Pattern Recognition

Group 12

Programming Assignment 1

Bayes Classifier

September 16, 2018

Group Members	
Name	**Roll No.**
Arpit Batra	B16047
Ayush Meghwani	B16127
Navneet Sharma	B16065

Faculty Mentor - Prof. Dileep A. D.

Contents

List of Figures

1 Introduction

Pattern recognition is the branch of machine learning which focuses on recognizing patterns or similarities in data. Classification is a supervised machine learning method of identifying to which of a set of classes (sub-populations) a new observation belongs, on the basis of a training set of data observations are made. Based on the observation drawn from the data class of new observations is decided.
Example: Assigning a given image of a person into "male" or "female" classes, assigning a diagnosis to a given patient as described by observed characteristics of the patient.

1.1 Problem Statement

Classification of given data sets using Bayes Classifier.

1.2 Data Sets Provided

- Linearly Separable Data
- Non-Linearly Separable Data
- Real Data

Data contains 3 classes and has only two features. 75% of the data is to be used as Training set and rest of the data is to be used as the Testing set.

1.3 Implemented Classifiers

- Bayes Classifier
  - Same co-variance matrix for all classes, of the form: $\sigma^2 I$.
    (where σ is same for all classes i.e., average of variances of all classes)
  - Same full co-variance matrix for all classes, of the form: Σ
    (where Σ is an average matrix of all classes)
  - co-variance matrix is different for each class and is diagonal
    (Σ is a forced-diagonal matrix of full co-variance matrix of class)
  - Full co-variance matrix is different for each class.

1.4 Learning Objective

Observe the decision boundaries for different datasets under different classifiers and explain the reasons for them. Observe performance accuracy of different classifiers for different types of data sets.

2 What is Bayes Classifier?

Bayesian decision theory is a fundamental statistical approach to the problem of pattern classification. This approach is based on quantifying the tradeoffs between various classification decisions using probability and the costs that accompany such decisions. It makes the assumption that the decision problem is posed in probabilistic terms, and that all of the relevant probability values are known.

3 Data Set Types

3.1 Linearly Separable Data

A set of data is linearly separable if there exists a line/plane (hyperplane in case of more than 3 dimensions) separating all the data points of one class from the data points of other classes. If X1 and X2 are two sets of points in n-dimensional space, Then X1 and X2 are linearly separable if there exists n+1 real numbers $w_1, w_2, w_3, \ldots, w_n, k$ such that every point $x \epsilon X_1$ satisfies

$$\Sigma_{i=1}^{n} \omega_i x_i < k \tag{1}$$

and every point $x \epsilon X_1$ satisfies

$$\Sigma_{i=1}^{n} \omega_i x_i > k \tag{2}$$

Equation of boundary between such linearly separable data points can be written as

$$\omega x - k = 0 \tag{3}$$

3.2 Non-Linearly Separable Data

A set of data is non-linearly separable if there exists a decision surface that has a quadratic form or hyper-quadric in nature (in case of more than 3 dimensions) separating all the data points of one class from the data points of other classes. These surfaces may have hyper-paraboloids, hyper-ellipsoids, hyperspheres, hyper-hyperboloids as their general forms.

3.3 Real Data

A set of data is real data if there is no particular shaped decision surface separating all the data points of one class from the data points of other classes. The decision boundary surface may be of polynomial or any other shape.

4 Cases for Classifiers

4.1 Case 1: $\Sigma_i = \sigma^2 I$

This is the simplest case when every class has equal variance i.e. σ^2. This case is realized when features are statistically independent. In this case, co-variance matrix is diagonal in nature with every element equal to σ^2. . Geometrically, this case corresponds to equal sized hyper-spherical clusters with ith class being centered around mean vector μ_i. In this case the discriminant function obtained is as

$$g_i(x) = \left(\frac{1}{\sigma^2}\mu_i\right)^T x - \frac{-1}{2\sigma^2}\mu_i^T\mu_i + ln(P(C_i)) \tag{4}$$

$$g_i(x) = w_i^t x + w_{i0} \tag{5}$$

where $w_i = \frac{1}{\sigma^2}\mu_i$, $w_{i0} = \frac{-1}{2\sigma^2}\mu_i^T\mu_i + ln(P(C_i))$.

Decision surface thus obtained will be of the form:

$$w^T(x - x_0) = 0 \tag{6}$$

where, $w = \frac{1}{\sigma^2}(\mu_1 - \mu_2)$, $x_0 = \frac{\mu_1+\mu_2}{2} - \frac{\sigma^2(\mu_1-\mu_2)}{||\mu_1-\mu_2||^2} ln\left(\frac{P(C_1}{P(C_2})\right)$

4.2 Case 2: $\Sigma_i = \Sigma$

This is the case when every class has same co-variance matrix. This can be achieved by calculating co-variance matrix for all classes and assigning co-variance matrix the value equal to average of all co-variance matrices. Mathematically,

$$\Sigma = \frac{1}{N}\Sigma_{i=0}^{N}\Sigma_i \tag{7}$$

In this case the discriminant function obtained is as

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i) + P(C_i) \tag{8}$$

$$g_i(x) = w_i^t x + w_{i0} \tag{9}$$

where, $w_i = \Sigma_{-1}\mu_i$, $w_{i0} = -\frac{1}{2}(\mu_i)^T\Sigma^{-1}(\mu_i) + P(C_i)$
Decision surface thus obtained will be of the form:

$$w^T(x-x_0) = 0 \tag{10}$$

where, $w = \Sigma^{-1}(\mu_1-\mu_2)$, $x_0 = \frac{\mu_1+\mu_2}{2} - \frac{(\mu_1-\mu_2)}{(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2)} ln\left(\frac{P(C_1}{P(C_2}\right)\right)$

4.3 Case 3: Σ_i = forced diagonal matrix

This is the case when every class has diagonal co-variance matrix which is extracted from full co-variance matrix. This can be achieved by equating non diagonal element of full co-variance matrix to zero. In this case the discriminant function obtained is as

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i) + P(C_i) \tag{11}$$

$$g_i(x) = \frac{1}{2}x^t(\Sigma^{-1})x + (\Sigma^{-1}\mu_i)^T x + \frac{1}{2}\mu_i^T\Sigma^{-1}\mu_i + ln(P(C_i)) \tag{12}$$

$$g_i(x) = x^t W_i x + w_i^t x + w_{i0} \tag{13}$$

where $W_i = \frac{1}{2}\Sigma^{-1}$, $w_i = \Sigma^{-1}\mu_i$, $w_{i0} = \frac{1}{2}\mu_i^T\Sigma^{-1}\mu_i + P(C_i)$
Decision surface thus obtained will be of the form:

$$g_i(x) = x^t W_i x + w_i^t x + w_{i0} \tag{14}$$

where, $W = \frac{1}{2}(\Sigma_1^{-1}+\Sigma_2^{-1})$, $w = (\Sigma^{-1}\mu_1)^T+(\Sigma^{-1}\mu_2)^T$, $w_0 = \frac{1}{2}(\mu_1^T\Sigma^{-1}\mu_1)+$ $\mu_2^T\Sigma^{-1}\mu_2 + ln\left(\frac{P(C_1)}{P(C_2)}\right)$

4.4 Case 4: Σ_i = arbitrary

This is the case when every class has different co-variance matrix. In this case the co-variance matrix is full co-variance matrix. In this case the discriminant function obtained is as

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i) - \frac{1}{2}ln\left(P(|\Sigma^{-1}|)\right) + P(C_i) \tag{15}$$

$$g_i(x) = \frac{1}{2}x^T\Sigma^{-1}x + (\Sigma^{-1}\mu_i)^Tx + \frac{1}{2}\mu_i^T\Sigma^{-1}\mu_i - \frac{1}{2}ln\left(P(|\Sigma^{-1}|)\right) + P(C_i) \quad (16)$$

$$g_i(x) = x^tW_ix + w_i^tx + w_{i0} \quad (17)$$

where, $W = \frac{1}{2}(\Sigma^{-1}$, $w = (\Sigma^{-1}\mu_i$, $w_{i0} = \frac{1}{2}\mu_i^T\Sigma^{-1}\mu_i - \frac{1}{2}ln\left(P(|\Sigma^{-1}|)\right) + P(C_i)$
Decision surface thus obtained will be of the form:

$$g_i(x) = x^tW_ix + w_i^tx + w_{i0} \quad (18)$$

where, $W = \frac{1}{2}(\Sigma_1^{-1} + \Sigma_2^{-1})$, $w = (\Sigma^{-1}\mu_1)^T + (\Sigma^{-1}\mu_2)^T$, $w_0 = \frac{1}{2}(\mu_1^T\Sigma^{-1}\mu_1) + \mu_2^T\Sigma^{-1}\mu_2 - \frac{1}{2}ln\left(\frac{P(|\Sigma_1|)}{P(|\Sigma_2|)}\right) + ln\left(\frac{P(C_1)}{P(C_2)}\right)$

5 Plots : Linear Separable

5.1 Case 1: $\Sigma_i = \sigma^2 I$

Here we calculated the three co-variance matrices for each case then took their average as co-variance matrix for all the classes and made it diagonal.

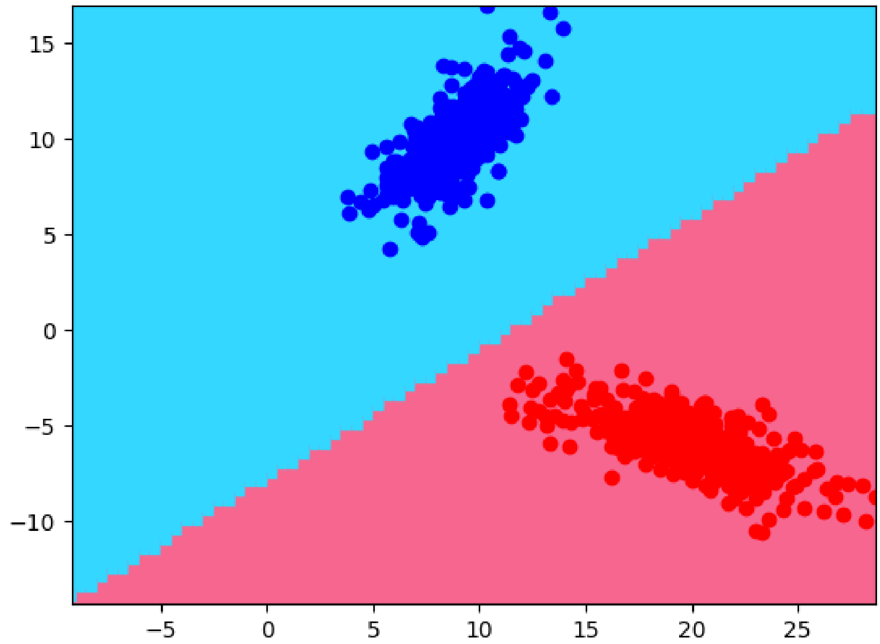

Figure 1: Plot for linearly separable data for class 1 and 2 (For same diagonal co-variance matrix)

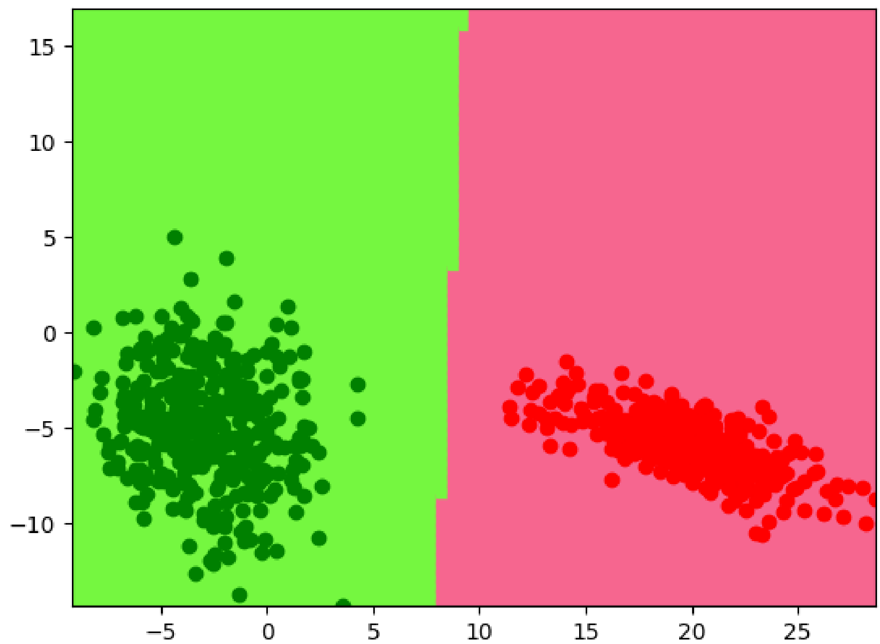

Figure 2: Plot for linearly separable data for class 1 and 3 (For same diagonal co-variance matrix)

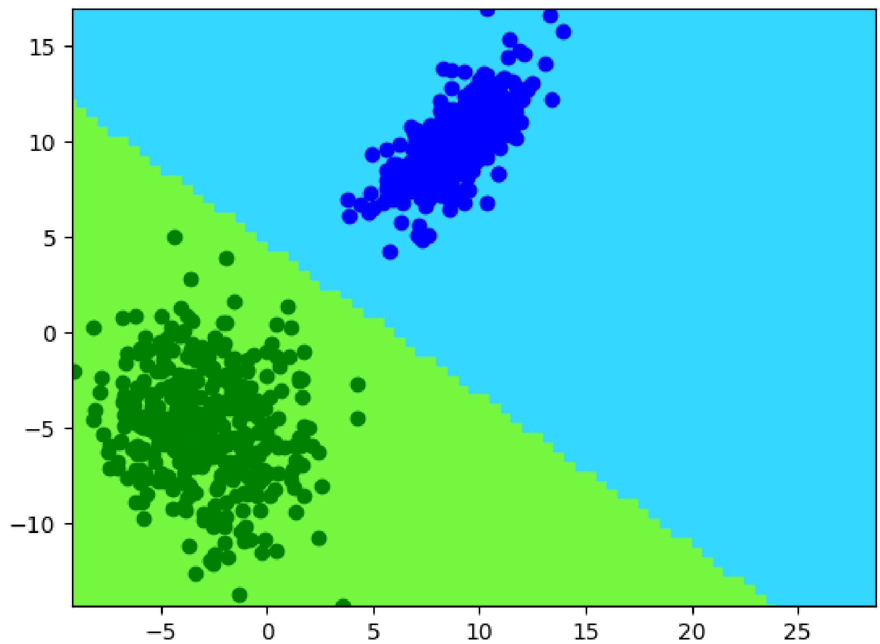

Figure 3: Plot for linearly separable data for class 2 and 3 (For same diagonal co-variance matrix)

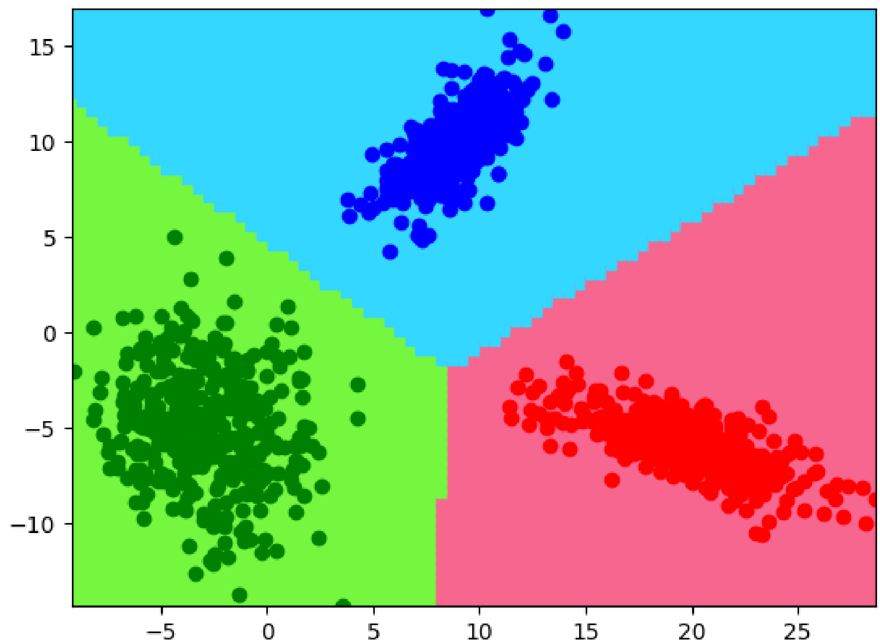

Figure 4: Plot for linearly separable data for all 3 classes (For same diagonal co-variance matrix)

5.1.1 Confusion Matrix based on performance for test data:

$$\begin{bmatrix} 125 & 0 & 0 \\ 0 & 125 & 0 \\ 0 & 0 & 125 \end{bmatrix}$$

5.1.2 Classification accuracy on test data

	Class1	Class2	Class3
Precision	1.00	1.00	1.00
Recall	1.00	1.00	1.00
F-measure	1.00	1.00	1.00

- Overall Accuracy - 100%
- Mean Precision - 1.00

- Mean Recall - 1.00
- Mean F-Measure - 1.00

5.1.3 Interpretation

1. Decision boundary is the plot comes out be linear which can be verified by Bayer's equation for this case which is also linear.
2. That's why it gives high accuracy for linearly separable(above) data.

5.2 Case 2: $\Sigma_i = \Sigma$

Here we calculated the three co-variance matrices for each case then took their average as co-variance matrix for all the classes i.e. Σ is same.

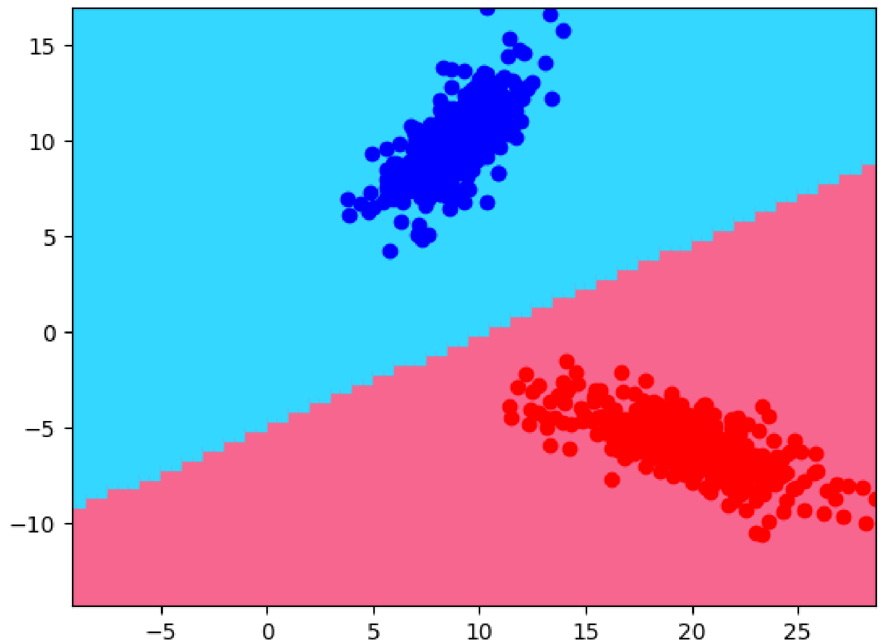

Figure 5: Plot for linearly separable data for class 1 and 2(For same full co-variance matrix)

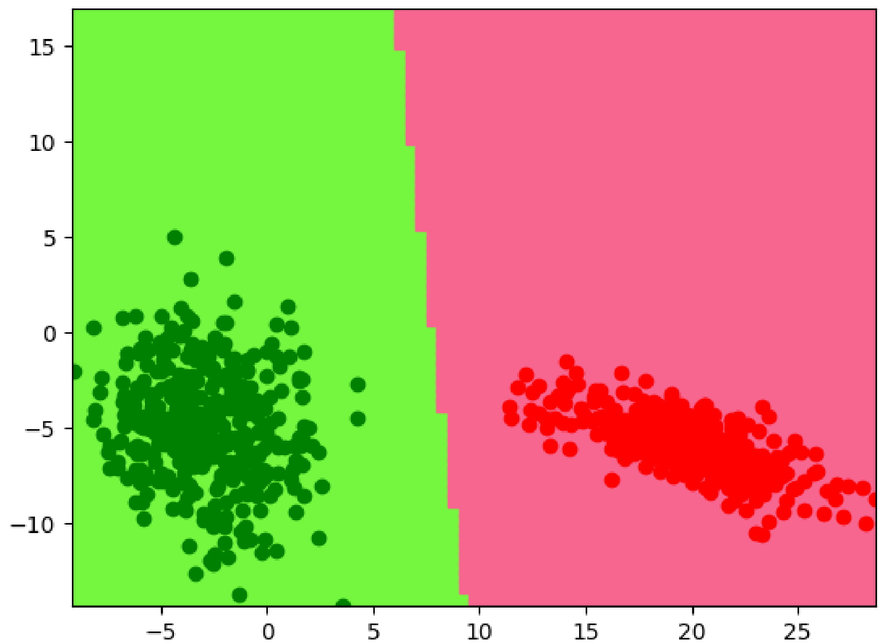

Figure 6: Plot for linearly separable data for class 1 and 3 (For same full co-variance matrix)

5.2.1 Confusion Matrix based on performance for test data:

$$\begin{bmatrix} 125 & 0 & 0 \\ 0 & 125 & 0 \\ 0 & 0 & 125 \end{bmatrix}$$

5.2.2 Classification accuracy on test data

	Class1	Class2	Class3
Precision	1.00	1.00	1.00
Recall	1.00	1.00	1.00
F-measure	1.00	1.00	1.00

- Overall Accuracy - 100%
- Mean Precision - 1
- Mean Recall - 1

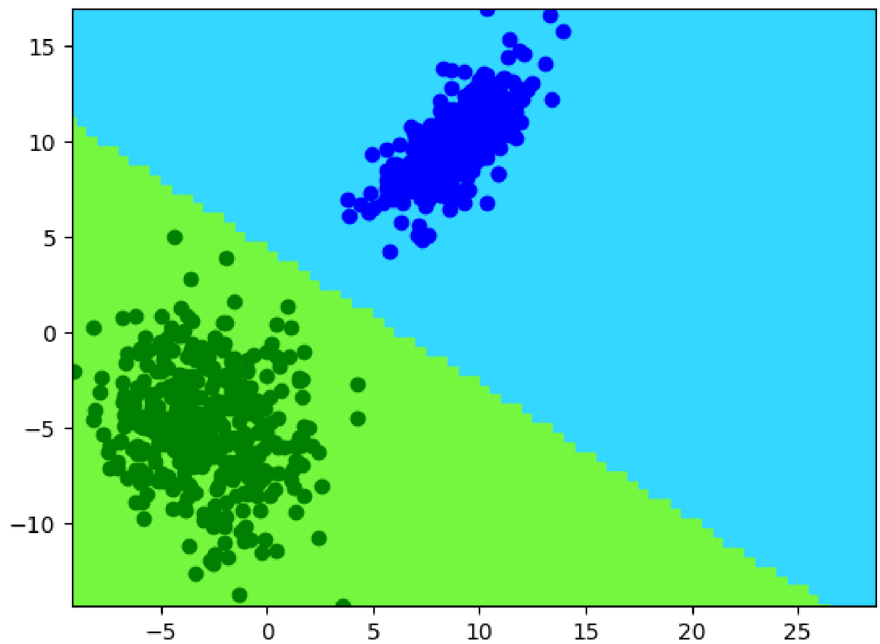

Figure 7: Plot for linearly separable data for class 2 and 3 (For same full co-variance matrix)

- Mean F-Measure - 1

5.2.3 Interpretation

1. Here we have taken all matrices to be equal so the nature of decision boundary is linear.
2. But data is still linearly separable so this gives high accuracy.

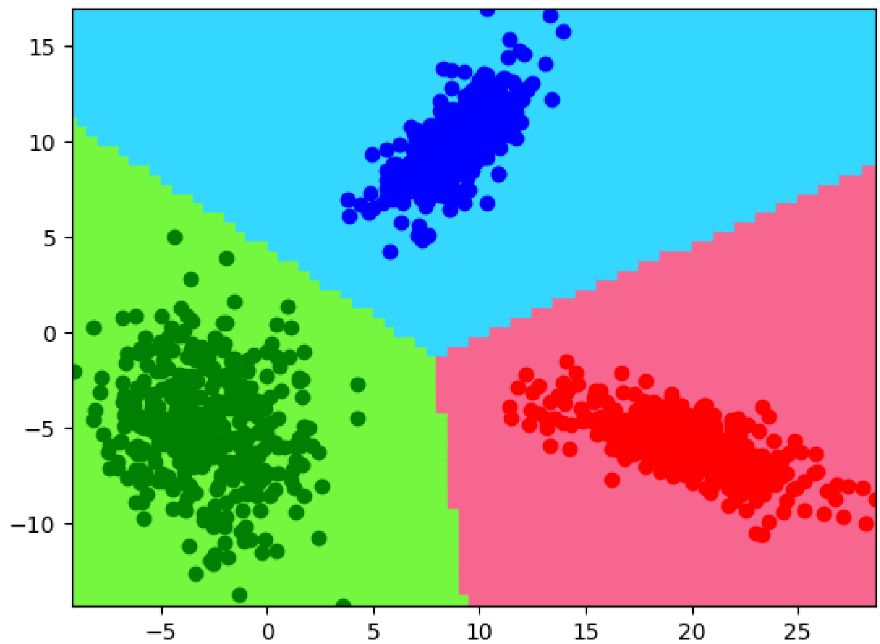

Figure 8: Plot for linearly separable data for all 3 classes(For same full co-variance matrix)

5.3 Case 3: Σ_i =Forced diagonal matrix

Here we calculated the three co-variance matrices for each case then forced them to diagonal co-variance matrix as co-variance matrix for the classes respectively i.e. is different

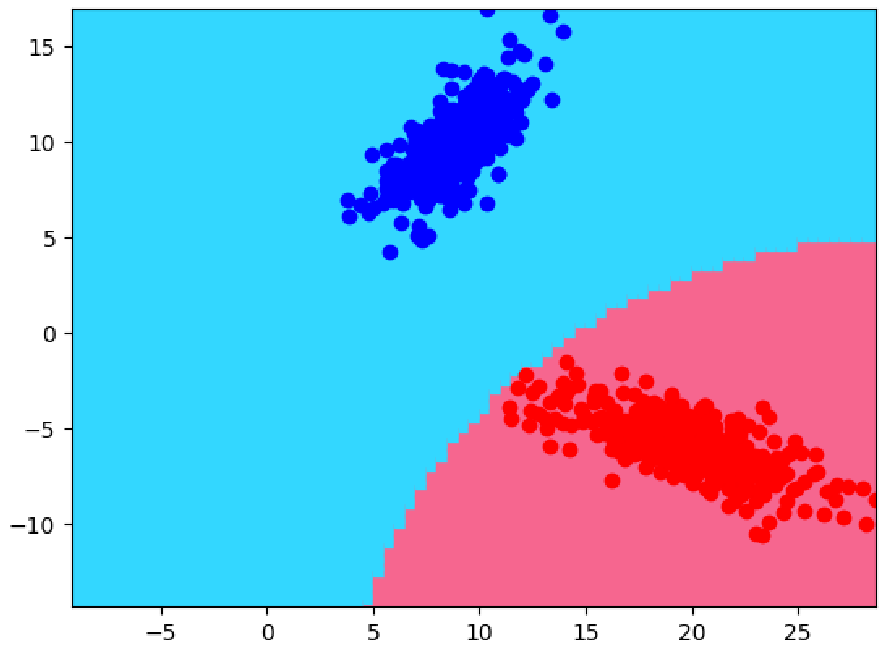

Figure 9: Plot for linearly separable data for class 1 and 2 (For different forced diagonal co-variance matrix)

5.3.1 Confusion Matrix based on performance for test data:

$$\begin{bmatrix} 125 & 0 & 0 \\ 0 & 125 & 0 \\ 0 & 0 & 125 \end{bmatrix}$$

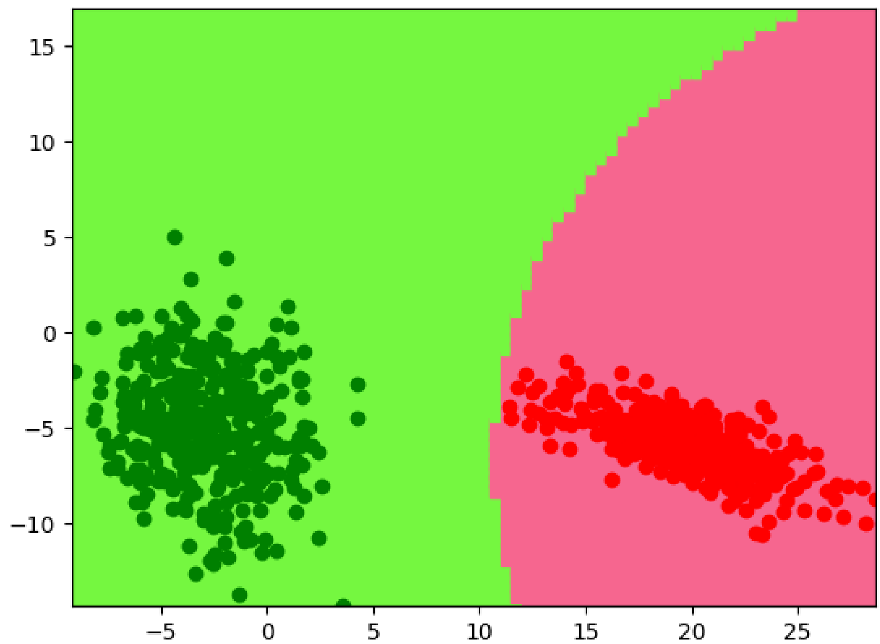

Figure 10: Plot for linearly separable data for class 1 and 3 (For different forced diagonal co-variance matrix)

5.3.2 Classification accuracy on test data

	Class1	Class2	Class3
Precision	1.00	1.00	1.00
Recall	1.00	1.00	1.00
F-measure	1.00	1.00	1.00

- Overall Accuracy - 100%
- Mean Precision - 1
- Mean Recall - 1
- Mean F-Measure - 1

5.3.3 Interpretation

In this case, off-diagonal elements are forced to zero so we get the result somewhat similar to Case 1, with boundaries having a linear form and but

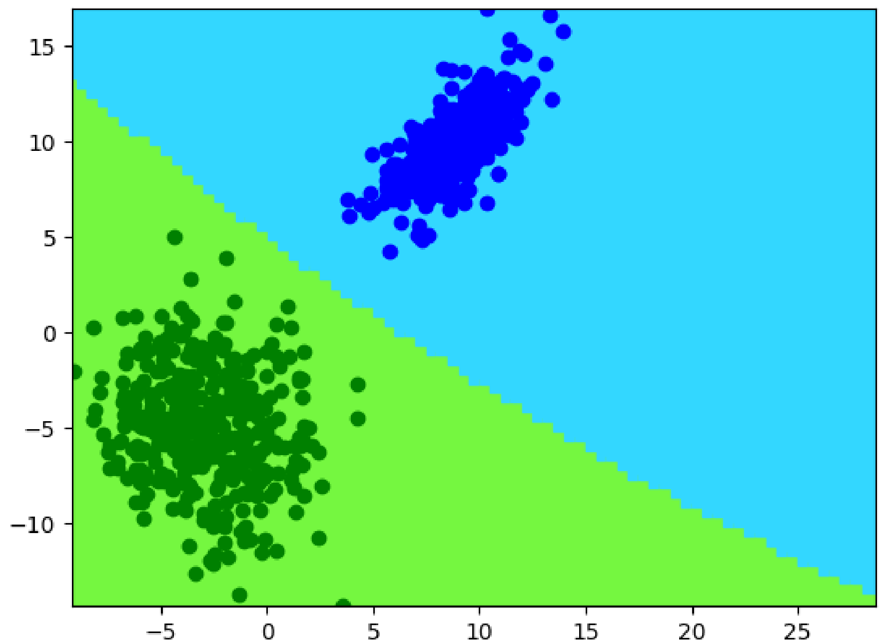

Figure 11: Plot for linearly separable data for class 2 and 3 (For different forced diagonal co-variance matrix)

there is a slight change in boundaries of classes which is due to diagonal entries of every class is not same which is in case 1.

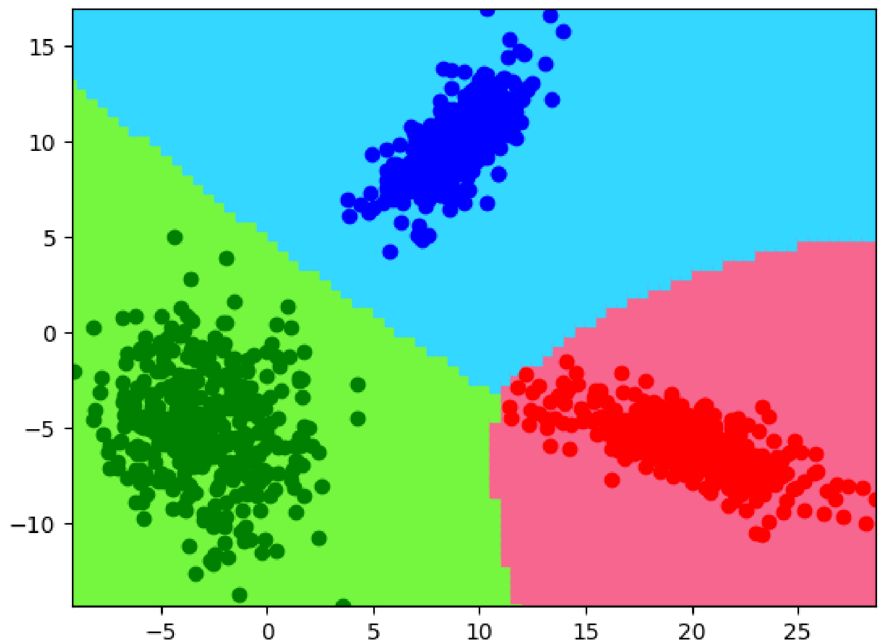

Figure 12: Plot for linearly separable data for all three classes(For different forced diagonal co-variance matrix)

5.4 Case 4: Σ_i = arbitrary

Here we calculated the three co-variance matrices for each class i.e. Σ is different.

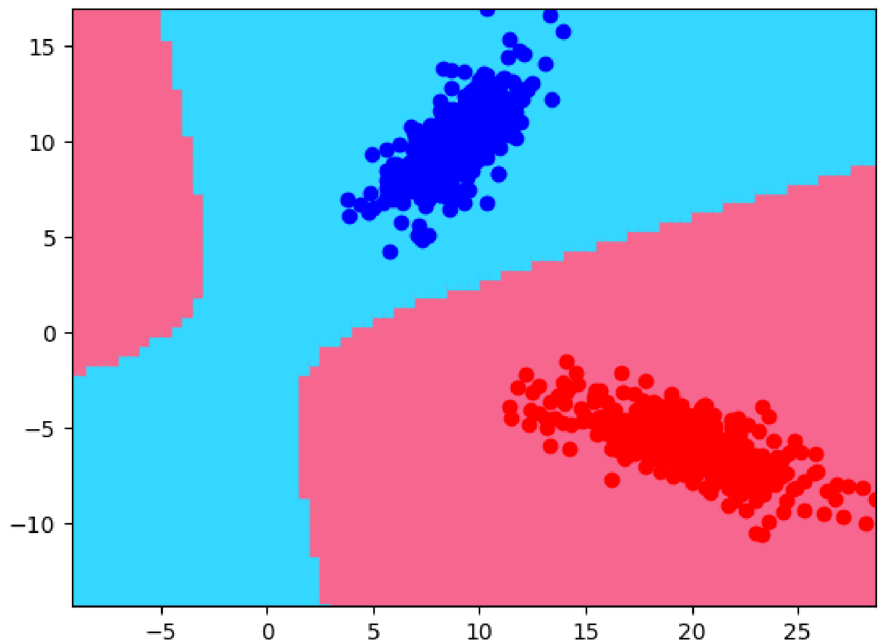

Figure 13: Plot for linearly separable data for class 1 and 2(For different full co-variance matrix)

5.4.1 Confusion Matrix based on performance for test data:

$$\begin{bmatrix} 125 & 0 & 0 \\ 0 & 125 & 0 \\ 0 & 0 & 125 \end{bmatrix}$$

5.4.2 Classification accuracy on test data

	Class1	Class2	Class3
Precision	1.00	1.00	1.00
Recall	1.00	1.00	1.00
F-measure	1.00	1.00	1.00

- Overall Accuracy - 100%
- Mean Precision - 1
- Mean Recall - 1

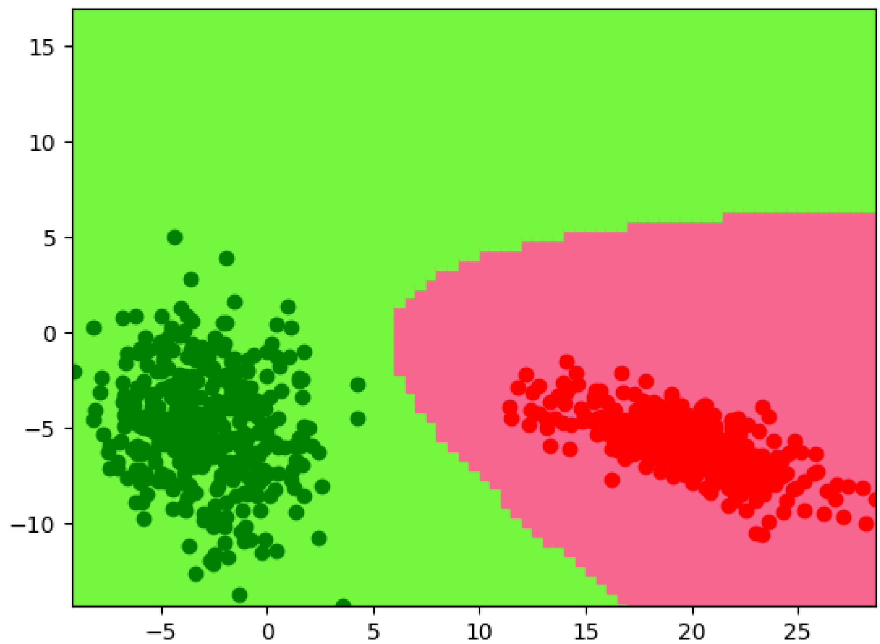

Figure 14: Plot for linearly separable data for class 1 and 3(For different full co-variance matrix)

- Mean F-Measure - 1

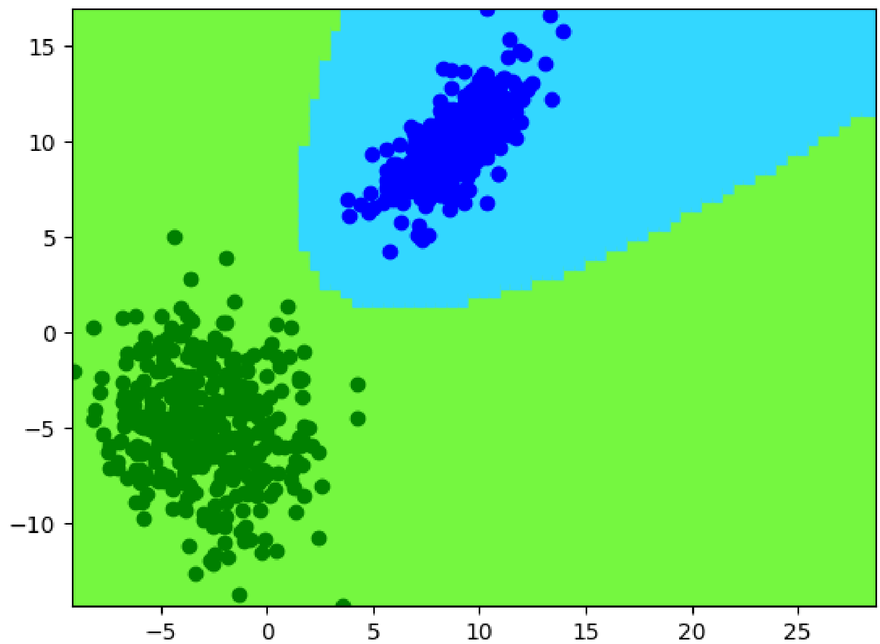

Figure 15: Plot for linearly separable data for class 2 and 3(For different full co-variance matrix)

5.4.3 Interpretation

Here the full co-variance matrices being different for each class we get decision boundary in a non-linear (quadratic) form, which can be seen through the equations. Data still being linearly separable giving high accuracy and good confusion parameters.

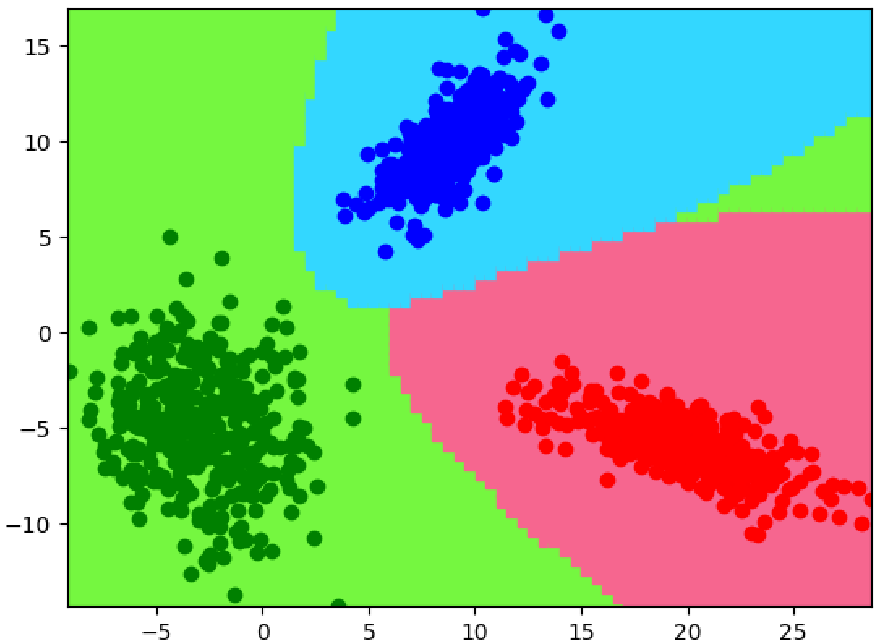

Figure 16: Plot for linearly separable data for all classes(For different full co-variance matrix)

6 Plots : Non-Linear Separable

6.1 Case 1: $\Sigma_i = \sigma^2 I$

Here we calculated the three co-variance matrices for each case then took their average as co-variance matrix for all the classes and made it diagonal.

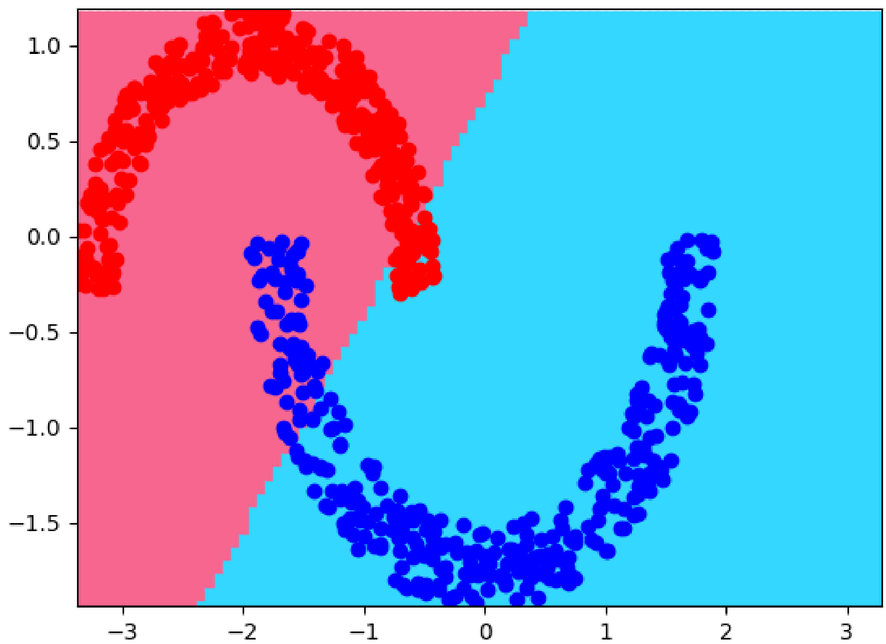

Figure 17: Plot for non-linearly separable data for class 1 and 2 (For same diagonal co-variance matrix)

6.1.1 Confusion Matrix based on performance for test data:

$$\begin{bmatrix} 113 & 12 & 0 \\ 26 & 77 & 22 \\ 0 & 10 & 115 \end{bmatrix}$$

6.1.2 Classification accuracy on test data

	Class 1	Class 2	Class 3
Precision	0.812	0.777	0.839
Recall	0.904	0.616	0.92
F-measure	0.856	0.6875	0.877

- Overall Accuracy - 81.33%
- Mean Precision - 0.8093
- Mean Recall - 0.8133

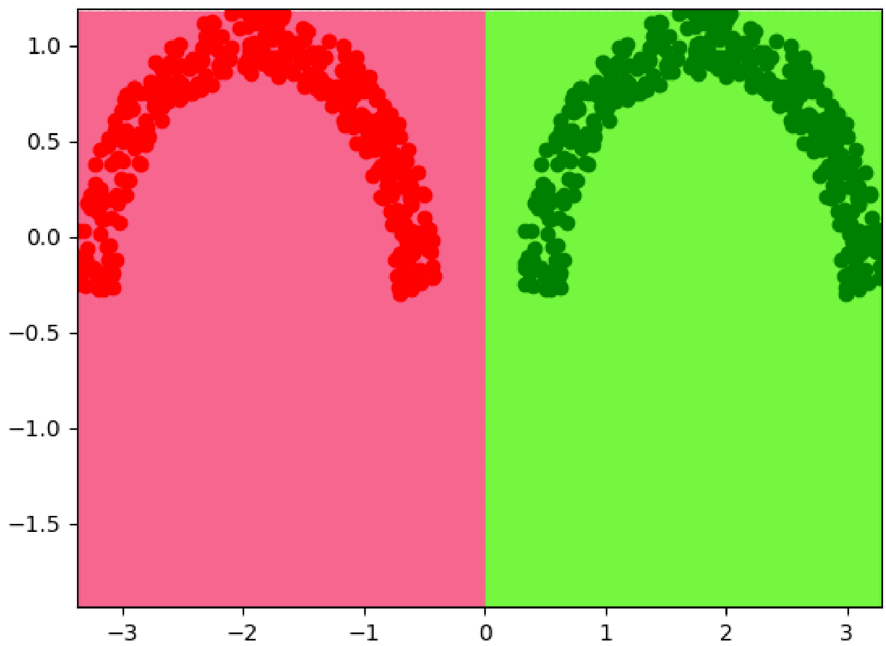

Figure 18: Plot for non-linearly separable data for class 1 and 3(For same diagonal co-variance matrix)

- Mean F-Measure - 0.8068

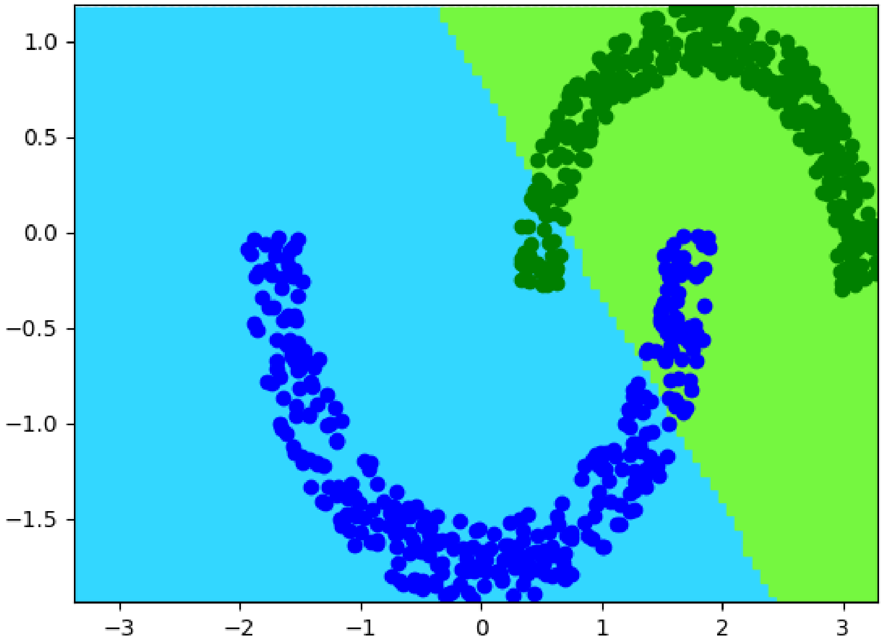

Figure 19: Plot for non-linearly separable data for class 2 and 3(For same diagonal co-variance matrix)

6.1.3 Interpretation

This time Data given is highly non-linear, hence the overall mean for the classes lies somewhere in interior of the parabolic/hyperbolic shape of the classes. And in case 1 the decision boundary is still linear because of the condition which has been imposed. Here too prior are equal. linear boundary equation and non-linear data, hence not an efficient classification, which can also be seen through confusion parameters.

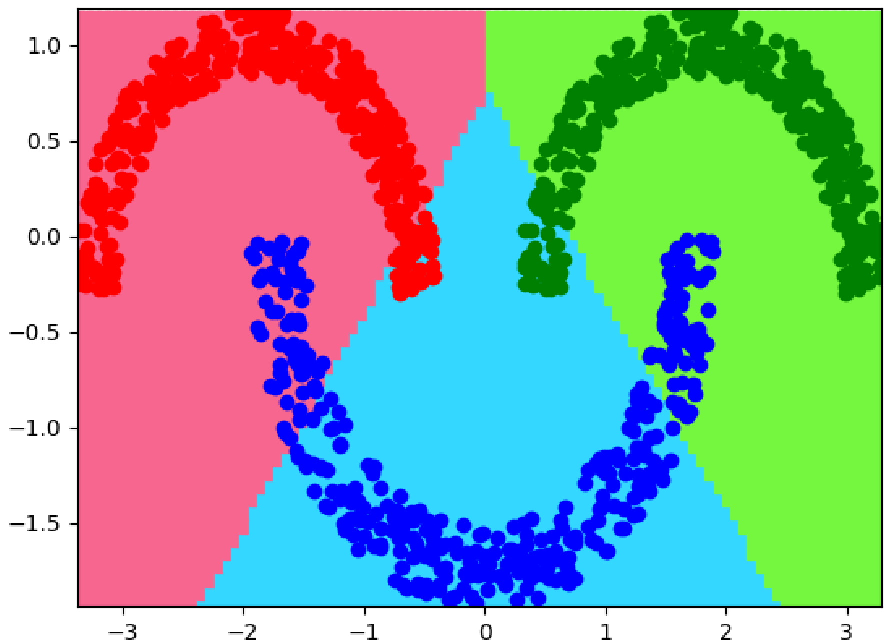

Figure 20: Plot for non-linearly separable data for all classes(For same diagonal co-variance matrix)

6.2 Case 2: $\Sigma_i = \Sigma$

Here we calculated the three co-variance matrices for each case then took their average as co-variance matrix for all the classes i.e. Σ is same.

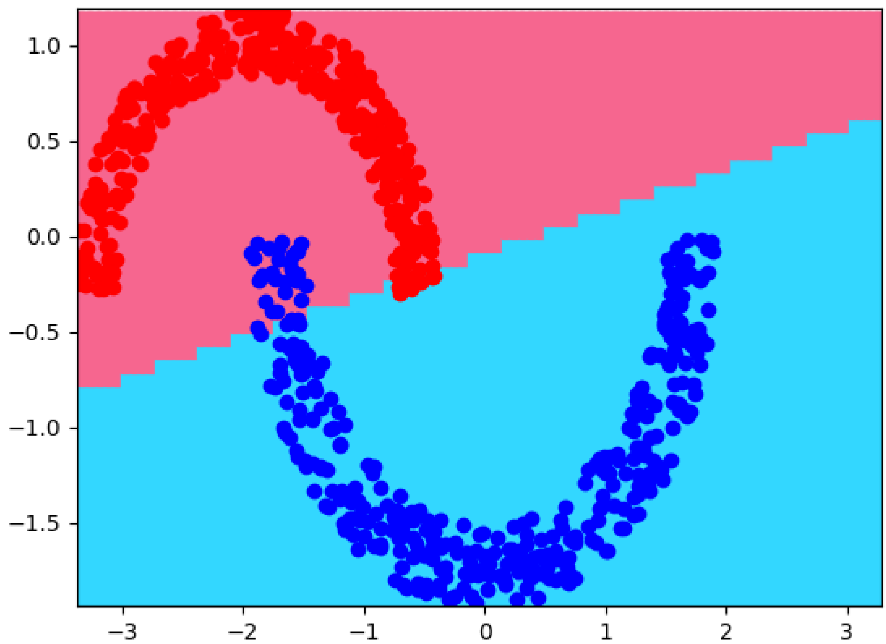

Figure 21: Plot for non-linearly separable data for class 1 and 2 (For same full co-variance matrix)

6.2.1 Confusion Matrix based on performance for test data:

$$\begin{bmatrix} 122 & 3 & 0 \\ 15 & 98 & 12 \\ 0 & 3 & 122 \end{bmatrix}$$

6.2.2 Classification accuracy on test data

	Class 1	Class 2	Class 3
Recall	0.976	0.784	0.976
Precision	0.89	0.942	0.91
F-measure	0.931	0.855	0.942

- Overall Accuracy - 91.2%
- Mean Precision - 0.914
- Mean Recall - 0.912

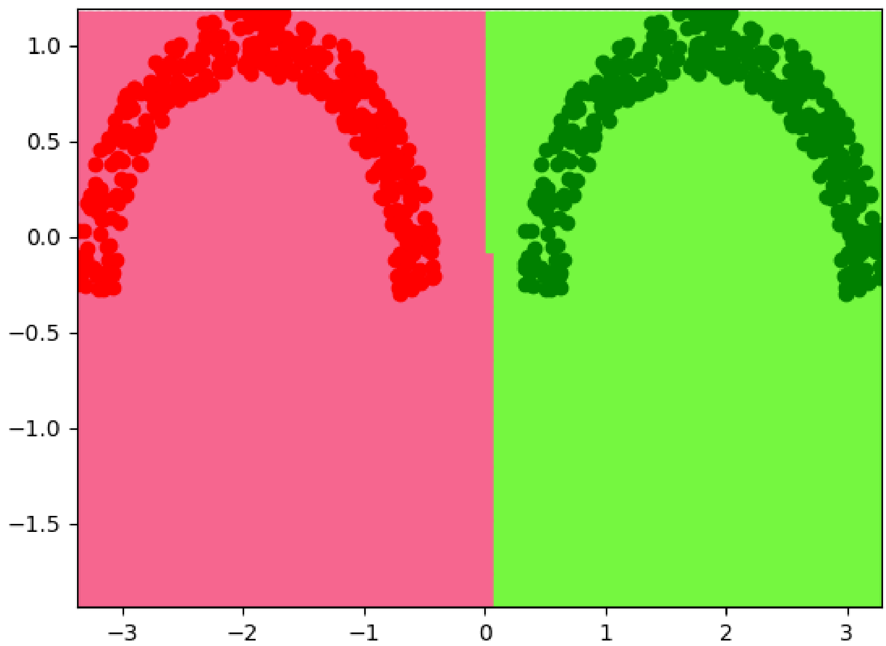

Figure 22: Plot for non-linearly separable data for class 1 and 3 (For same full co-variance matrix)

- Mean F-Measure - 0.9093

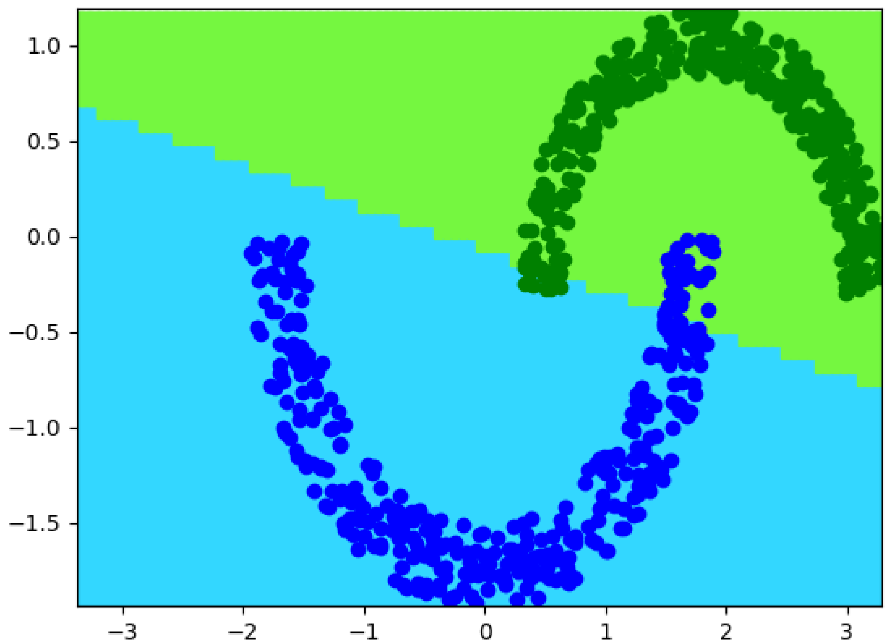

Figure 23: Plot for non-linearly separable data for class 2 and 3 (For same full co-variance matrix)

6.2.3 Interpretation

This time also decision boundary comes out to be linear,so we can infer that the decision boundaries and confusion parameters are not good to solve this classification problem. One of the reasons for that is because of our first assumption of assuming the entire data set as a Gaussian distribution. And the given spiral distribution nowhere resembles a Gaussian distribution hence the likelihoods calculated (based on Gaussian distribution) which gave posteriors predicted wrong class labels many times.

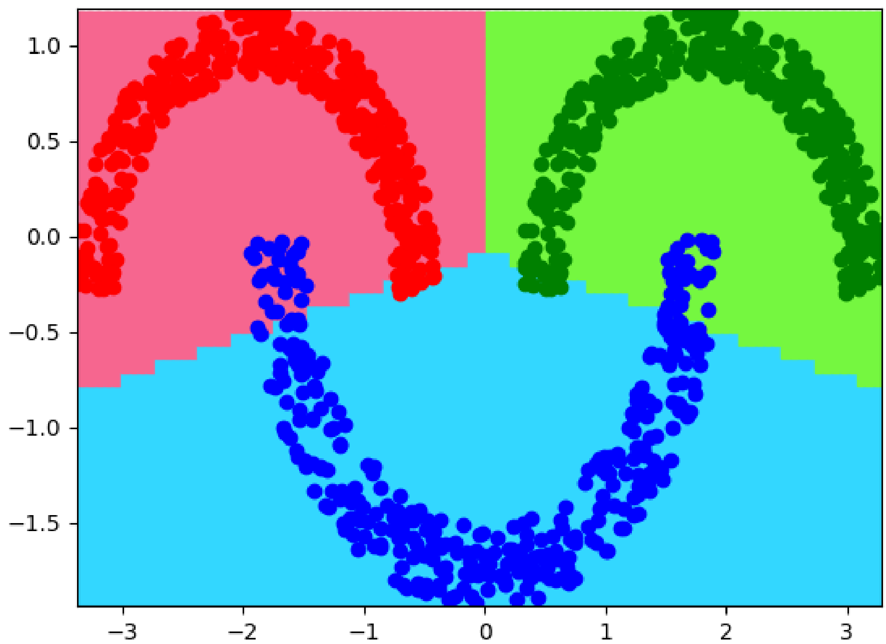

Figure 24: Plot for non-linearly separable data for all classes(For same full co-variance matrix)

6.3 Case 3: Σ_i =Forced diagonal matrix

Here we calculated the three co-variance matrices for each case then forced them to diagonal co-variance matrix as co-variance matrix for the classes respectively i.e. is different

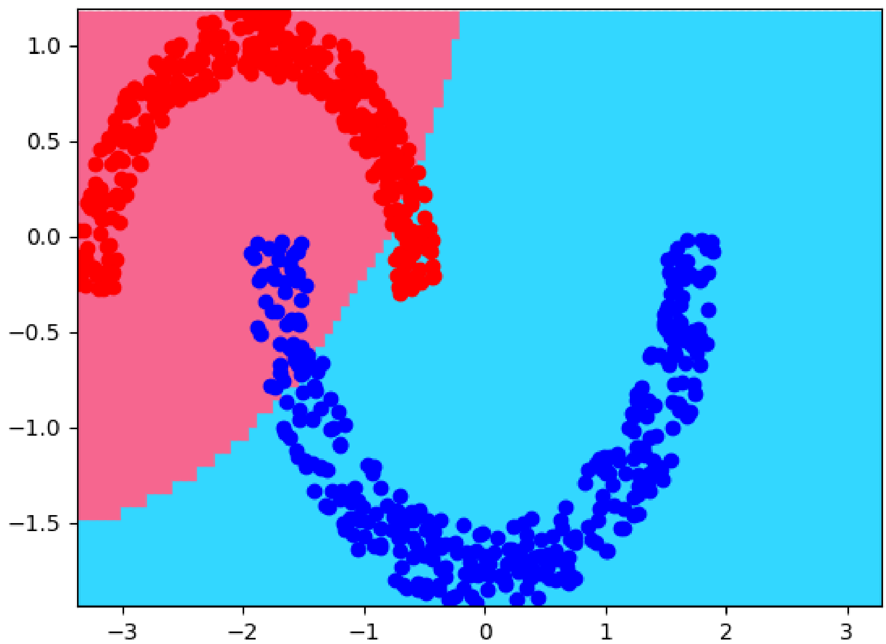

Figure 25: Plot for non-linearly separable data for class 1 and 2 (For different forced diagonal co-variance matrix)

6.3.1 Confusion Matrix based on performance for test data:

$$\begin{bmatrix} 110 & 15 & 0 \\ 21 & 84 & 20 \\ 0 & 15 & 110 \end{bmatrix}$$

6.3.2 Classification accuracy on test data

	Class 1	Class 2	Class 3
Recall	0.88	0.672	0.88
Precision	0.839	0.736	0.846
F-measure	0.859	0.7025	0.862

- Overall Accuracy - 81.06%
- Mean Precision - 0.807
- Mean Recall - 0.8106

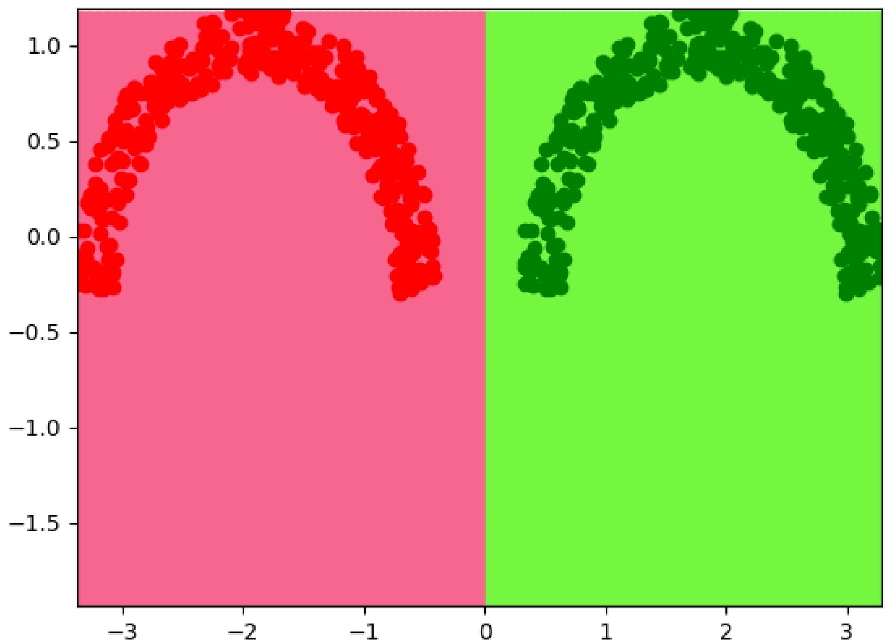

Figure 26: Plot for non-linearly separable data for class 1 and 3 (For different forced diagonal co-variance matrix)

- Mean F-Measure - 0.8078

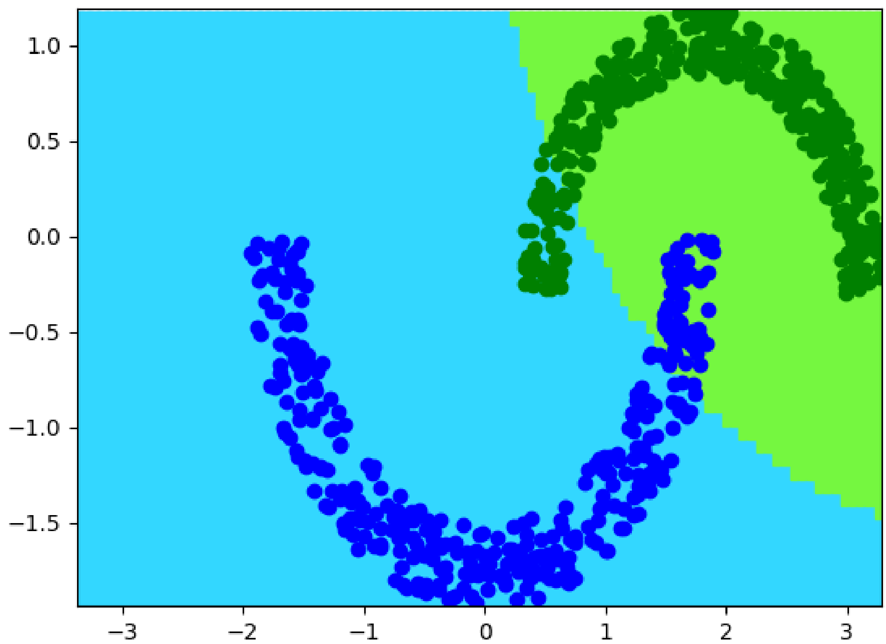

Figure 27: Plot for non-linearly separable data for class 2 and 3 (For different forced diagonal co-variance matrix)

6.3.3 Interpretation

Here too we can come to same conclusion as discussed in case1 and case2. Here we can at least infer why we get such a surface because the (nature of) co-variance matrices of each of these spiral classes is similar on top of it we are imposing a Gaussian distribution over its mean vectors which results in a linear boundary.

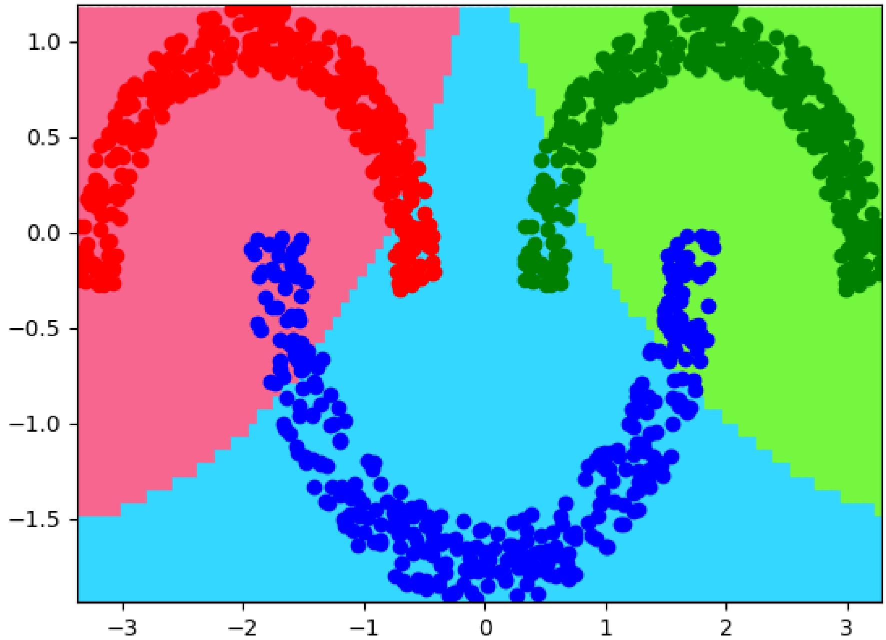

Figure 28: Plot for non-linearly separable data for all classes(For different forced diagonal co-variance matrix)

6.4 Case 4: Σ_i = arbitrary

Here we calculated the three co-variance matrices for each class i.e. Σ is different

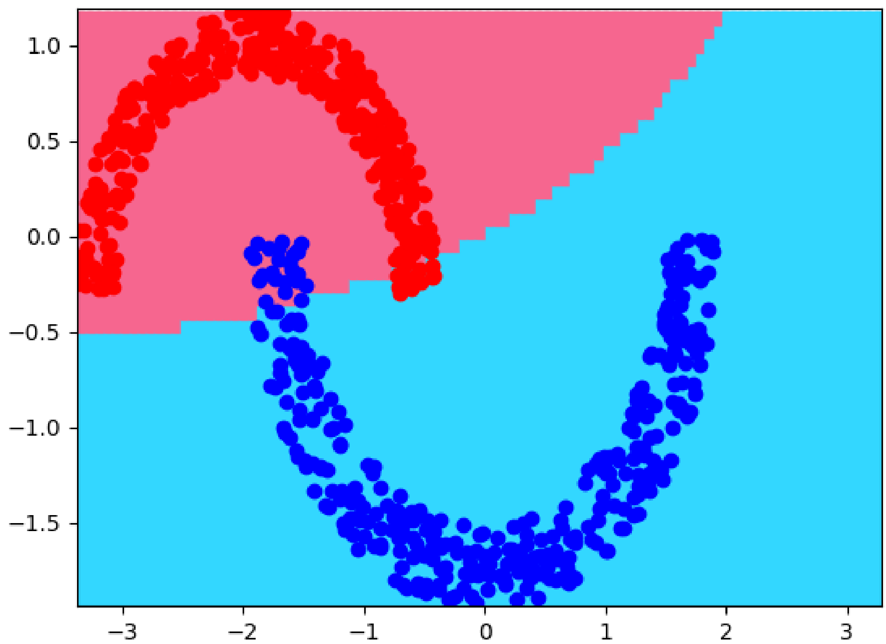

Figure 29: Plot for non-linearly separable data for class 1 and 2 (For different full co-variance matrix)

6.4.1 Confusion Matrix based on performance for test data:

$$\begin{bmatrix} 117 & 8 & 0 \\ 12 & 107 & 6 \\ 0 & 4 & 121 \end{bmatrix}$$

6.4.2 Classification accuracy on test data

	Class 1	Class 2	Class 3
Recall	0.936	0.856	0.968
Precision	0.907	0.899	0.952
F-measure	0.921	0.877	0.96

- Overall Accuracy - 92%
- Mean Precision - 0.9193
- Mean Recall - 0.92

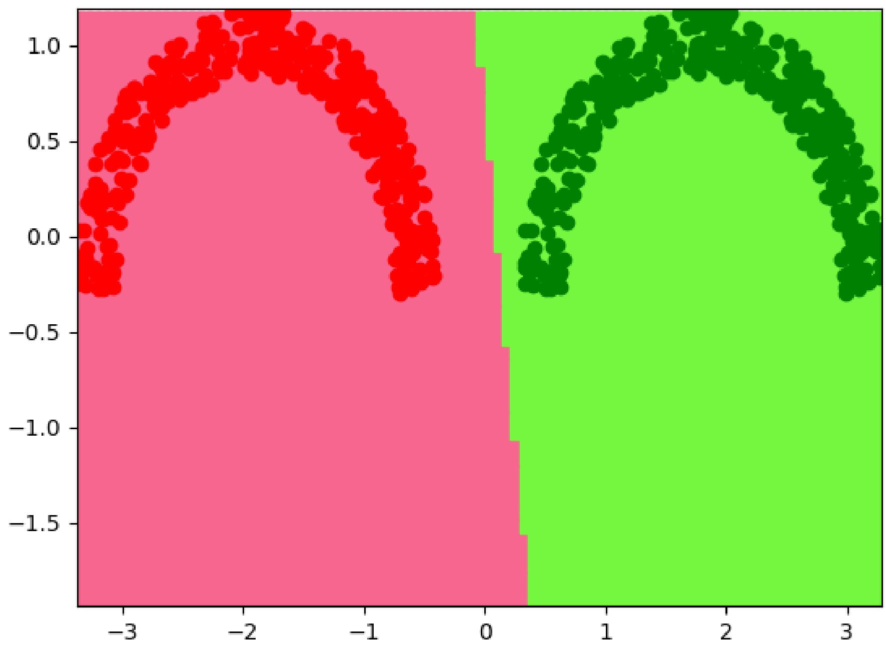

Figure 30: Plot for non-linearly separable data for class 1 and 3 (For different full co-variance matrix)

- Mean F-Measure - 0.9193

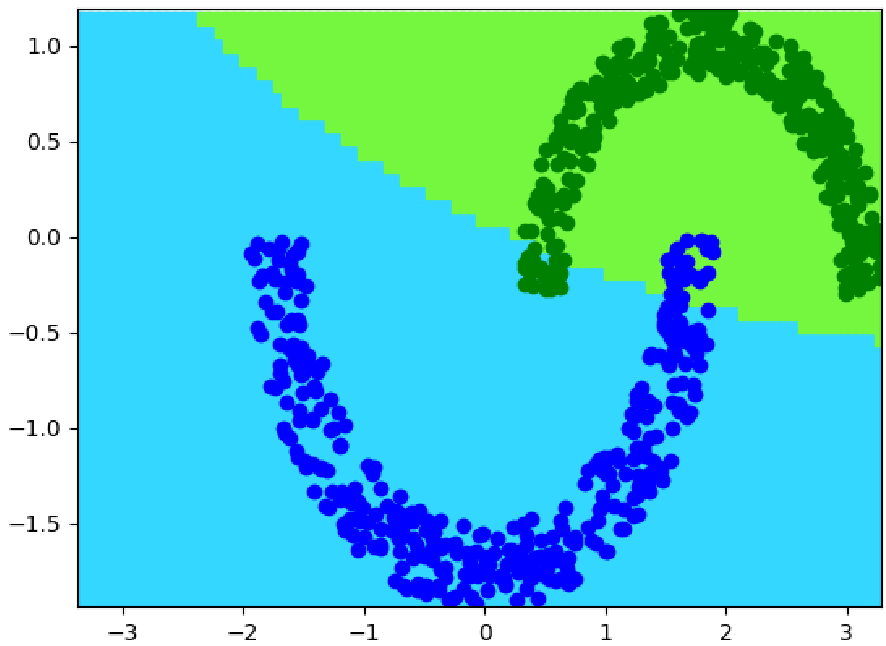

Figure 31: Plot for non-linearly separable data for class 2 and 3 (For different full co-variance matrix)

6.4.3 Interpretation

Here too we can extend above interpretations with some changes that can be noted as at the central part some non-linearity is observed of the decision surface where class2 is more pulled down at center where its more clustered compared to class1. This is because of the co-variance matrices which we have assumed different. Hence we can infer that complete co-variance matrices different for every class in a non-linear distribution might be able to give a close and a good solution to classification problem. Another thing which we can suggest here is that we can divide such class in many Gaussian surfaces and then apply this formula which may help us in refining our prediction about a classs label. But from the values present in the confusion matrix still this cant be said empirically here.

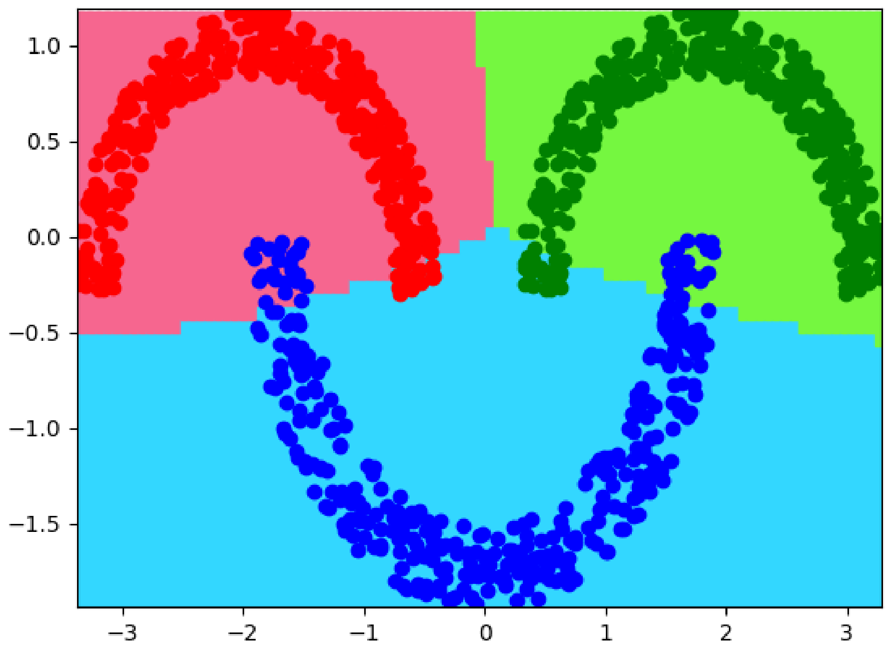

Figure 32: Plot for non-linearly separable data for all classes (For different full co-variance matrix)

7 Plots : Real

7.1 Case 1: $\Sigma_i = \sigma^2 I$

Here we calculated the three co-variance matrices for each case then took their average as co-variance matrix for all the classes and made it diagonal.

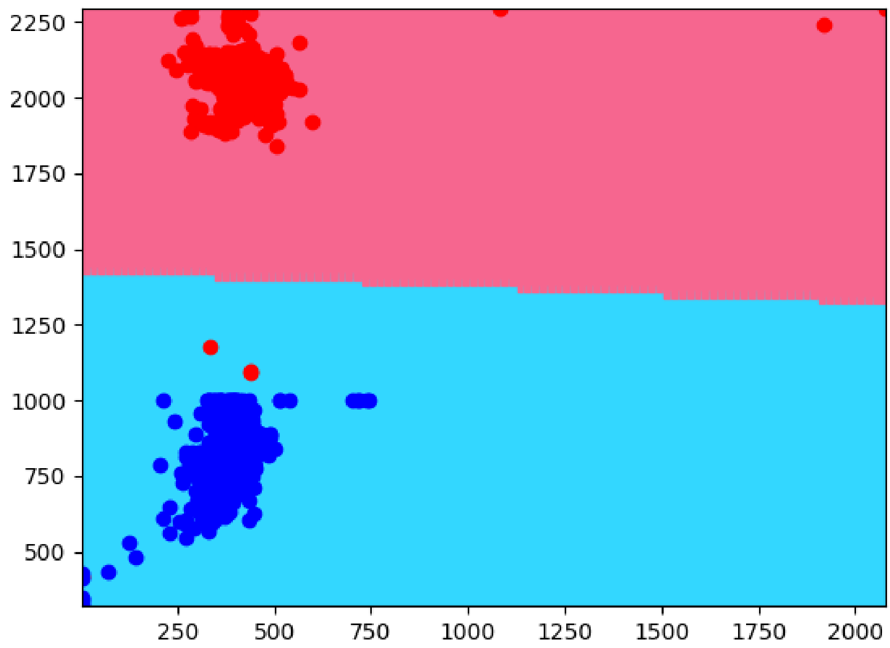

Figure 33: Plot for real data for class 1 and 2(For same diagonal co-variance matrix)

7.1.1 Confusion Matrix based on performance for test data:

$$\begin{bmatrix} 554 & 13 & 5 \\ 0 & 572 & 50 \\ 0 & 178 & 435 \end{bmatrix}$$

7.1.2 Classification accuracy on test data

	Class1	Class2	Class3
Recall	0.968	0.919	0.709
Precision	1.0	0.749	0.8877
F-measure	0.984	0.825	0.7887

- Overall Accuracy - 86.38%
- Mean Precision - 0.8789
- Mean Recall - 0.8638

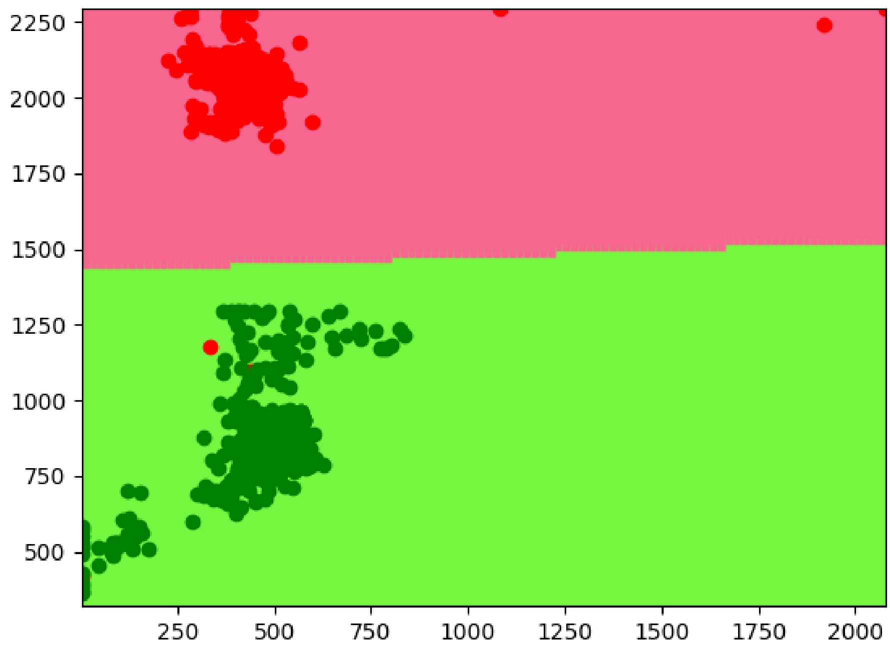

Figure 34: Plot for real data for class 1 and 3(For same diagonal co-variance matrix)

- Mean F-Measure - 0.8659

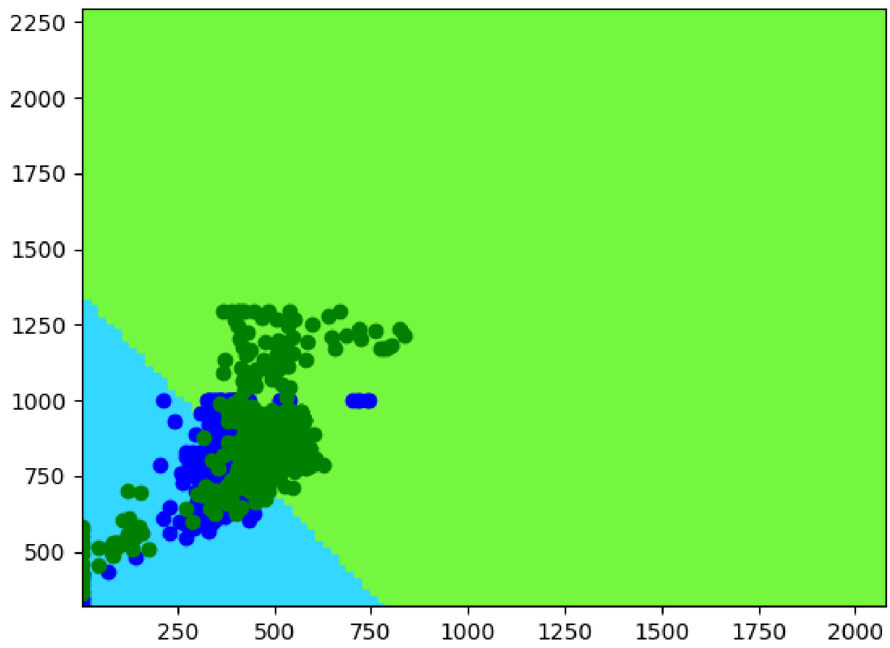

Figure 35: Plot for real data for class 2 and 3(For same diagonal co-variance matrix)

7.1.3 Interpretation

We get a linear decision surface which is consistent with the formula and conditions which we have imposed in this case. This being quite a good approximation of Gaussian distribution. The decision boundaries which we get are quite useful in classifying the data points of class-3 with class-1 and class-2. Thus, confusion matrix gives us here a balanced data with a high precision for the third class.

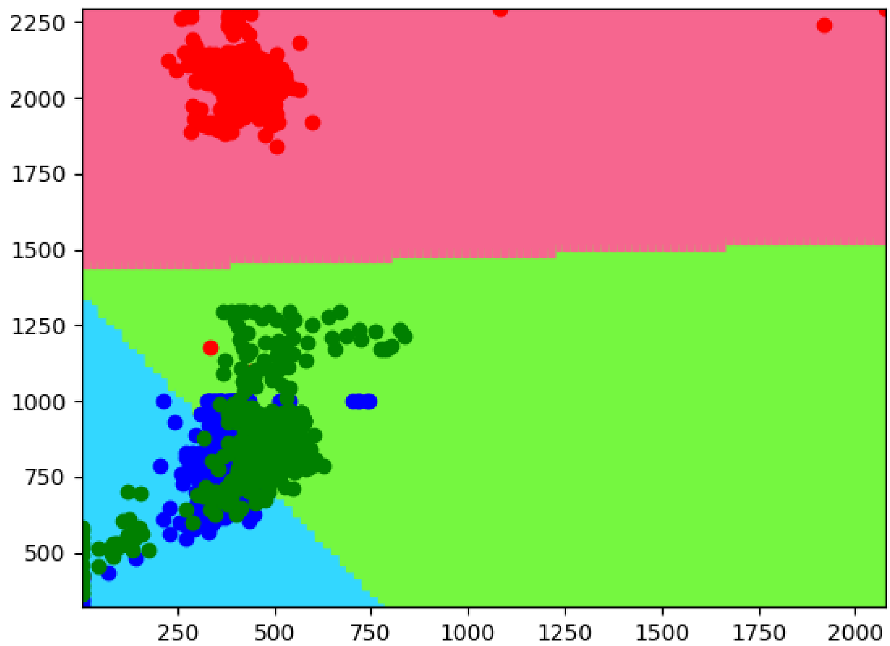

Figure 36: Plot for real data for all 3 classes(For same diagonal co-variance matrix)

7.2 Case 2: $\Sigma_i = \Sigma$

Here we calculated the three co-variance matrices for each case then took their average as co-variance matrix for all the classes i.e. Σ is same.

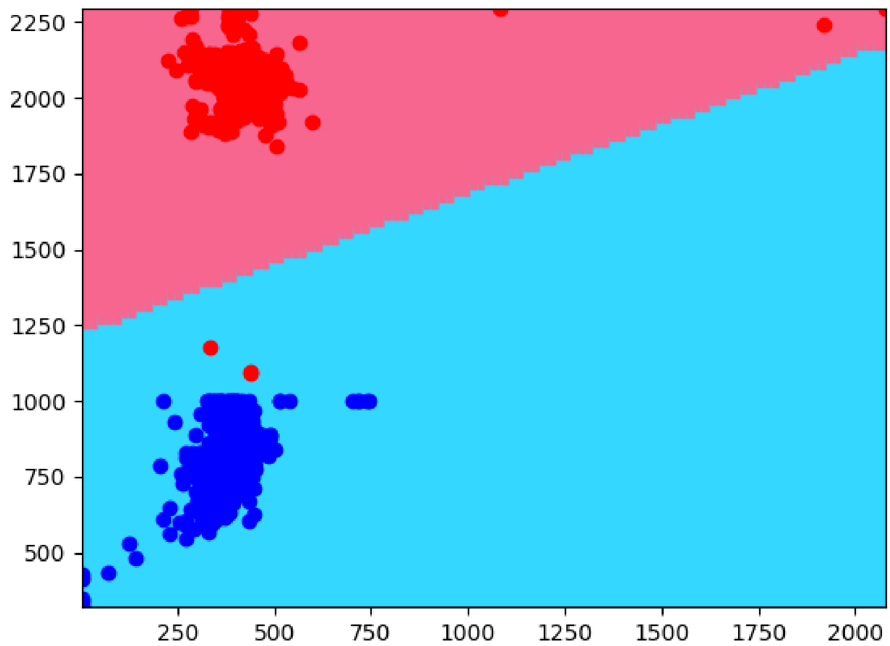

Figure 37: Plot for real data for class 1 and 2(For same full co-variance matrix)

7.2.1 Confusion Matrix based on performance for test data:

$$\begin{bmatrix} 554 & 15 & 3 \\ 0 & 593 & 29 \\ 0 & 19 & 422 \end{bmatrix}$$

7.2.2 Classification accuracy on test data

	Class1	Class2	Class3
Recall	0.968	0.953	0.6884
Precision	1	0.742	0.929
F-measure	0.984	0.834	0.791

- Overall Accuracy - 86.82%
- Mean Precision - 0.89
- Mean Recall - 0.8682

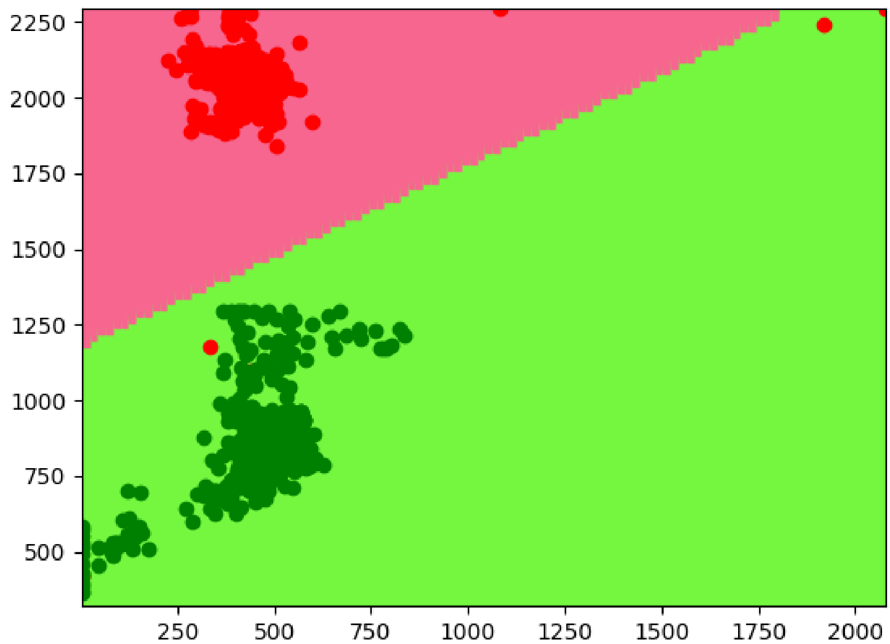

Figure 38: Plot for real data for class 1 and 3(For same full co-variance matrix)

- Mean F-Measure - 0.8696

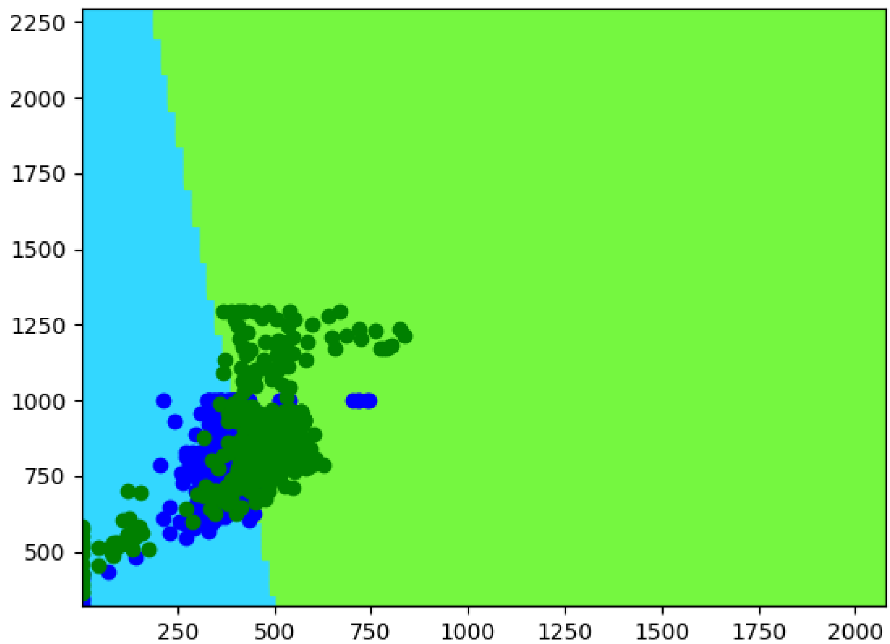

Figure 39: Plot for real data for class 2 and 3(For same full co-variance matrix)

7.2.3 Interpretation

Boundary which we get here is linear but the imposed conditions are quite not efficient to solve this classification problem. As data points are not easily linearly separable hence using same co-variance matrix in all classes is not efficient as this completely disables the clustering properties for some classes. Hence we are getting such different distribution. The confusion matrix tells us that the class 2 and class 1 doesnt have a balanced precision and recall values.

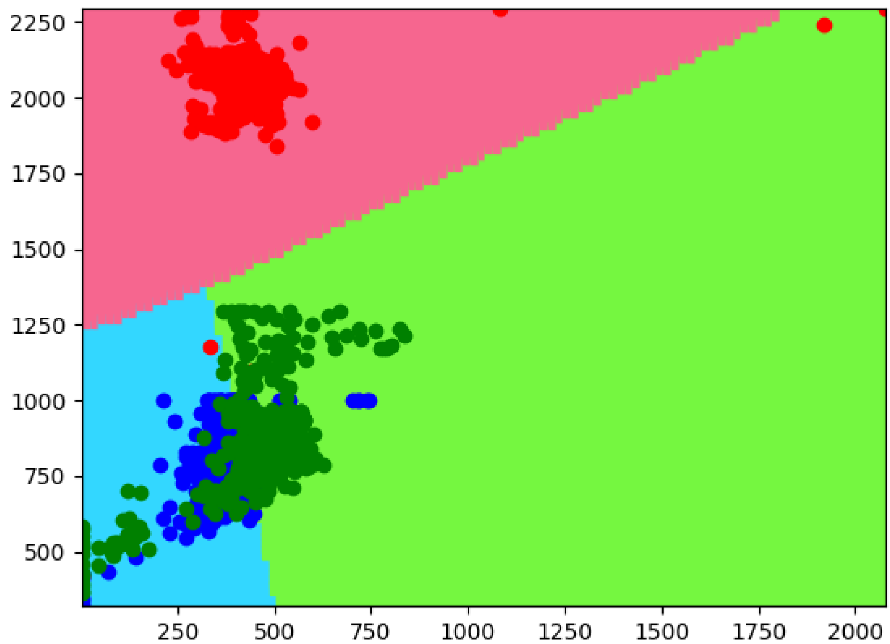

Figure 40: Plot for real data for all classes(For same full co-variance matrix)

7.3 Case 3: Σ_i =Forced diagonal matrix

Here we calculated the three co-variance matrices for each case then forced them to diagonal co-variance matrix as co-variance matrix for the classes respectively i.e. is different

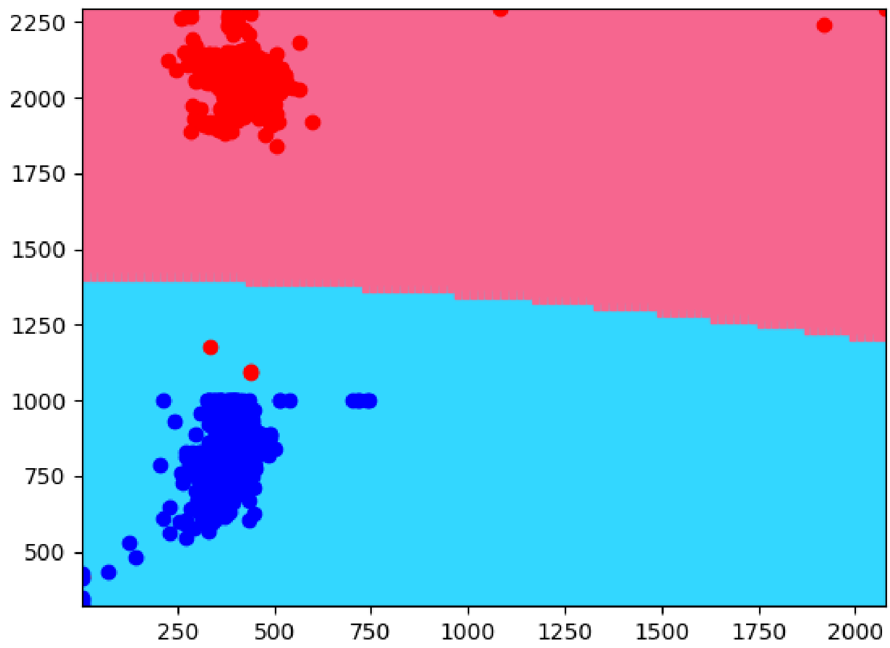

Figure 41: Plot for real data for classes 1 and 2(For different forced diagonal co-variance matrix)

7.3.1 Confusion Matrix based on performance for test data:

$$\begin{bmatrix} 548 & 7 & 17 \\ 0 & 550 & 72 \\ 0 & 204 & 409 \end{bmatrix}$$

7.3.2 Classification accuracy on test data

	Class1	Class2	Class3
Recall	0.958	0.884	0.667
Precision	1	0.722	0.8212
F-measure	0.978	0.795	0.736

- Overall Accuracy - 83.39%
- Mean Precision - 0.8479
- Mean Recall - 0.8339

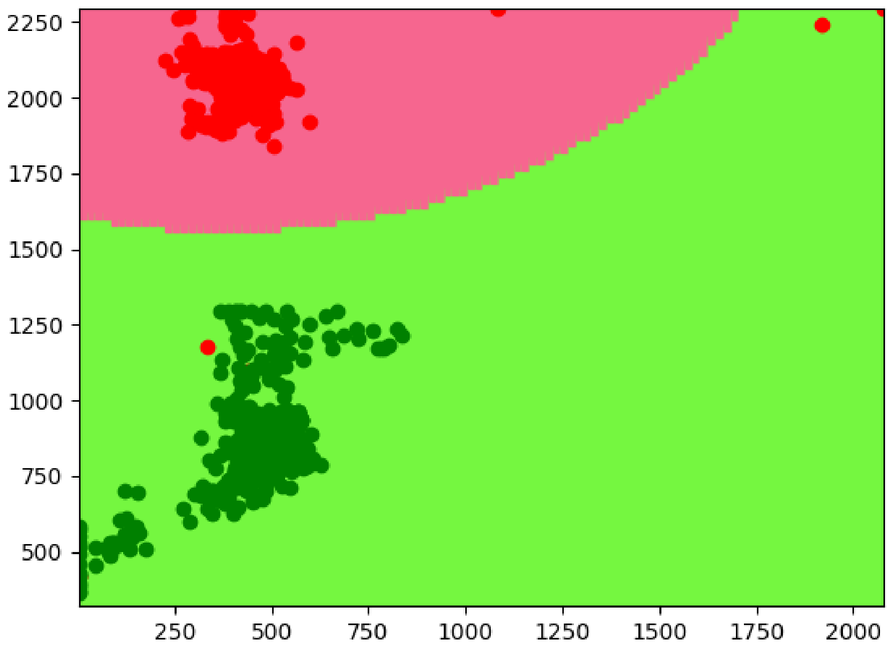

Figure 42: Plot for real data for classes 1 and 3(For different forced diagonal co-variance matrix)

- Mean F-Measure - 0.8363

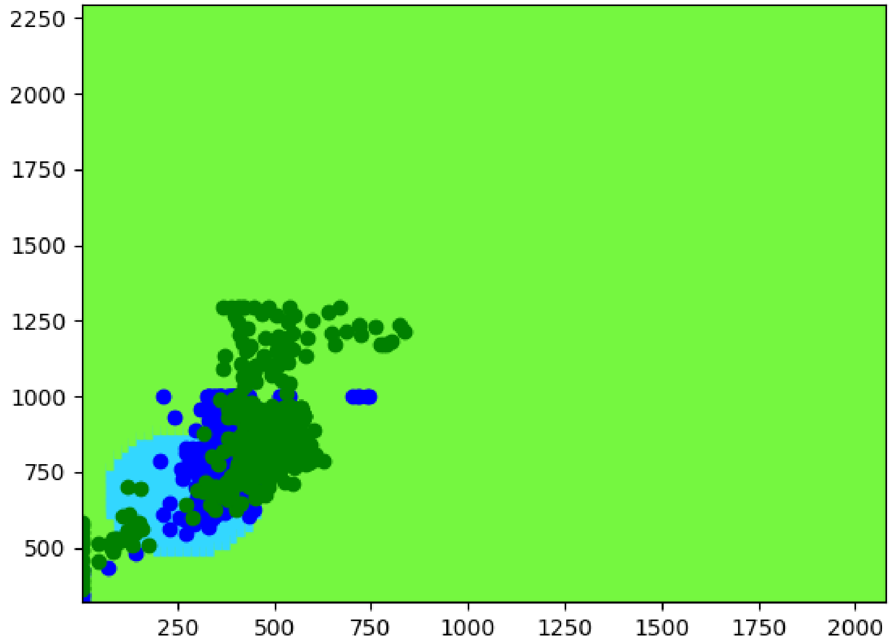

Figure 43: Plot for real data for classes 2 and 3 (For different forced diagonal co-variance matrix)

7.3.3 Interpretation

Interpretation of this case too can be seen as a continuation of above case 2 as here we have completely neglected the clustering property of some classes because its off-diagonal elements are not taken into consideration. Hence here too class labeling is not efficient given the conditions we have imposed in calculating the co-variance matrices of the classes.

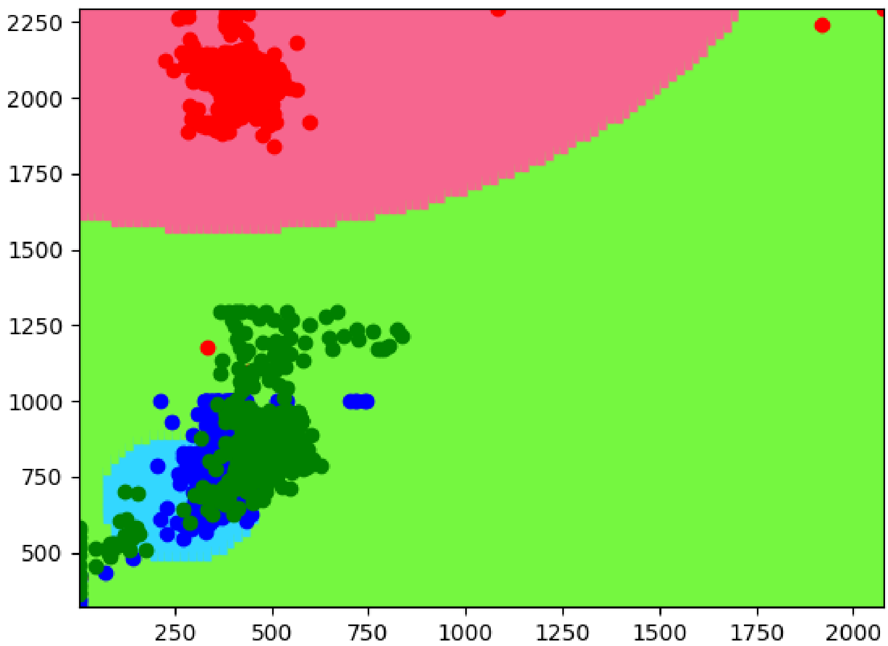

Figure 44: Plot for real data for all classes(For different forced diagonal covariance matrix)

In this when we plot contours, they will be elliptical but will be oriented to the axis, as there are diagonal entries in the matrix.

7.4 Case 4: Σ_i = arbitrary

Here we calculated the three co-variance matrices for each class i.e. Σ is different

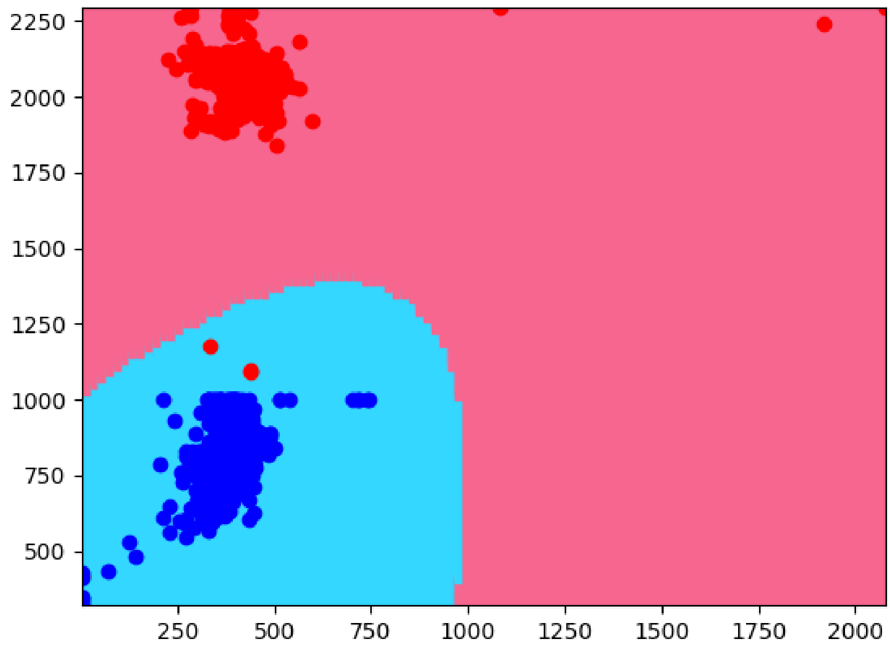

Figure 45: Plot for real data for classes 1 and 2(For different full co-variance matrix)

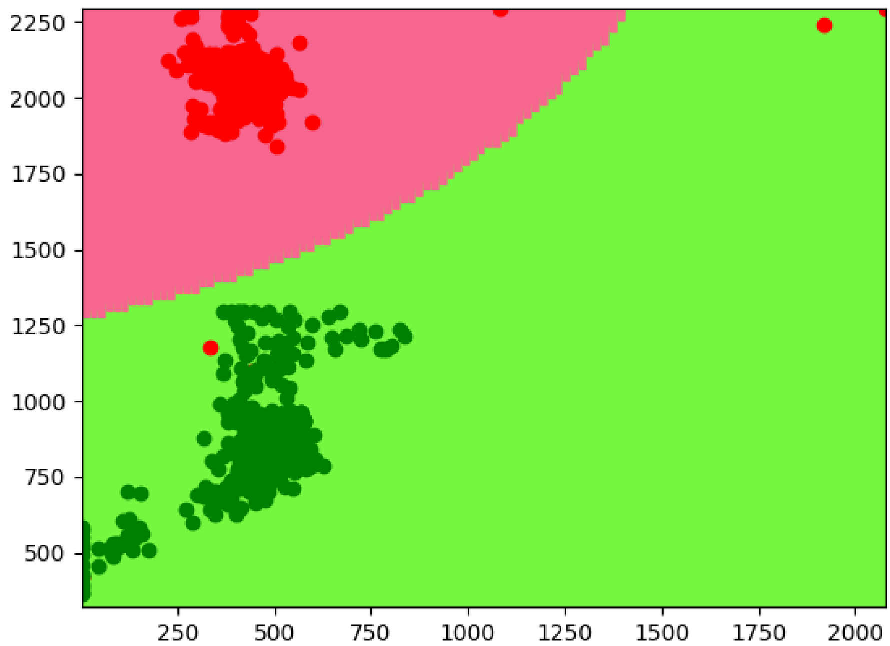

Figure 46: Plot for real data for classes 1 and 3(For different full co-variance matrix)

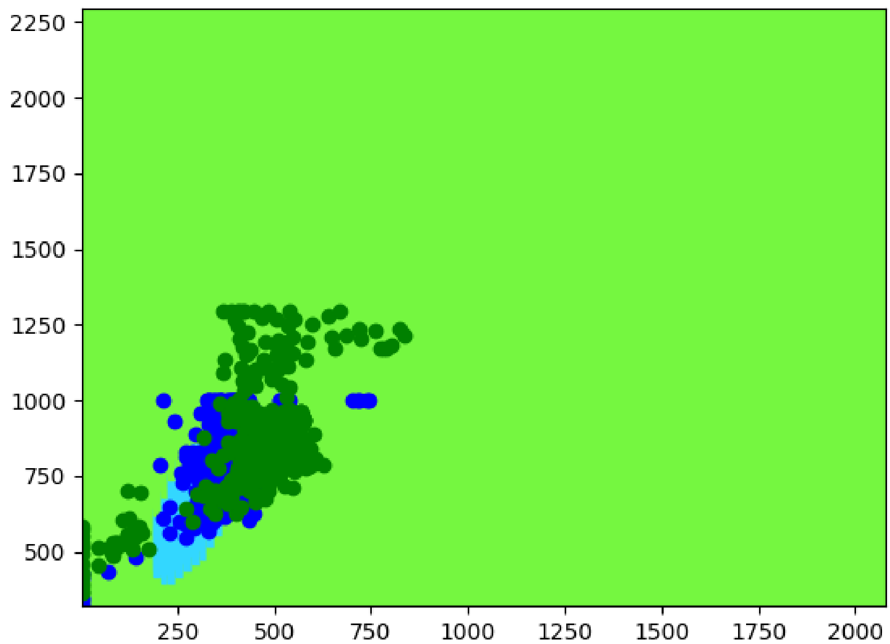

Figure 47: Plot for real data for classes 2 and 3(For different full co-variance matrix)

7.4.1 Confusion Matrix based on performance for test data:

$$\begin{bmatrix} 553 & 8 & 11 \\ 0 & 531 & 91 \\ 0 & 208 & 405 \end{bmatrix}$$

7.4.2 Classification accuracy on test data

	Class1	Class2	Class3
Recall	0.966	0.853	0.660
Precision	1	0.710	0.798
F-measure	0.983	0.7757	0.7232

- Overall Accuracy - 82.4%
- Mean Precision - 0.836
- Mean Recall - 0.824

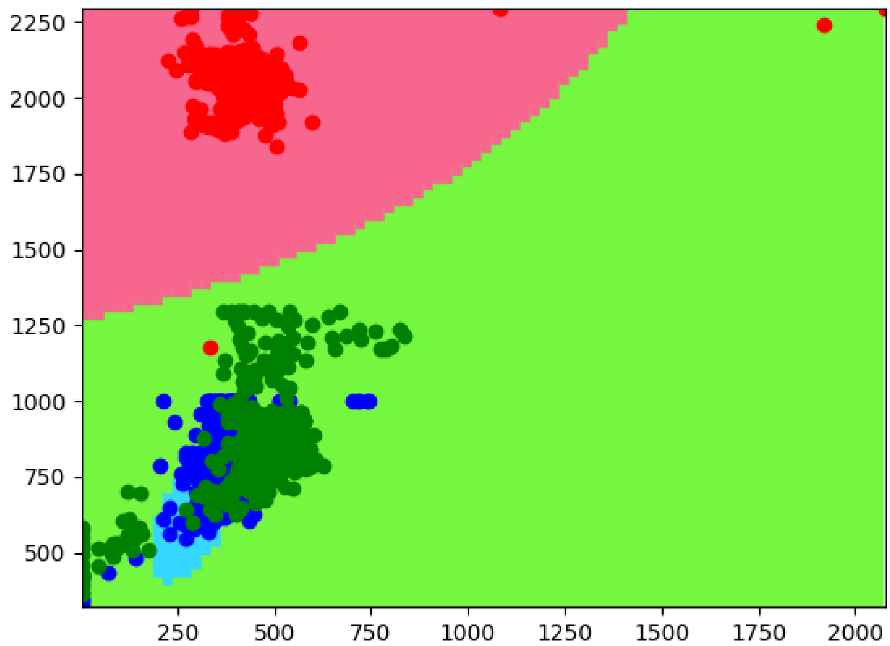

Figure 48: Plot for real data for all classes(For different full co-variance matrix)

- Mean F-Measure - 0.8273

7.4.3 Interpretation

Here the decision boundary has a quadratic form. The surface formed is a better solution to this classification problem. Though decision boundary is non-linear, it is separating class1 and class2 data in quite a good manner, but the other two classes are overlapping which is not covered in the case of Gaussian distribution. Thus overlapping data is not classified much efficiently.
In this when we plot contours, they will be elliptical but will be oriented in the random orientation.

8 Contour Plots

8.1 Linear Separable

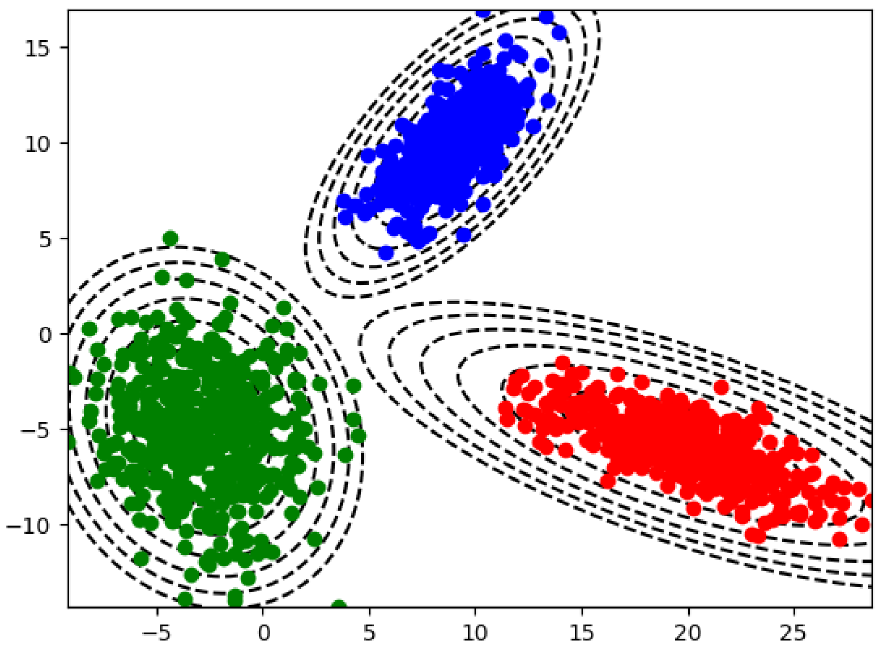

Figure 49: Contour plot for linearly separable data)

8.2 Non-Linear Separable

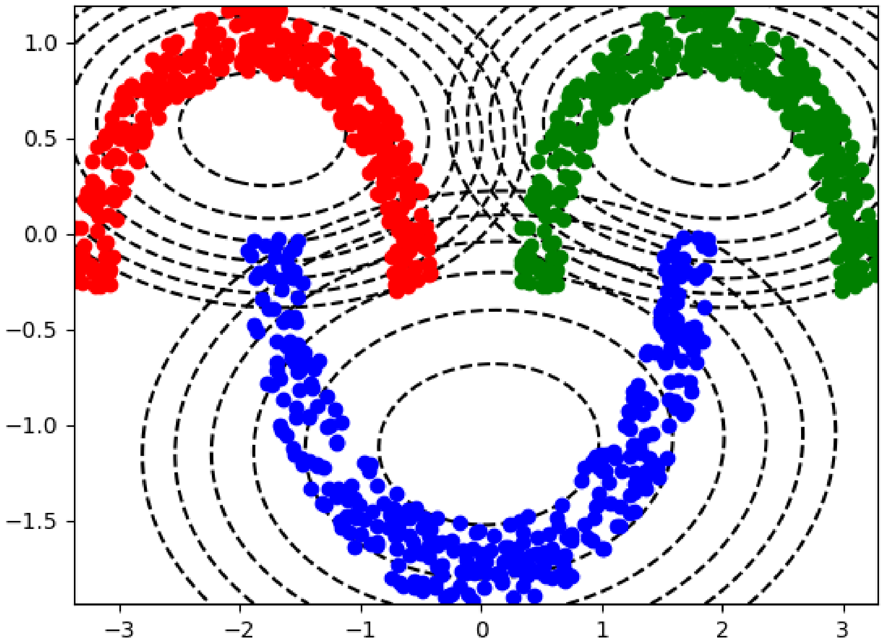

Figure 50: Contour plot for non linearly separable data

8.3 Real data

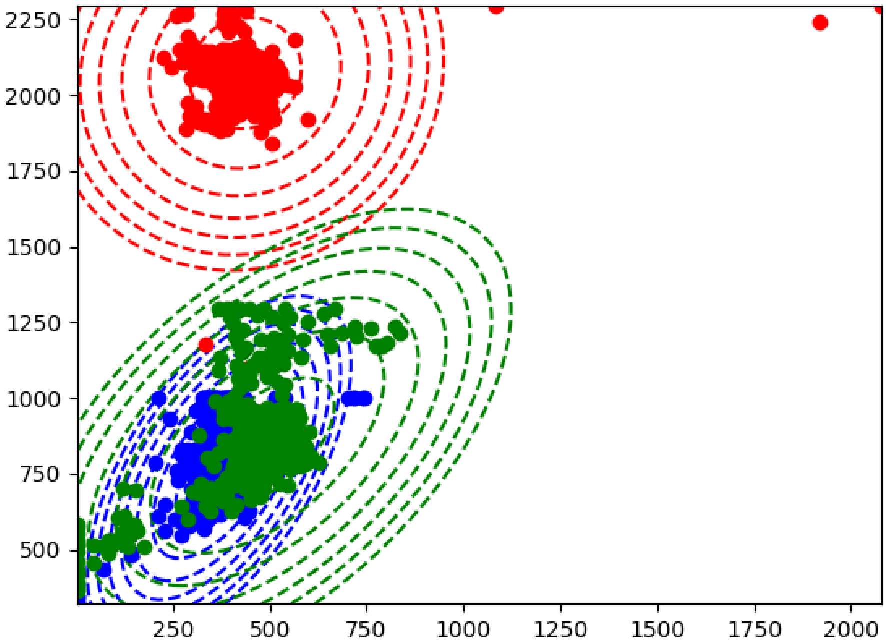

Figure 51: Contour plot for real data)

9 References

1. PATTERN CLASSIFICATION Second Edition by Richard O. Duda ,PETER E. Hart ,David G. Stork.

2. Wikipedia - https://en.wikipedia.org/wiki/Linear$_separability$