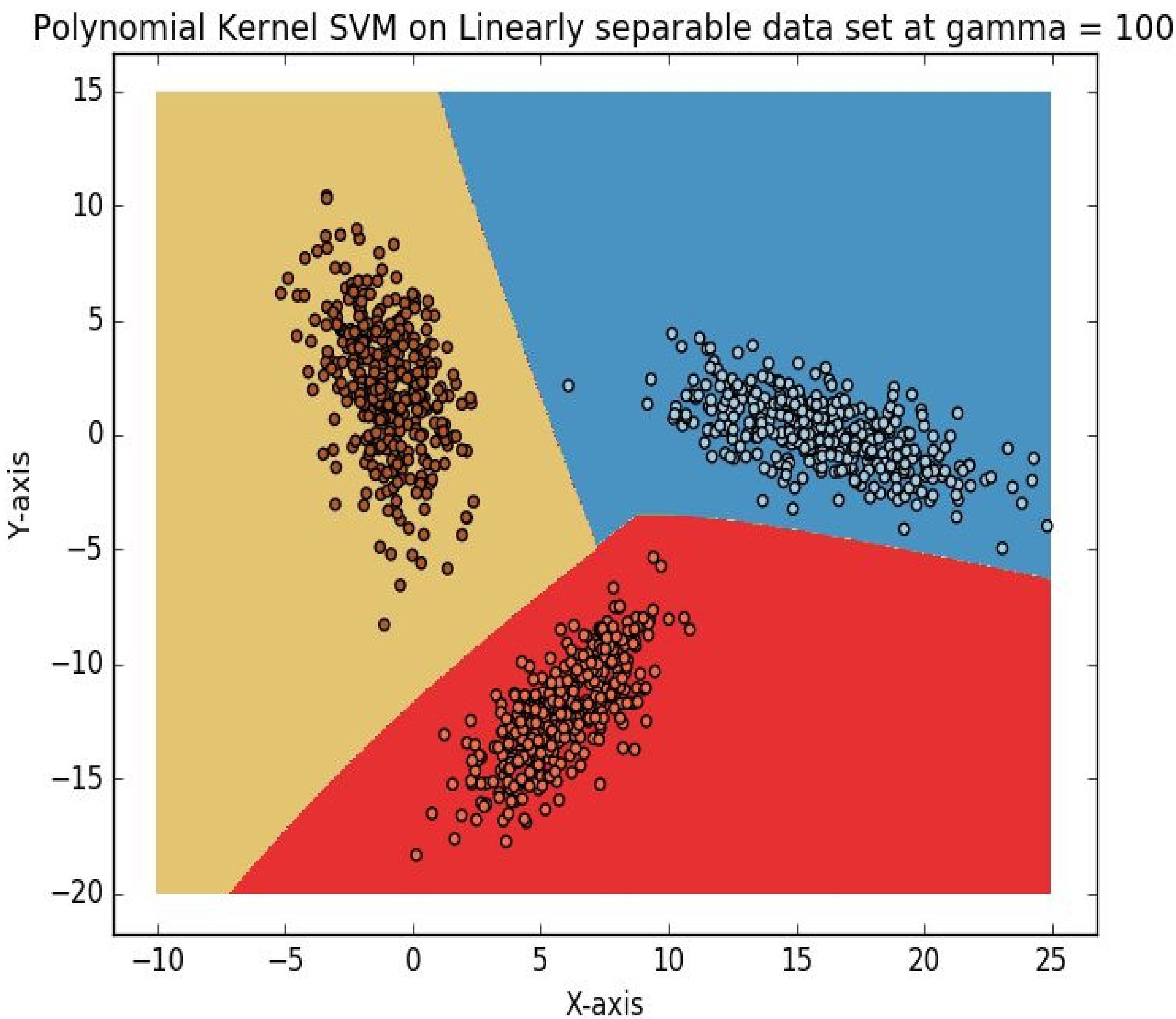


c. Gamma (Kernel Coefficient) = 100

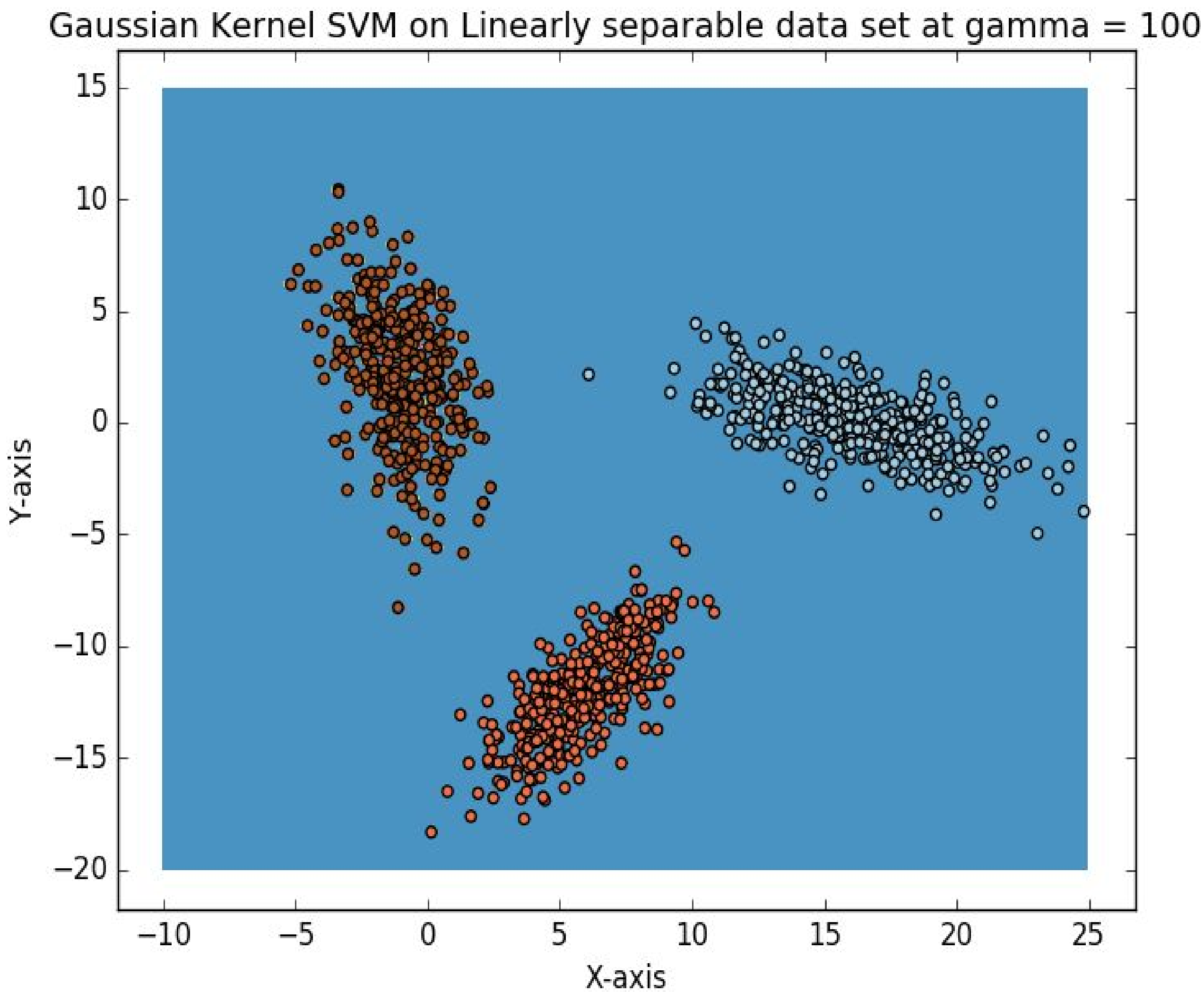
i. Linear Kernel



ii. Polynomial Kernel



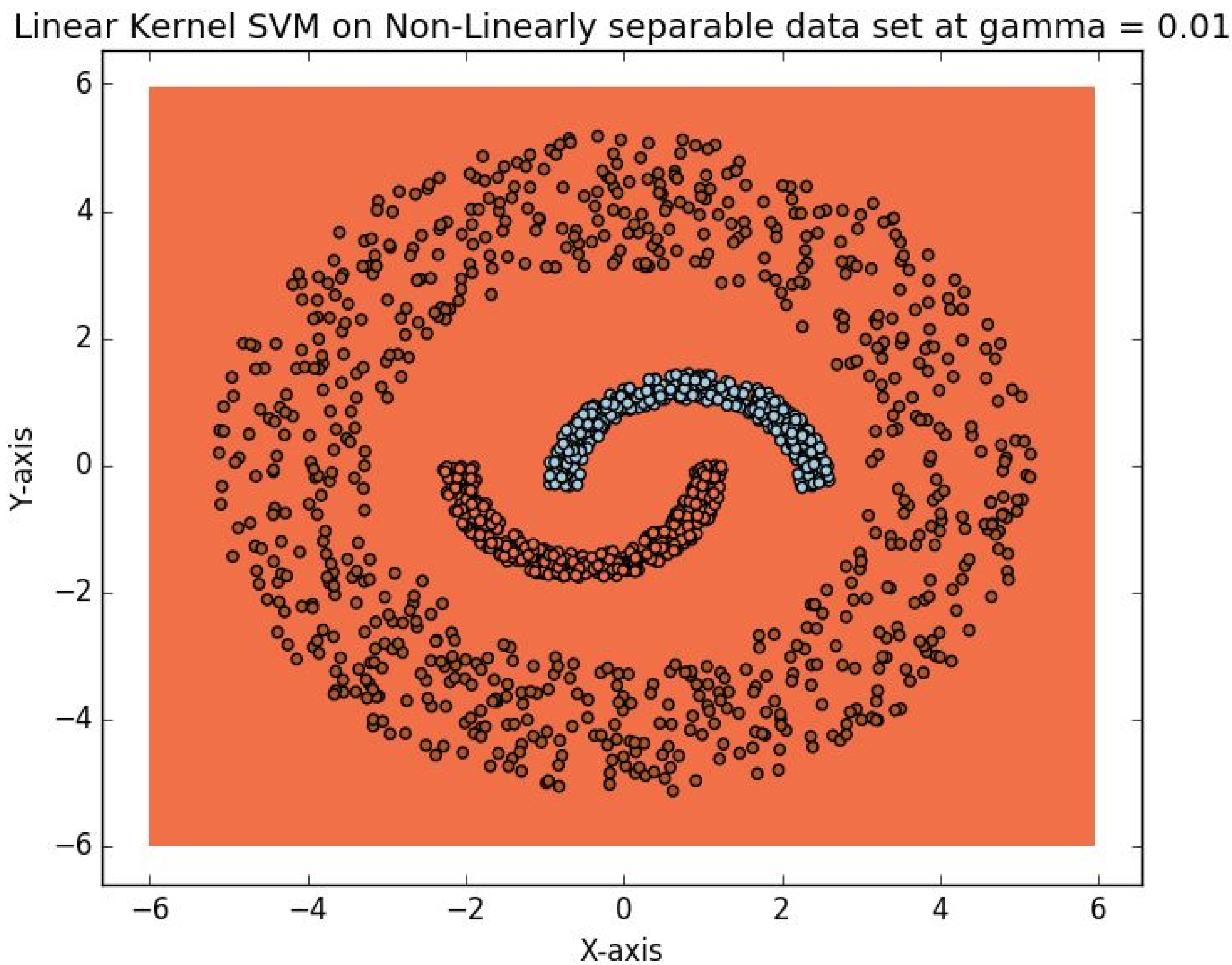
iii. Gaussian Kernel



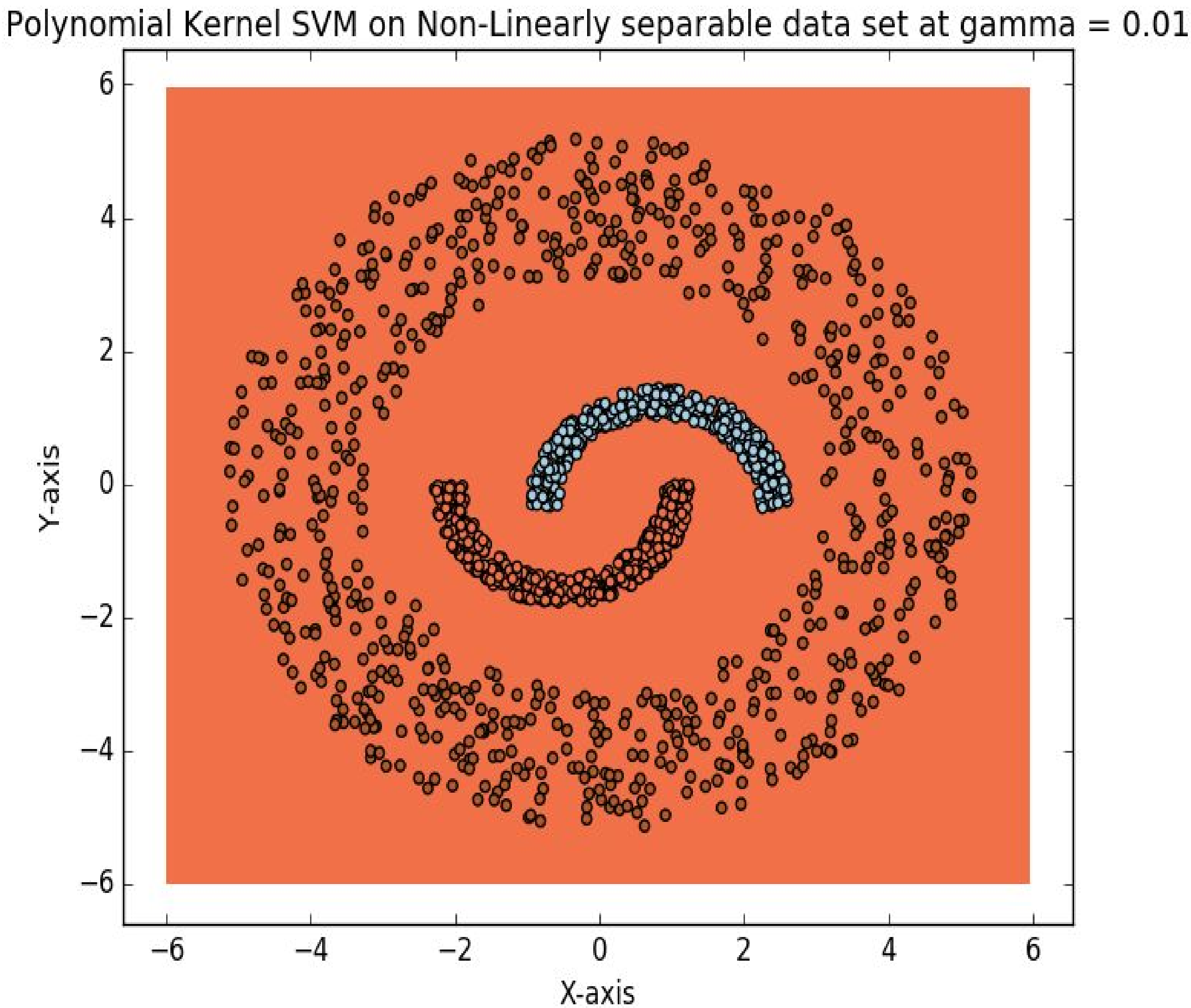
2. **Dataset -1(b) : Non-Linearly Separable Data :**

   a. **Gamma (Kernel Coefficient) = 0.01**

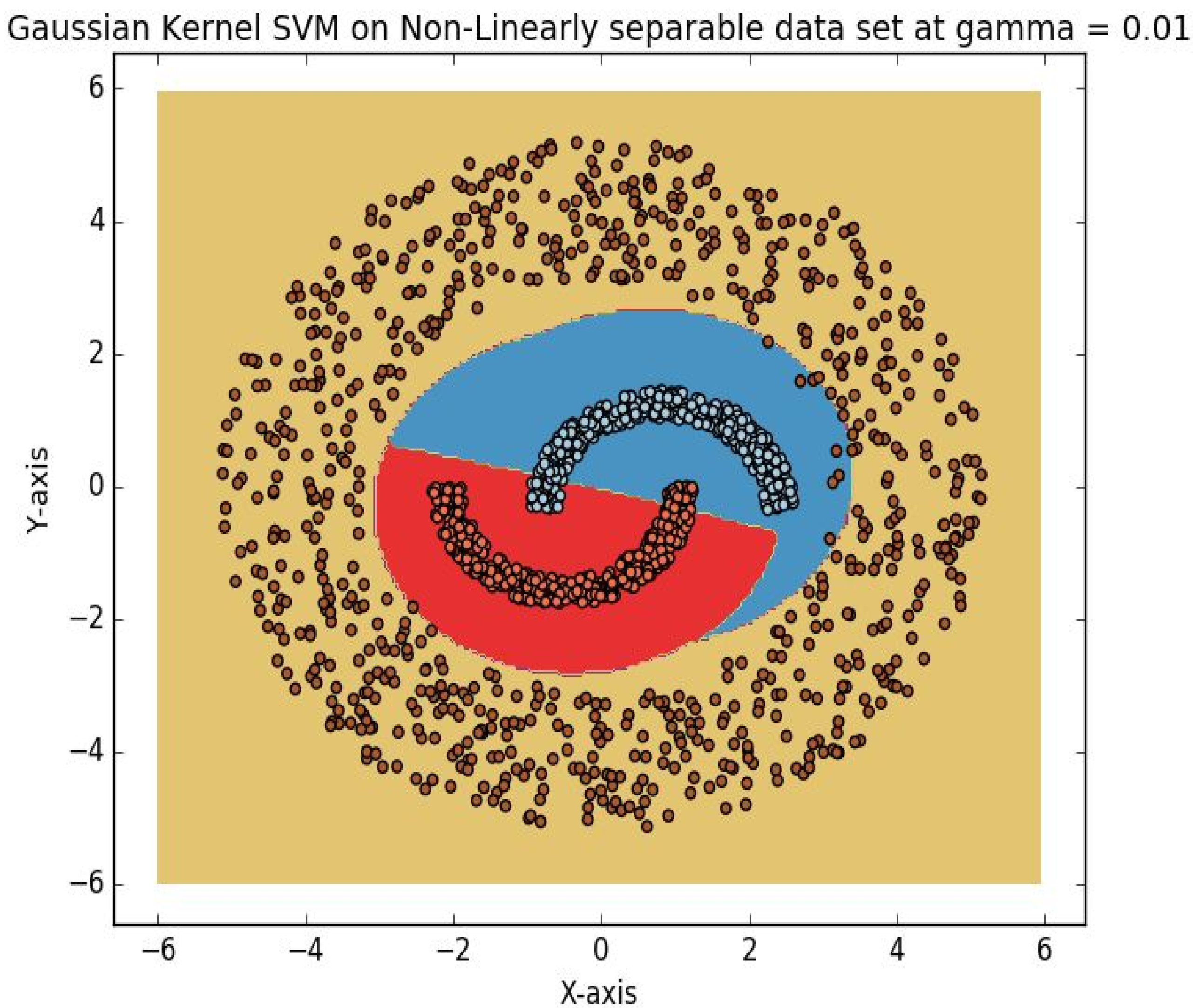
      i. **Linear Kernel**



ii. Polynomial Kernel

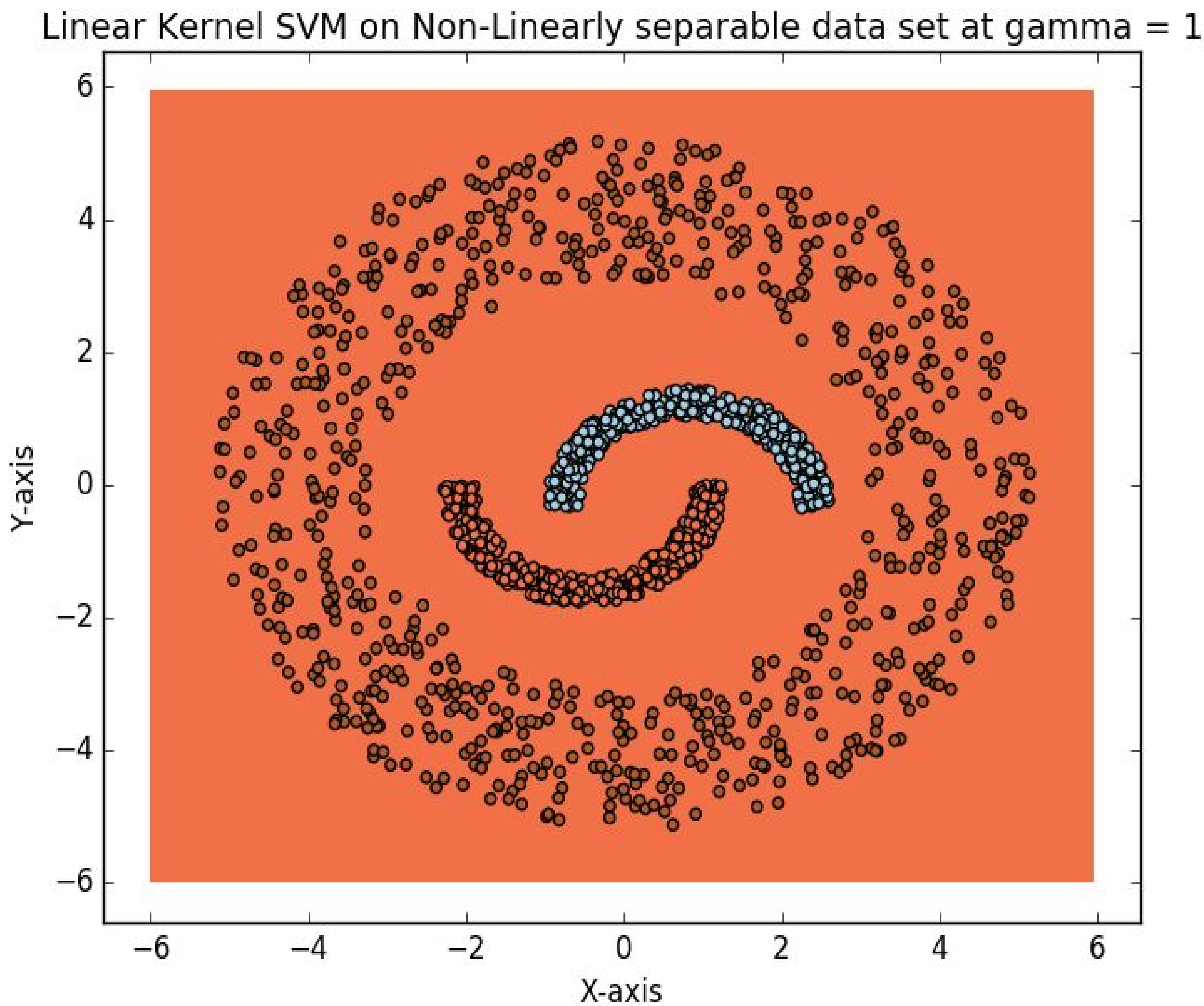


iii. Gaussian Kernel (RBF)

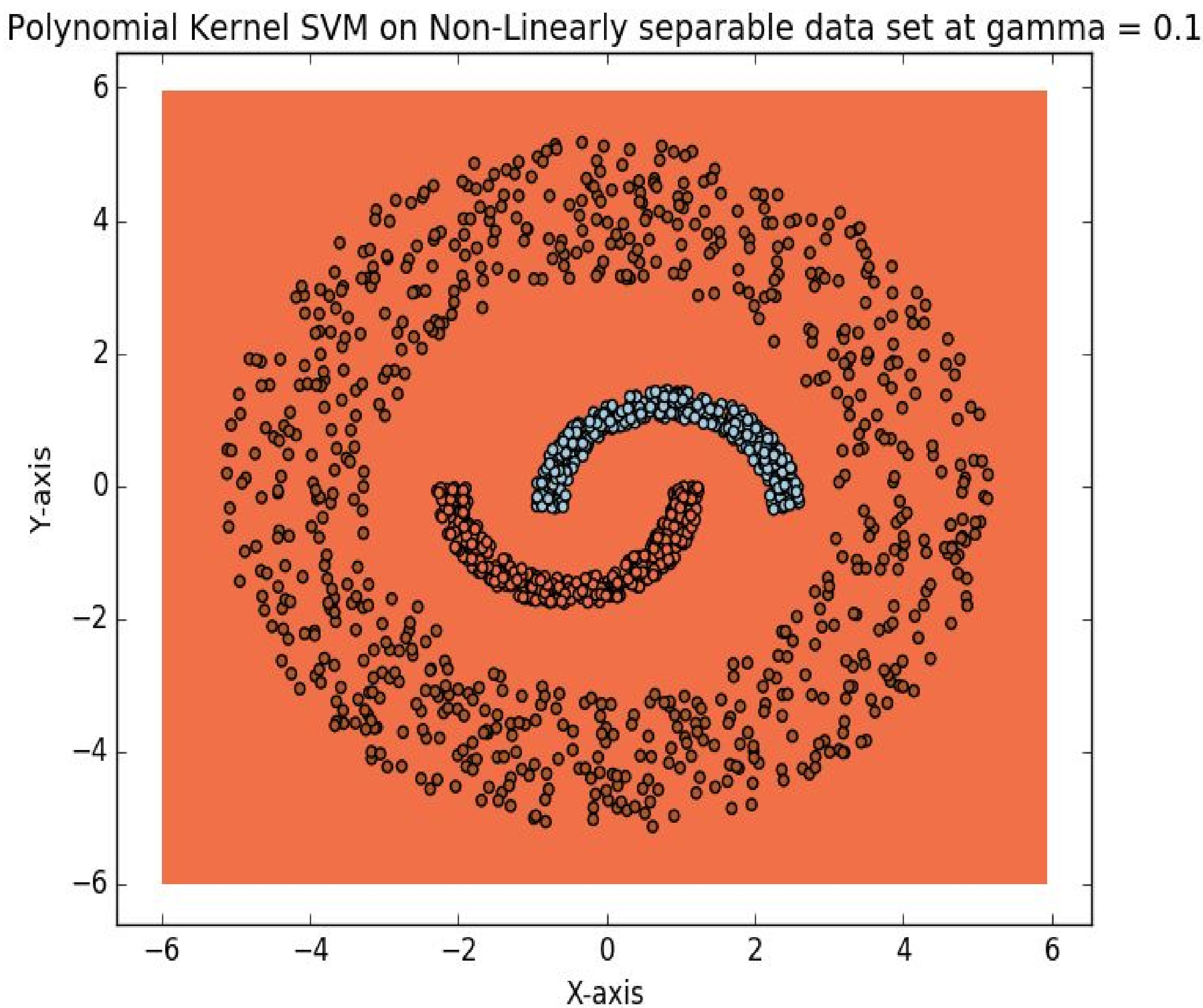


b. Gamma (Kernel Coefficient) = 1

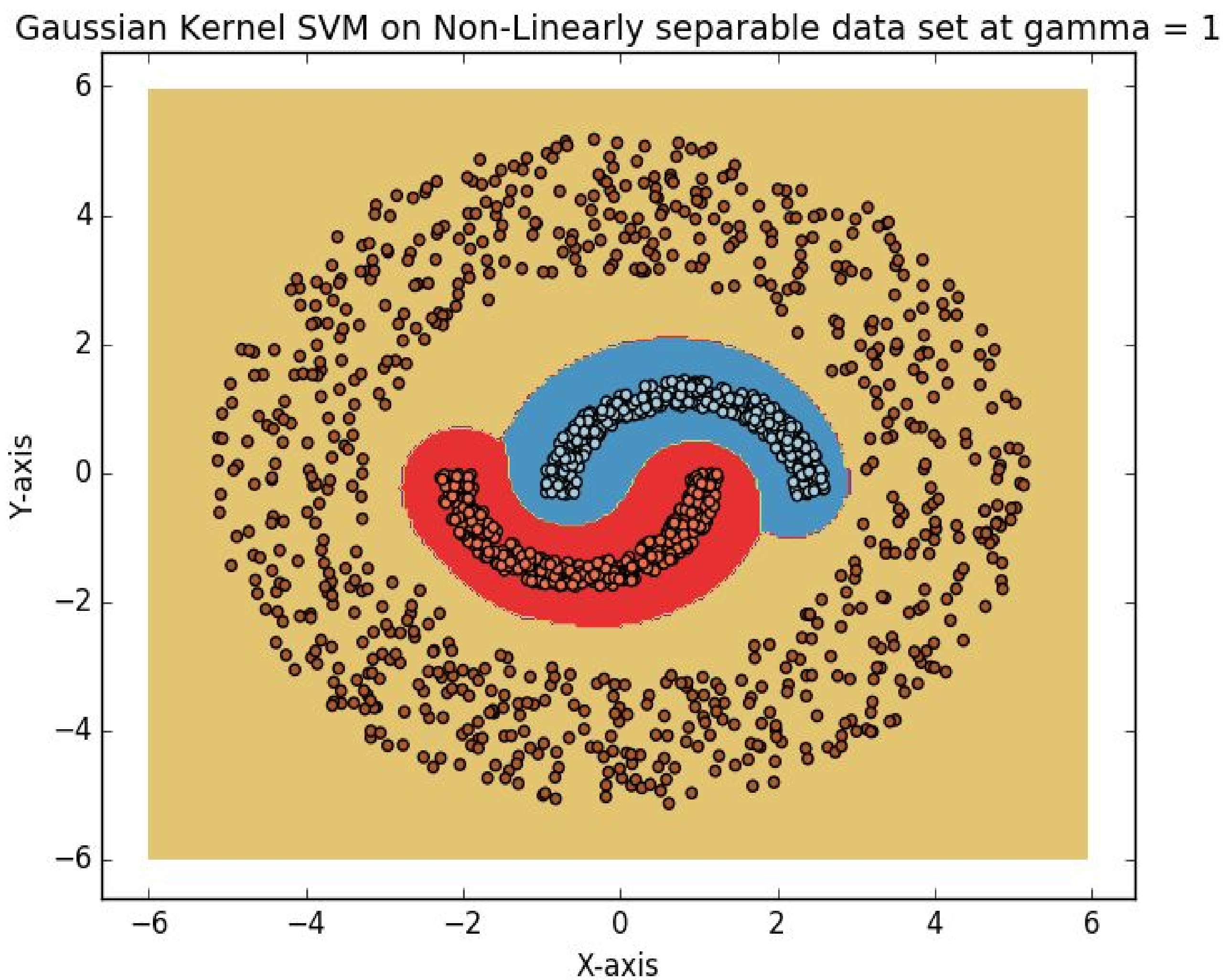
i. Linear Kernel



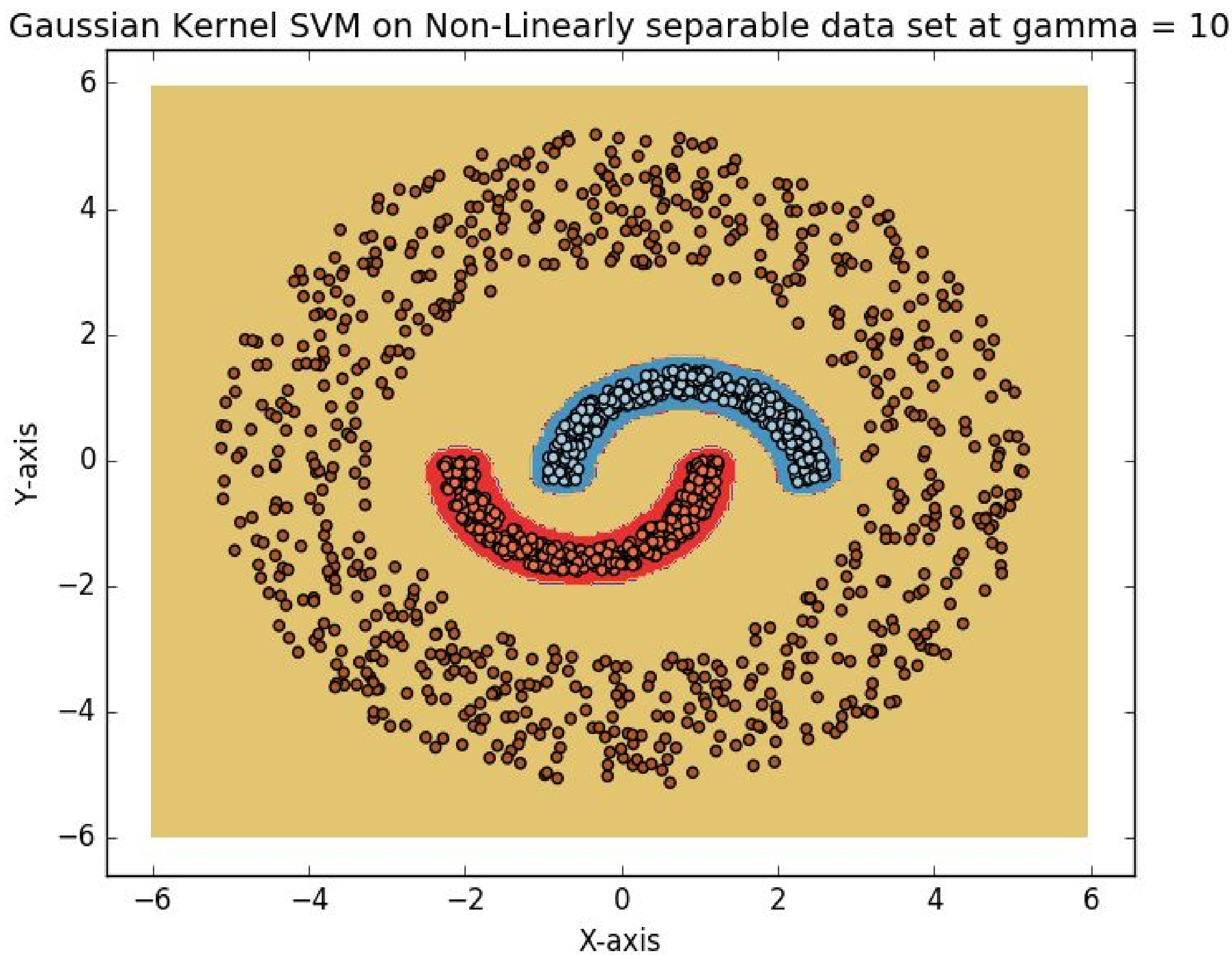
ii. Polynomial Kernel



iii. Gaussian Kernel



c. Gaussian Kernel For Non-Linearly Separable data with gamma = 10



Inferences :

1. For linearly separable dataset the plot for gaussian kernel is better than non-linear kernel and linear kernel.

2. As the accuracy for gaussian kernel is better than linear and non-linear kernel which is shown above.

3. On increasing the value of gamma from 0.01 to 1, it doesn't affect much for the case of linear and non-linear kernel but for gaussian kernel the increase in the value of gamma gives better decision boundary.

4. On further increase in the value of gamma decreases the accuracy of linear and polynomial kernel but for gaussian it makes more compact decision boundary in short increases accuracy.

5. For non-linear separable data, the plots, decision boundary or linear kernel and non-linear kernel is very close to the data points thus classifying the data god enough.

6. For gaussian Kernel the accuracy for non-linearly separable is not better than linear kernel.

7. The increase in the value of gamma does not affect much the decision boundary for linear and polynomial kernel.

8. But increase in gamma increases the accuracy for gaussian kernel thus improve the decision boundary plot for non-linearly separable data.

9. Further increase in the value of gamma, increase the accuracy of gaussian kernel SVM-based classifier for non-linearly separable data.

Comparison of all classifiers on basis of accuracy

1. For linearly separable dataset :

    a. Bayes classifier using unimodal gaussian - 100%
    b. Perceptron Based Classifier - 99.73%
    c. SVM-Based Classifier - 100%
    d. Bayes classifier using GMM (FDA) - 100%

For linearly separable dataset the accuracy of all the classifiers is same approximately which means for linear data all the classifier works well.

2. For non-linearly separable data :

    a. Bayes classifier using unimodal gaussian - 94.4%
    b. Bayes classifier using GMM - 96.56%
    c. Bayes classifier using GMM (FDA) - 80.04%
    d. SVM-Based Classifier - 96.00%

For non-linearly separable data, bayes classifier using GMM works better than all other classifiers.

3. For Real World Dataset :

    a. Bayes Classifier using unimodal gaussian - 84.68%
    b. Bayes Classifier using GMM - 84.04%
    c. Bayes Classifier using KNN - 76.03%
    d. Bayes Classifier using DHMM - 67.76%
    e. Bayes Classifier using GMM (FDA) - 58.86%
    f. SVM-Based Classifier - 64.00%

c,d (i.e. KNN, DHMM) are the classifiers for speech data and e,f (i.e. GMM(FDA) , SVM) classifiers for class scene image data.

Conclusion :

1. For linearly separable dataset, bayes classifier and SVM-based classifier works excellent with high accuracy (close to 100%) which is good enough to classify data into their corresponding classes.

2. For non-linearly separable dataset, bayes classifier with gaussian mixture model and SVM-based classifier works well with accuracy nearly 96% which is favourable to classify data.

3. For real world dataset, the bayes classifier for unimodal gaussian is better than gaussian mixture model giving accuracy about 84.68% which is acceptable to classify data to respective classes.

4. For CV segment speech dataset, bayes classifier using KNN works better than discrete hidden markov model giving accuracy 76.03%.

5. For class scene image dataset,SVM-based classifier is better than bayes classifier using GMM with reduced dimension using FDA.

6. From the observations it can be observed the for one type of data, some classifiers give better classification accuracy than others and for some other type of data, some other classifiers give better accuracy.

7. SVM-based classifier is giving reasonable results for classification as this classifier is free from the curse of dimensionality which helps to classify data in higher dimensions.

8. A non-linearly separable data can be seen as linearly separable data in higher dimensions thus helps to increase the accuracy of the training model in case of SVM-based classifier.

References :

1) Class Notes

2) Lecture slides

3) https://en.wikipedia.org/wiki/Perceptron

4) https://www.stat.cmu.edu/~cshalizi/350/lectures/25/lecture-25.pdf

5) https://en.wikipedia.org/wiki/Support_vector_machine

6) https://en.wikipedia.org/wiki/Linear_discriminant_analysis

7) http://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html

8) https://matplotlib.org/

9) https://www.youtube.com/watch?v=azXCzI57Yfc

10) https://www.youtube.com/watch?v=wL2aVUjDdoo

11) https://www.youtube.com/watch?v=1NxnPkZM9bc