CS 669: Pattern Recognition

Professor: Dr. Dileep AD

Assignment 1: Bayes Classifier

Abhinav Dixit (B16003)
Aj R Laddha (B16003)
Hritik Gupta (B16097)

Due Sunday, September 16, 2018

Contents

1 Problem Definition and Dataset Description

1. To build Bayes Classifier and classify the given datasets -
   - Linearly separable artificial data
   - Non-linearly separable artificial data
   - Real World data
2. Few features of the datasets:
   - The Linear and Real-World data consists of 3 classes, whereas, the Non Linear data consists of 2 classes
   - Each data point consists of two features

Training Set : The first 75% data of each class has been taken as training set. The classifier is designed on the basis of parameters determined from this set.
Test Set : The remaining 25% data of each class has been taken as test set. The classifier takes this data as input and gives the class label for this data as output.

2 Task

We need to design classifiers considering varying constraints on covariance matrices of all classes resulting to four distinct cases. For every case we consider each dataset separately. For each point in each test dataset, our classifier predicts the class it belongs to. The results of this classification for each dataset is presented in the form of a confusion matrix which is further analyzed in terms of classification accuracy, precision, recall and F-measure.

Assumption : Class-conditional densities are Gaussian.
Based on the assumption to the covariance matrices of all classes, we are left with the following four cases :

1. Covariance matrix for all the classes is the same and is $\sigma^2 I$.
2. Full Covariance matrix for all the classes is the same and is Σ
3. Covariance matrix is diagonal and is different for each class.
4. Full Covariance matrix for each class is different.

3 Classification Criteria

1. Classifying between two classes(C_i and C_j) at a time : For every data point x in test data we calculate g_i(x) for i and j. Then we provide class-label for x as :
   L(x) : argmax{i}g_i(x)

2. Classifying among three classes(C_i , C_j and C_k) at a time : For every data point x in test data we calculate g_i(x) for i, j and k. Then we provide class-label for x :
   L(x) : argmax{i}g_i(x)

The following is a detailed insight into the aforementioned cases :

1. Covariance matrix for all the classes is the same and is $\sigma^2 I$.
   The general form of covariance matrix for a class having data which has two features is following

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

   σ_{11} : Variance of feature 1
   σ_{12} : Covariance of feature 1 with feature 2.
   σ_{21} : Covariance of feature 2 with feature 1.
   σ_{22} : Variance of feature 2

   For the above class average covariance :
   $\sigma_i^2 = \frac{(\sigma_{11}+\sigma_{12}+\sigma_{21}+\sigma_{22})}{4}$
   σ_i^2 : Average covariance for each class.
   Furthermore we define σ^2 for n classes as :
   $\sigma^2 = \frac{(\sigma_1^2+\sigma_2^2 \cdots +\sigma_n^2)}{n}$
   For our particular case with three classes this reduces to :
   $\sigma^2 = \frac{(\sigma_1^2+\sigma_2^2+\sigma_3^2)}{3}$

   With this idea the training data for all three datasets is analyzed to get their corresponding covariances i.e. σ_1^2 , σ_2^2, and σ_3^2 These covariance give the required σ^2 . The equation for discriminating function for each class in this case is :
   $g_i(x)$ = -[x^tx - u_ix + $u_i^t u_i$]/$2\sigma^2$ + lnP(C_i)
   u_i : Mean vector of i^{th} class.
   lnP(C_i) : natural logarithm of prior probability of i^{th} class.

   In our case we know a prior that every dataset has classes which have equal number of data points in them, thus prior probability for each class P(C_i) is same and can be neglected from the discriminating function. Therefore, the discriminating function for i^{th} class now becomes
   $g_i(x) = -\frac{[x^t x - u_i x + u_i^t u_i]}{2\sigma^2}$

2. Full Covariance matrix for all the classes is the same and is Σ.
   $\Sigma = \frac{\Sigma_1+\Sigma_2+\Sigma_3}{3}$
   $g_i(x)$ = -(x-u_i^t)Σ^{-1}(x-u_i) + lnP(C_i)

3. Covariance matrix is diagonal and is different for each class.

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

   σ_{11} : Variance of feature 1
   σ_{12} : Covariance of feature 1 with feature 2.
   σ_{21} : Covariance of feature 2 with feature 1.
   σ_{22} : Variance of feature 2
   We consider $\sigma_{12} = \sigma_{21} = 0$
   So the covariance matrix is now ;

$$\Sigma = \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix}$$

   The discriminating equation thus produced is :
   $g_i(X) = x^t W_i x + w_i^t$x + C_{i0}
   W_i =-$\frac{\Sigma_i^-1}{2}$
   $w_i = \Sigma_i^{-1} u_i$
   $C_{i0} = -u_i^t \Sigma_i^{-1} u_i - (\ln|\Sigma_i|)/2 + \ln P(C_i)$
   u_i : Mean vector of i^{th} class.
   lnP(C_i) : natural logarithm of prior probability of i^{th} class.

4. Full Covariance matrix for each class is different. The general form of covariance matrix for a class having data which has two features is:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

   σ_{11} : Variance of feature 1
   σ_{12} : Covariance of feature 1 with feature 2.
   σ_{21} : Covariance of feature 2 with feature 1.
   σ_{22} : Variance of feature 2

   The discriminating equation thus produced is :
   $g_i(X) = x^t W_i x + w_i^t$x + C_{i0}
   W_i =-$\frac{\Sigma_i^-1}{2}$

$w_i = \Sigma_i^{-1} u_i$
$C_{i0} = -u_i^t \Sigma_i^{-1} u_i - (\ln|\Sigma_i|)/2 + \ln P(C_i)$
u_i : Mean vector of i^{th} class.
$\ln P(C_i)$: natural logarithm of prior probability of i^{th} class.

Formulas to determine Result-

1. **Accuracy**= $(\frac{Total\ correct\ classification}{Total\ classification}) \times 100$
2. **Precision**
   For a class i
   $P_i = (\frac{Correct\ classification\ for\ class\ i}{Total\ classification\ for\ class\ i}) \times 100$
3. **Recall**
   For a class i
   $RC_i = (\frac{Correct\ classification\ for\ class\ i}{Total\ data\ points\ in\ class\ i}) \times 100$
4. **Mean Recall** = $\frac{(RC_1 + RC_2 + \cdots + RC_n)}{n} \times 100$
5. **F-Measure**
   For a class i :
   $FM_i = \frac{(PC_i \times RC_i \times 2)}{PC_i + RC_i}$
6. **Mean F-Measure** = $\frac{(FC_1 + FC_2 + \cdots + FC_n)}{n} \times 100$

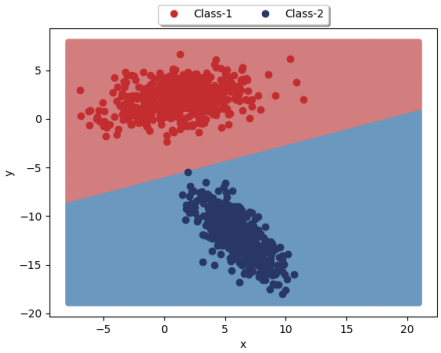
4 Procedure

1. Data for each class is partitioned into 75% for training and 25% for testing.
2. The data set for each class is assumed to be coming from Gaussian distribution.
3. In case 1 ($\Sigma = \sigma^2 I$), mean of covariance matrix for each class was calculated and its off diagonal terms were assumed to be 0 for further calculations.
4. In case 2 ($\Sigma_i = \Sigma$ for every class), mean of covariance matrix for each class was calculated for further calculations.
5. In case 3 (Σ_i is diagonal matrix), covariance matrix for each class was different and its off diagonal terms were assumed to be 0 for further calculations.
6. In case 4 (Σ_i is unique), no assumptions were made for further calculations.
7. Based of assumptions, the discriminant function ($g_i(x)$) was calculated for each class and decision region and Contour was plotted
8. The remaining 25 % data was tested for each case and further analysis was made.

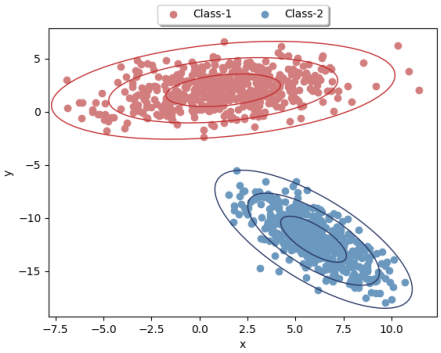
5 Observations and Results

5.1 Case 1 - $\Sigma = \sigma^2 I$

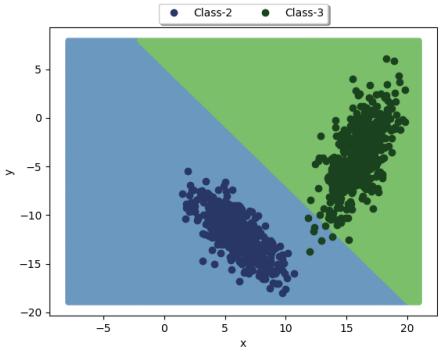
5.1.1 Linear Data



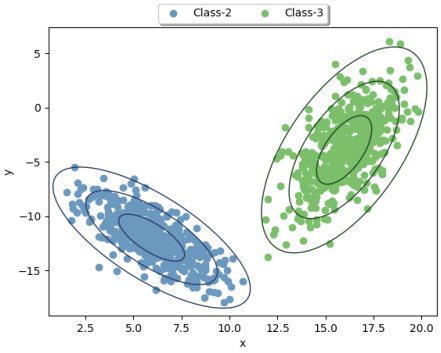
Case1 v/s Case2



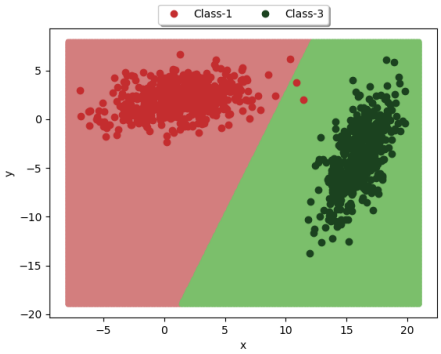
Case1 v/s Case 2



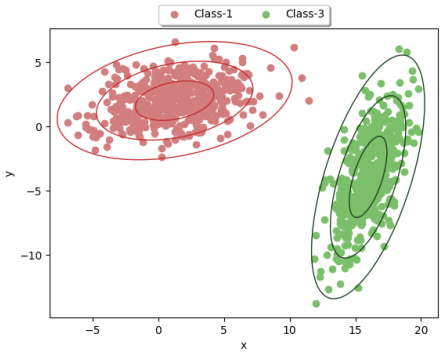
Case2 v/s Case3



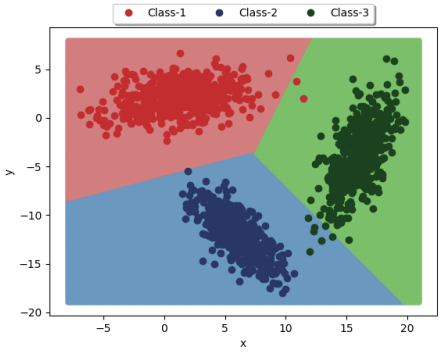
Case2 v/s Case3



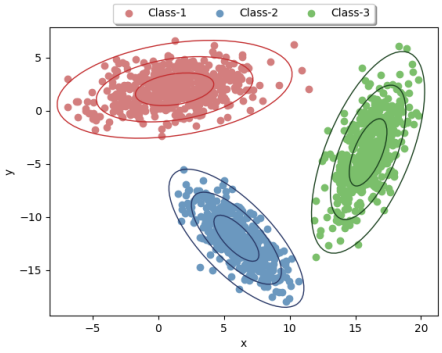
Case1 v/s Case3



Case1 v/s Case3



All cases



All cases

Table 1: Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	123	0	2
Class 2	1	124	0
Class 3	0	2	123

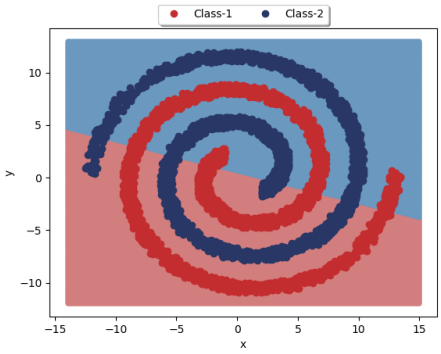
Table 2: Analysis

	Class 1	Class 2	Class 3
Precision	0.9919	0.9841	0.984
Recall	0.984	0.992	0.984
F-Measure	0.9879	0.9980	0.9840

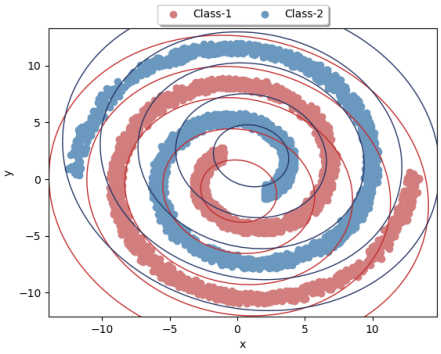
Table 3: Result

	Average Precision	Average Recall	Average F-Mean	Overall Accuracy
Values	0.9866	0.9867	0.9866	0.987

5.1.2 Non-Linear Data



All cases



All cases

Table 4: Confusion Matrix

	Class 1	Class 2
Class 1	387	225
Class 2	232	380

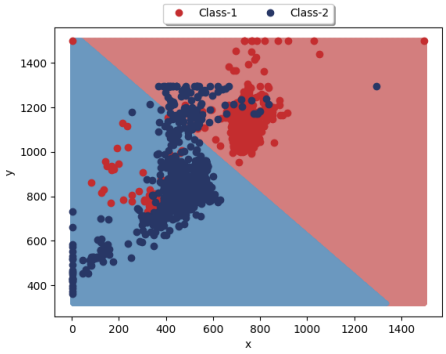
Table 5: Analysis

	Class 1	Class 2
Precision	0.6252	06281
Recall	0.6323	0.6209
F-Measure	0.6287	0.6244

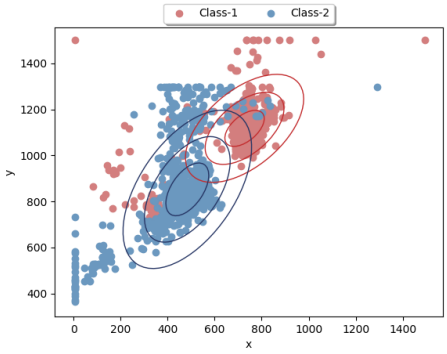
Table 6: Result

	Average Precision	Average Recall	Average F-Mean	Overall Accuracy
Values	0.62665	0.62663	0.62664	0.6266

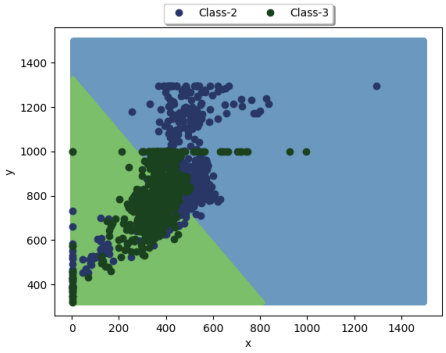
5.1.3 Real- World Data



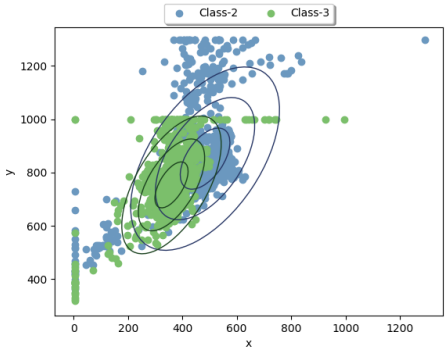
Class1 v/s Class2



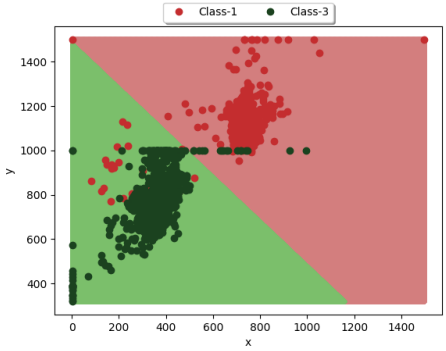
Class1 v/s Class2



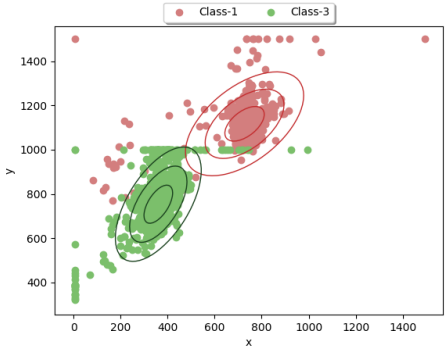
Class2 v/s Class3



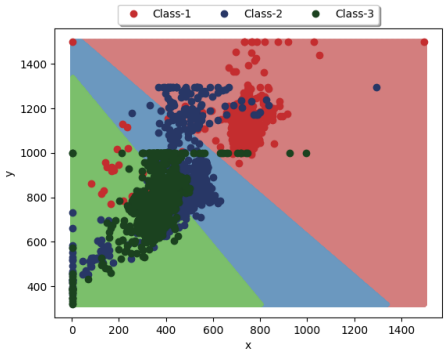
Class2 v/s Class3



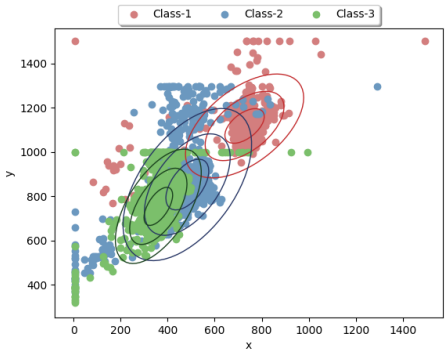
Class1 v/s Class3



Class1 v/s Class3



All classes



All classes

Table 7: Confusion Matrix

	Class 1	**Class 2**	**Class 3**
Class 1	512	5	24
Class 2	44	392	178
Class 3	7	43	572

Table 8: Analysis

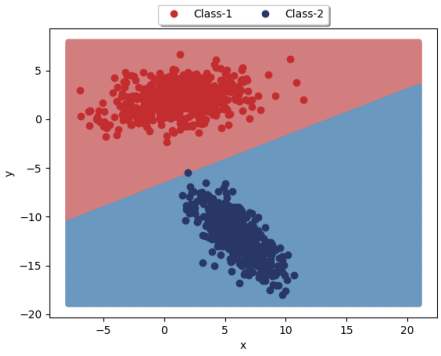
	Class 1	**Class 2**	**Class 3**
Precision	0.9094	0.8909	0.7390
Recall	0.9463	0.6384	0.9196
F-Measure	0.9275	0.7438	0.8194

Table 9: Result

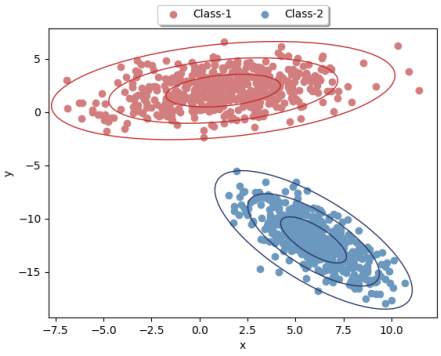
	Average Precision	**Average Recall**	**Average F-Mean**	**Overall Accuracy**
Values	0.84644	0.83481	0.8406	0.8306

5.2 Case 2 - $\Sigma_i = \Sigma$

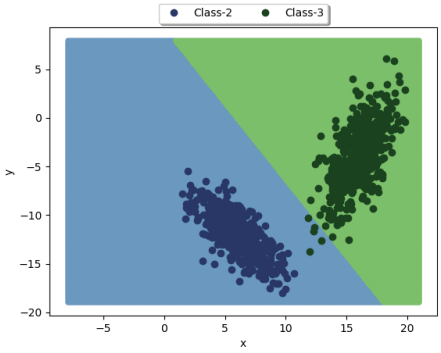
5.2.1 Linear Data



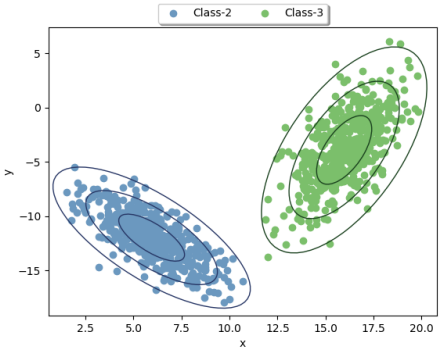
Class1 v/s Class2



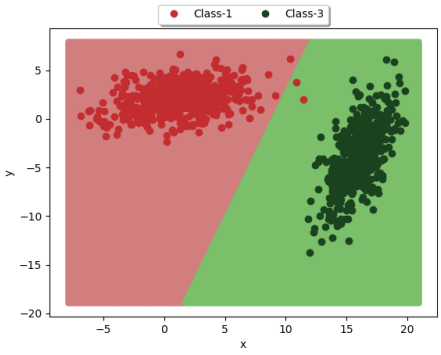
Class1 v/s Class2



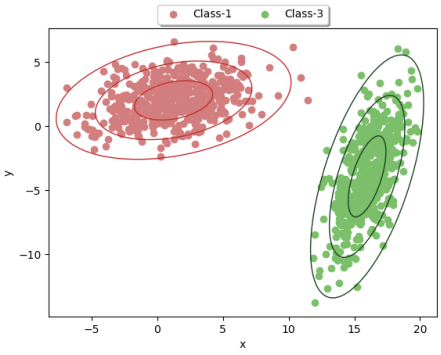
Class2 v/s Class3



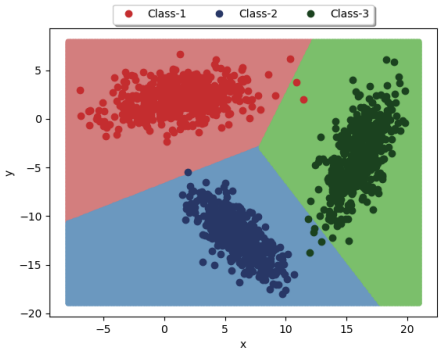
Class2 v/s Class3



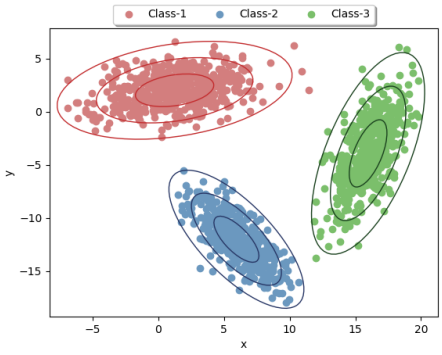
Class1 v/s Class3



Class1 v/s Class3



All classes



All classes

Table 10: Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	123	0	2
Class 2	1	124	0
Class 3	0	2	123

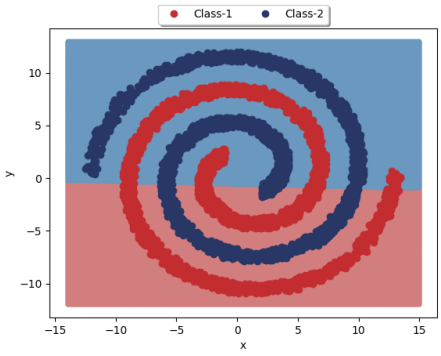
Table 11: Analysis

	Class 1	Class 2	Class 3
Precision	0.9919	0.9841	0.984
Recall	0.984	0.992	0.984
F-Measure	0.9879	0.9980	0.9840

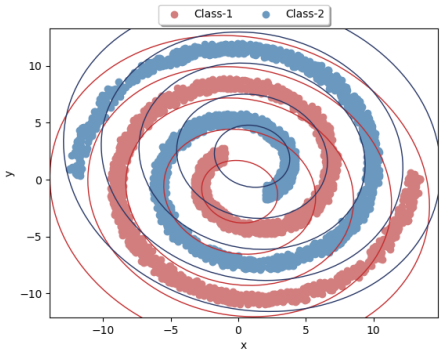
Table 12: Result

	Average Precision	Average Recall	Average F-Mean	Overall Accuracy
Values	0.9866	0.9867	0.9866	0.987

5.2.2 Non-Linear Data



All classes



All classes

Table 13: Confusion Matrix

	Class 1	**Class 2**
Class 1	384	228
Class 2	235	377

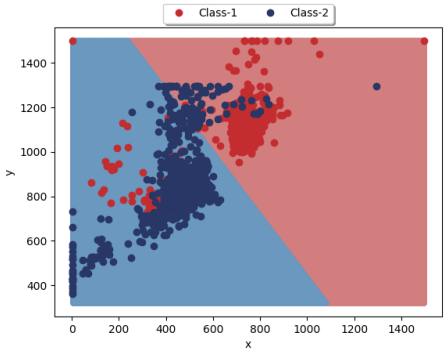
Table 14: Analysis

	Class 1	**Class 2**
Precision	0.6203	06231
Recall	0.6274	0.6160
F-Measure	0.6238	0.6195

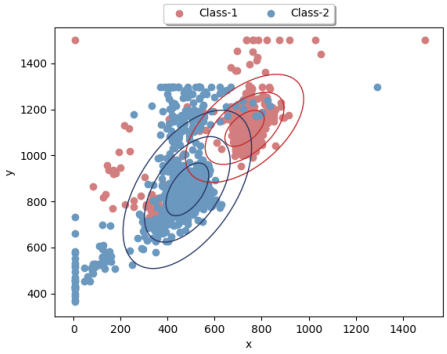
Table 15: Result

	Average Precision	**Average Recall**	**Average F-Mean**	**Overall Accuracy**
Values	0.62174	0.62173	0.62174	0.62173

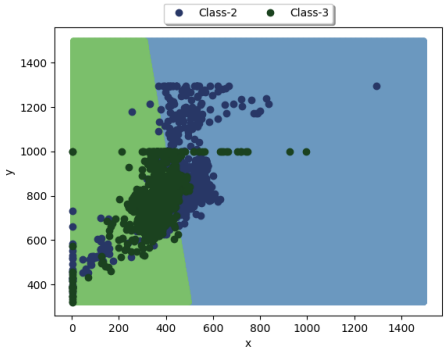
5.2.3 Real- World Data



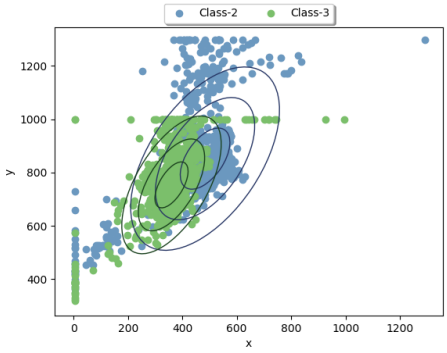
Class1 v/s Class2



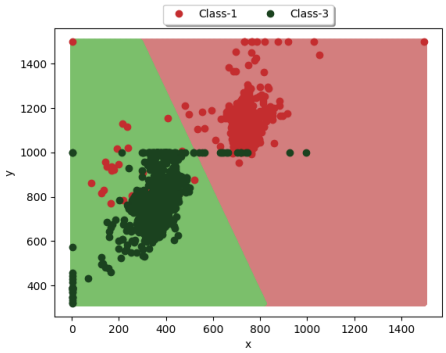
Class1 v/s Class2



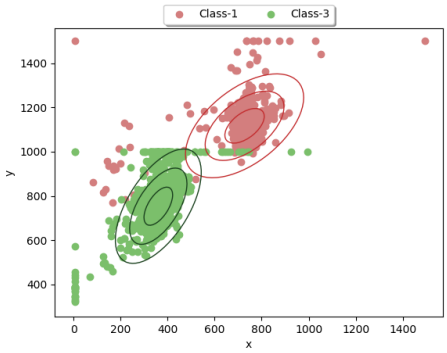
Class2 v/s Class3



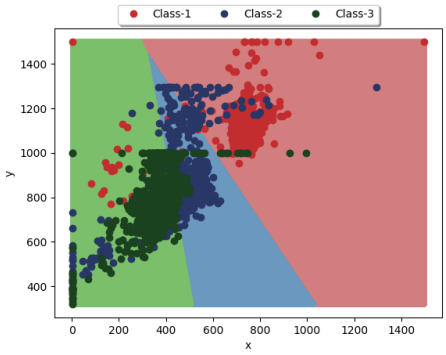
Class2 v/s Class3



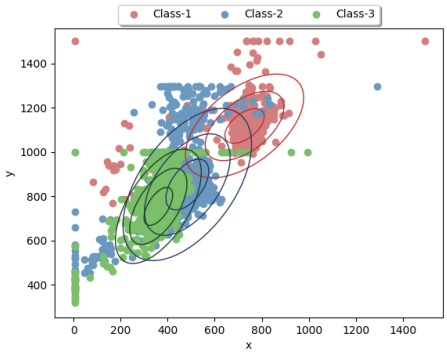
Class1 v/s Class3



Class1 v/s Class3



All classes



All classes

Table 16: Confusion Matrix

	Class 1	**Class 2**	**Class 3**
Class 1	510	5	26
Class 2	19	393	202
Class 3	7	13	602

Table 17: Analysis

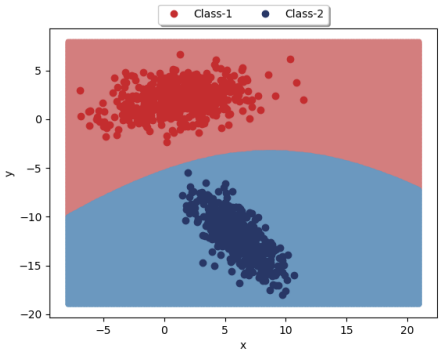
	Class 1	**Class 2**	**Class 3**
Precision	0.9515	0.9562	0.7253
Recall	0.9426	0.6400	0.9678
F-Measure	0.9470	0.7668	0.8292

Table 18: Result

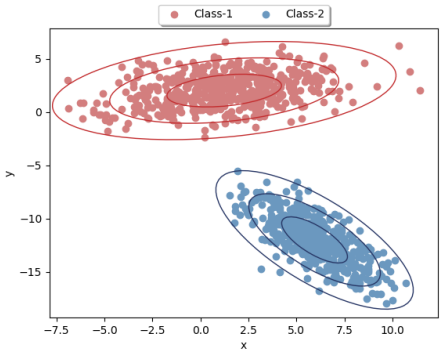
	Average Precision	**Average Recall**	**Average F-Mean**	**Overall Accuracy**
Values	0.8776	0.8502	0.8637	0.8469

5.3 Case 3 - $\Sigma_i$is a diagonal matrix and different for each class

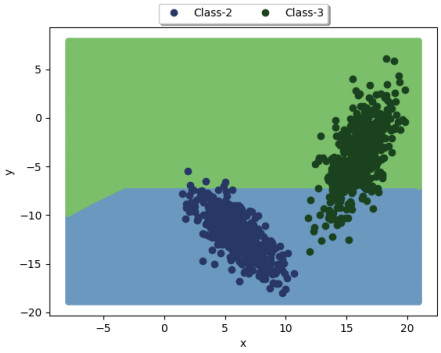
5.3.1 Linear Data



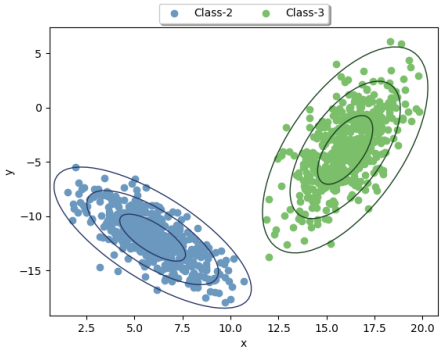
Class1 v/s Class2



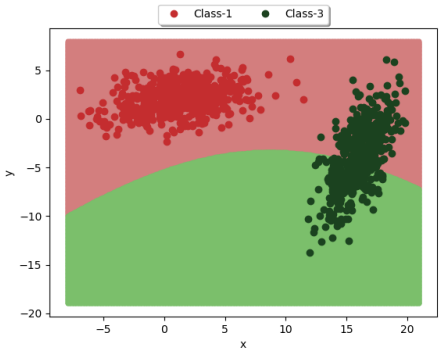
Class1 v/s Class2



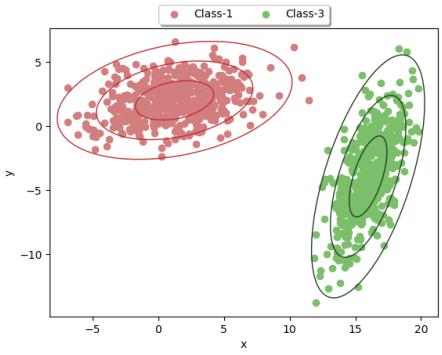
Class2 v/s Class3



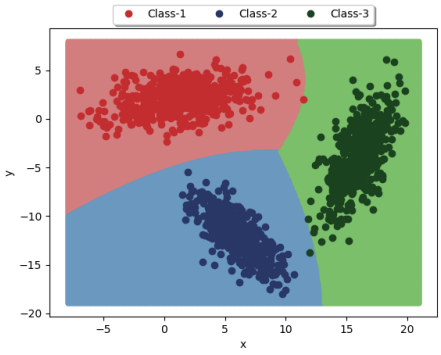
Class2 v/s Class3



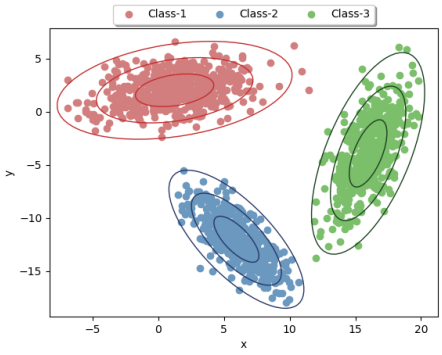
Class1 v/s Class3



Class1 v/s Class3



All classes



All classes

Table 19: Confusion Matrix

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

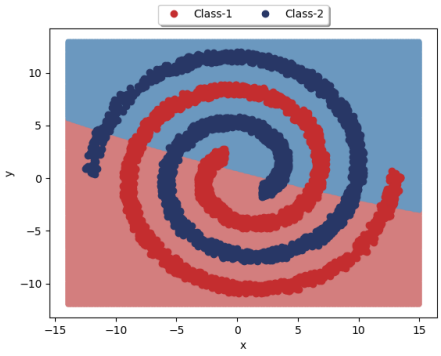
Table 20: Analysis

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

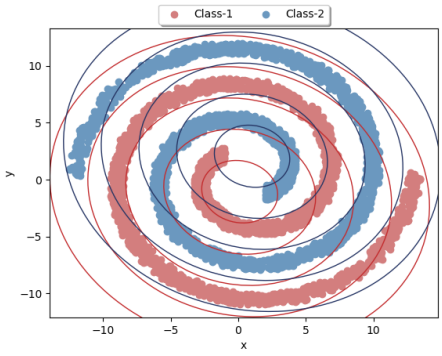
Table 21: Result

	Average Precision	Average Recall	Average F-Mean	Overall Accuracy
Values	1.0	1.0	1.0	1.0

5.3.2 Non-Linear Data



All classes



All classes

Table 22: Confusion Matrix

	Class 1	**Class 2**
Class 1	397	215
Class 2	246	366

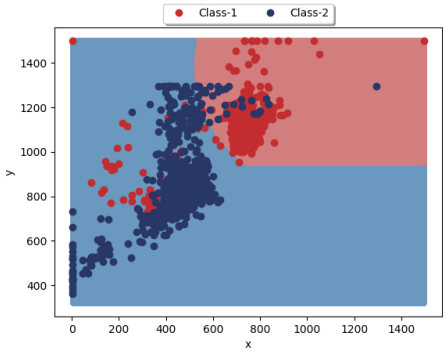
Table 23: Analysis

	Class 1	**Class 2**
Precision	0.6174	0.6299
Recall	0.6486	0.5980
F-Measure	0.632	0.613

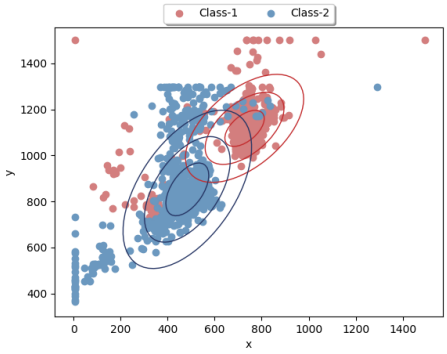
Table 24: Result

	Average Precision	**Average Recall**	**Average F-Mean**	**Overall Accuracy**
Values	0.6236	0.6233	0.6235	0.6233

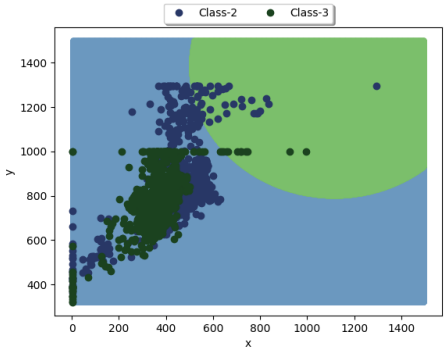
5.3.3 Real- World Data



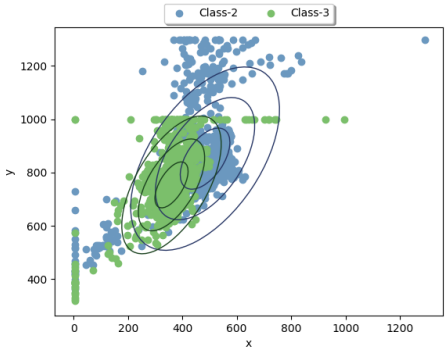
Class1 v/s Class2



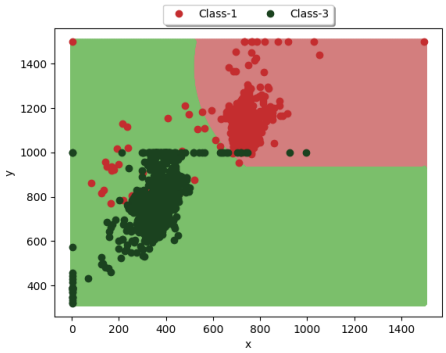
Class1 v/s Class2



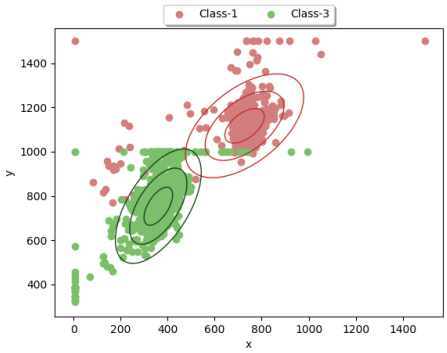
Class2 v/s Class3



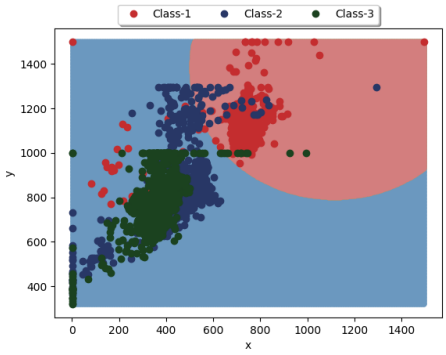
Class2 v/s Class3



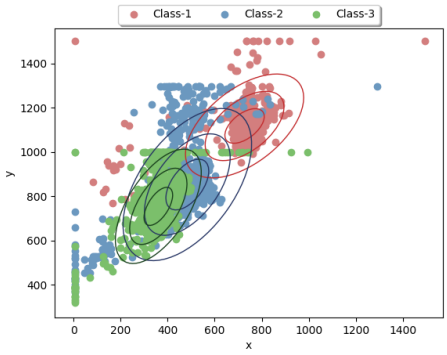
Class1 v/s Class3



Class1 v/s Class3



All classes



All classes

Table 25: Confusion Matrix

	Class 1	**Class 2**	**Class 3**
Class 1	509	21	11
Class 2	15	569	30
Class 3	3	119	500

Table 26: Analysis

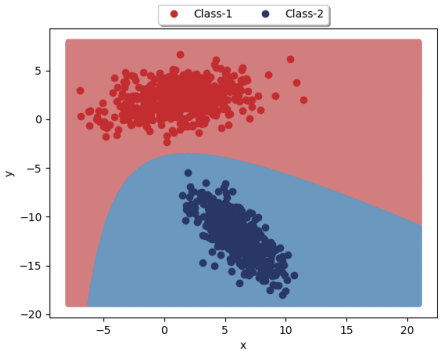
	Class 1	**Class 2**	**Class 3**
Precision	0.965	0.802	0.924
Recall	0.940	0.926	0.803
F-Measure	0.953	0.792	0.707

Table 27: Result

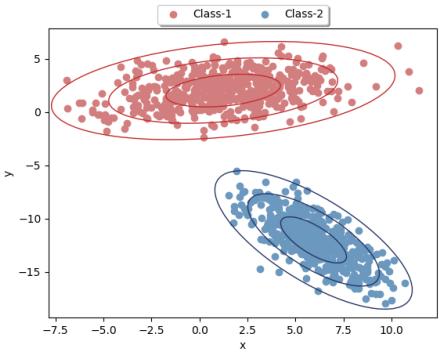
	Average Precision	**Average Recall**	**Average F-Mean**	**Overall Accuracy**
Values	0.869	0.825	0.846	0.819

5.4 Case 4 - $\Sigma_i$is unique

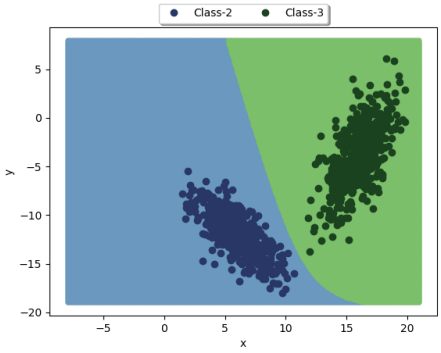
5.4.1 Linear Data



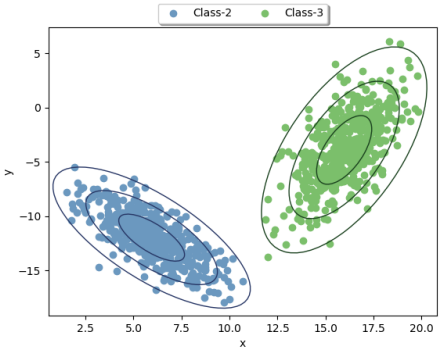
Class1 v/s Class2



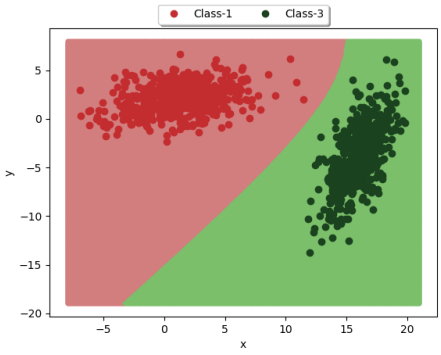
Class1 v/s Class2



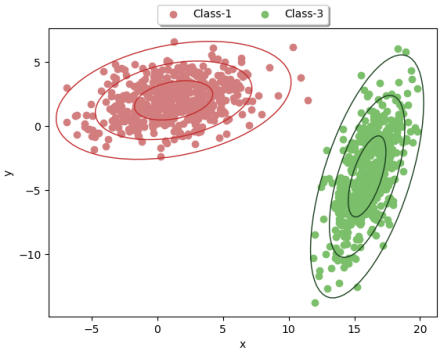
Class2 v/s Class3



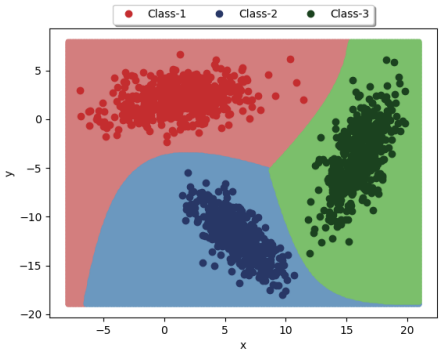
Class2 v/s Class3



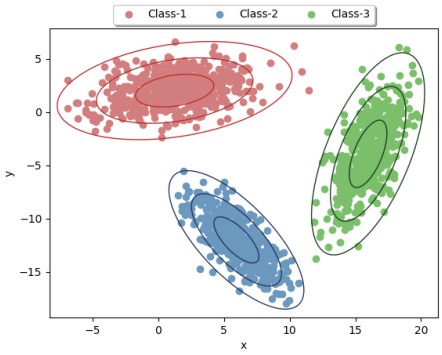
Class1 v/s Class3



Class1 v/s Class3



All classes



All classes

Table 28: Confusion Matrix

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

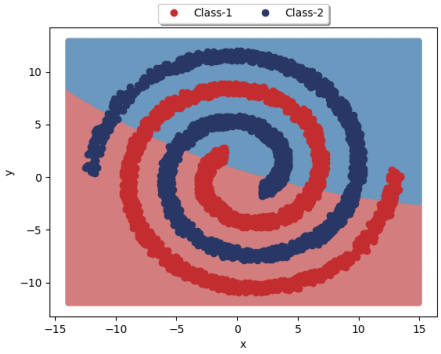
Table 29: Analysis

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

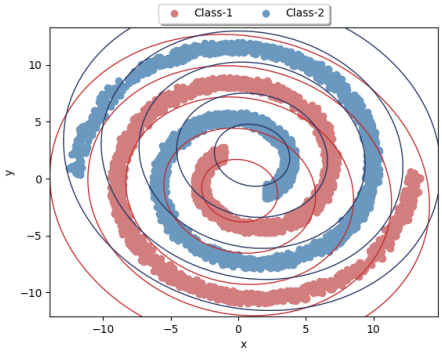
Table 30: Result

	Average Precision	**Average Recall**	**Average F-Mean**	**Overall Accuracy**
Values	1.0	1.0	1.0	1.0

5.4.2 Non-Linear Data



All classes



All classes

Table 31: Confusion Matrix

	Class 1	**Class 2**
Class 1	400	202
Class 2	242	370

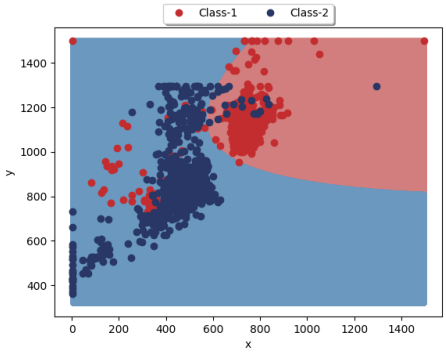
Table 32: Analysis

	Class 1	**Class 2**
Precision	0.62305	0.637
Recall	0.653	0.604
F-Measure	0.6379	0.6197

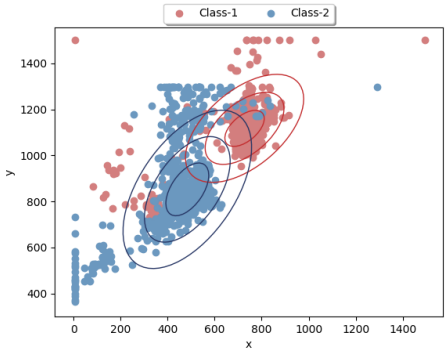
Table 33: Result

	Average Precision	**Average Recall**	**Average F-Mean**	**Overall Accuracy**
Values	0.6293	0.62908	0.62924	0.62908

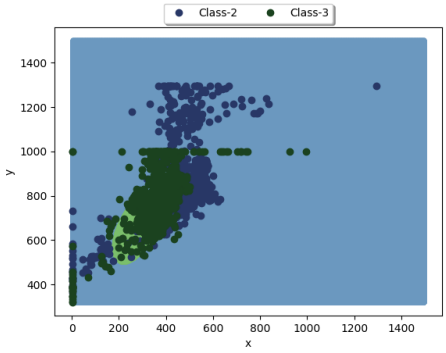
5.4.3 Real- World Data



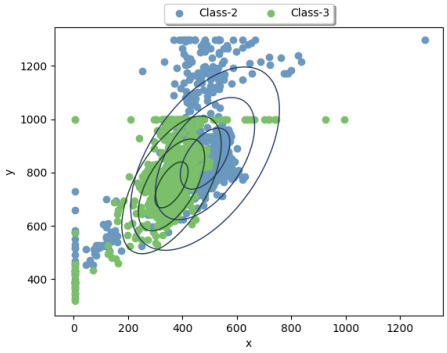
Class1 v/s Class2



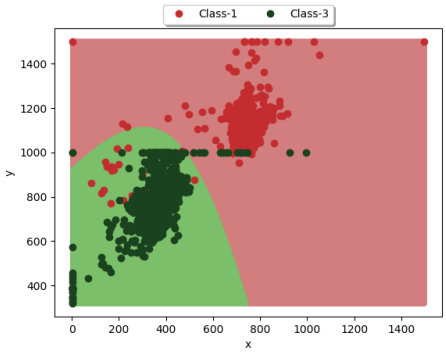
Class1 v/s Class2



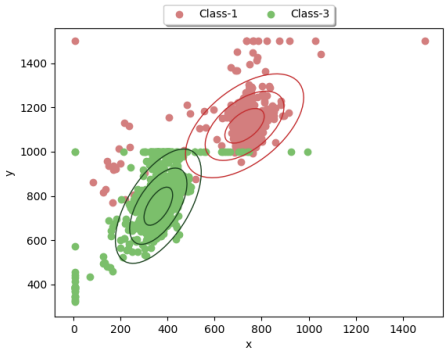
Class2 v/s Class3



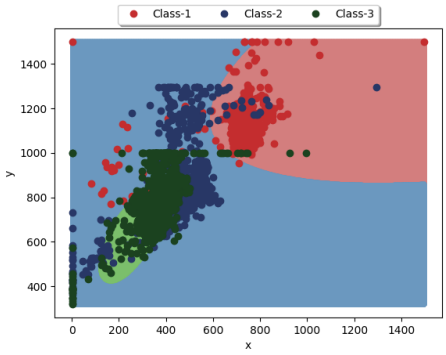
Class2 v/s Class3



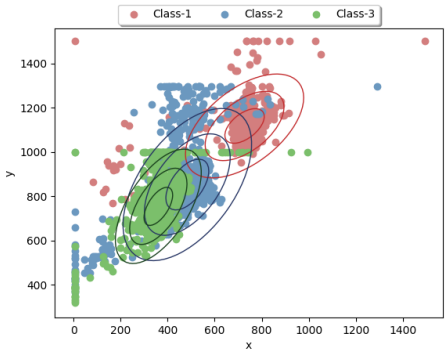
Class1 v/s Class3



Class1 v/s Class3



All classes



All classes

Table 34: Confusion Matrix

	Class 1	**Class 2**	**Class 3**
Class 1	508	27	6
Class 2	5	598	11
Class 3	2	270	350

Table 35: Analysis

	Class 1	**Class 2**	**Class 3**
Precision	0.986	0.668	0.953
Recall	0.939	0.973	0.562
F-Measure	0.962	0.792	0.707

Table 36: Result

	Average Precision	**Average Recall**	**Average F-Mean**	**Overall Accuracy**
Values	0.869	0.825	0.846	0.819

6 Discussion

We have built four different types of Bayes classifier constrained on covariance on three different sets of two dimensional data. From the distribution of linearly separable data itself it can be inferred that all the classes are ellipsoidal which indicates equal non-zero covariance. The major and minor axis of the ellipse tell us the direction of distribution. For instance the ellipsoidal distribution of class 1 and class 3 are inclined with a positive slope of major axis indicating positive covariance whereas class 2 distribution is inclined in the opposite direction for its negative covariance. Again the minor axis of class 1 distribution ellipse is smaller compared to class 2 and 3 and the major axis is more horizontal, with which one can conclude that the variance of first dimension of data is more than the second for class 1 compared with other two classes. Similarly the spread of class 3 is along vertical direction meaning variance of second dimension of class 3 is more. All of these inferences are verified by the calculation of covariance matrix of three different classes and has been shown in the table. From this the best classifier that should give better accuracy is the one with arbitrary covariance matrix and hence the decision boundary must be hyper-quadratic for distinct classification. For non-linear artificial data all the four types of classifier gave linear boundary because the distribution of both the classes are identical which is verified by calculation of covariance matrix resulting to be nearly equal. However, the distribution of real data is completely random, but still mostly clustered in their distribution and hence was satisfactorily classified. The naive Bayes classifier which assumes diagonal covariance matrix with different variances for each class gave the best accuracy as the classes are highly correlated. From the above experimentation we infer that the choice of Bayes classifier depends on the distribution of data and hence on their covariance matrix. The decision boundary for Bayes classifier with same variance and/or same covariance gave a linear decision boundary whereas Naive Bayes and Bayes classifier with arbitrary covariance matrix gave hyper-quadratic decision boundary.