

# CS669 - Pattern Recognition

---

## **Bayes Classifier**

Assignment 1

**Group 14**

---

Team:

Aashish Kumar (B16001)

Priyanshu Khandelwal (B16027)

Vishal Anand (B16040)

<b>Bayes Classifier</b>	<b>1</b>
<b>1. Objective</b>	<b>4</b>
<b>2. Procedure</b>	<b>4</b>
<b>3. Observations</b>	<b>5</b>
3.1 Case 1: $\Sigma = \sigma^2 I$	5
3.1.1 Linear Separable Data	5
3.1.2 Non-Linear Data	7
3.1.3 Real World Data	9
3.2 Case 2: $\Sigma_i = \Sigma$	11
3.2.1 Linear Separable Data	11
3.2.2 Non-Linear Data	13
3.2.3 Real World Data	15
3.3 Case 3: $\Sigma_i$ is a diagonal matrix	17
3.3.1 Linear Separable Data	17
3.3.2 Non-Linear Data	19
3.3.3 Real World Data	21
3.4 Case 4: $\Sigma_i$ is unique	23
3.4.1 Linear Separable Data	23
3.4.2 Non-Linear Data	25
3.4.3 Real World Data	27
<b>4. Conclusion</b>	<b>29</b>

# 1. Objective

1. To build Bayes Classifier and classify the following datasets -
  - a. 2D Dataset - 1 (Artificial)
    - i. Linearly Separable Dataset
    - ii. Non Linearly Separable Dataset
  - b. 2D Dataset - 2 (Real World)
2. Plot Decision Region for all pairs of classes.
3. Contour Region Plots for all pairs of classes.
4. Calculate Accuracy, Precision, mean Recall, F Measure and Confusion Matrix.

# 2. Procedure

1. For each class, 75% of data is used for training and 25% of data is used for testing.
2. It is assumed that the data for each class is obtained from Gaussian distribution.
3. In case 1 ( $\Sigma = \sigma^2 I$ ), the diagonal terms are all taken to be equal of the covariance matrix and was the mean for each class and it's off diagonal terms were assumed to be zero for simplicity in 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 ( $\Sigma_i$  is diagonal matrix), covariance matrix for each class was different and it's off diagonal terms were assumed to be 0 for further calculations.
6. In case 4 ( $\Sigma_i$  is unique), all individual covariance matrices were taken without any assumptions.
7. On the basis of assumptions taken, 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 analysis table were filled for each case.

### 3. Observations

#### 3.1 Case 1: $\Sigma = \sigma^2 I$

##### 3.1.1 Linear Separable Data

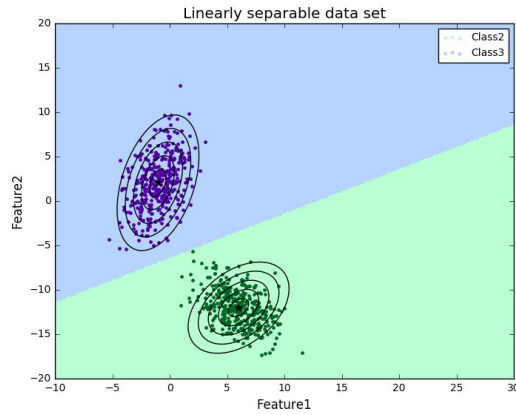


Fig. 1a

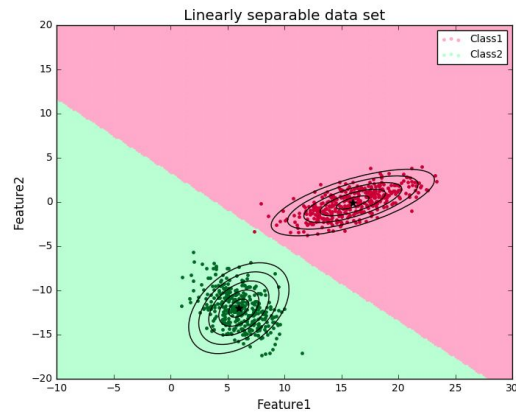


Fig. 1b

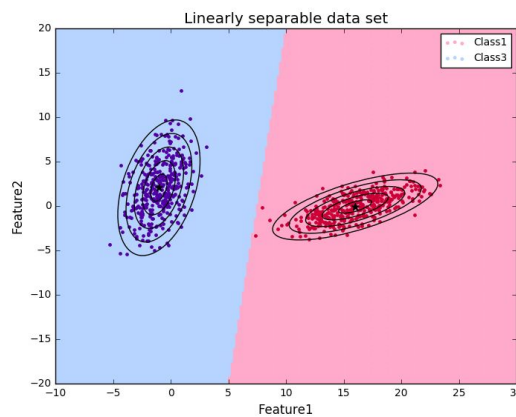


Fig. 1c

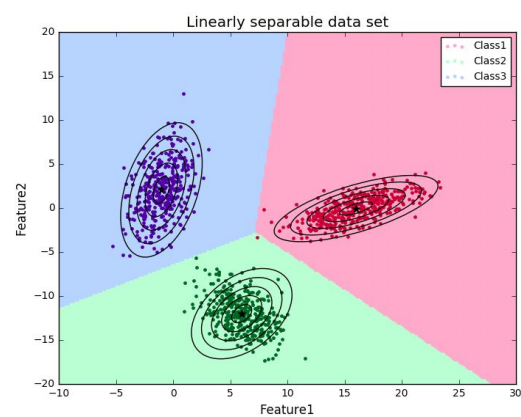


Fig. 1d

Fig1. Decision boundary for Case 1 Linear Separable data depicting between : a) Class 2 and Class 3 b) Class 1 and Class 2 c) Class 1 and Class 3 d) Class 1,2 and 3

**Confusion Matrix:**

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

Accuracy: 1.0

F Measure: 1.0

**Analysis:**

	Class 1	Class 2
Precision	1.0	1.0
Recall	1.0	1.0

**Confusion Matrix:**

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

Accuracy: 1.0

F Measure: 1.0

**Analysis:**

	Class 2	Class 3
Precision	1.0	1.0
Recall	1.0	1.0

**Confusion Matrix:**

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

Accuracy: 1.0

F Measure: 1.0

**Analysis:**

	Class 1	Class 3
Precision	1.0	1.0
Recall	1.0	1.0

**Confusion Matrix:**

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

Accuracy: 1.0

F Measure: 1.0

**Analysis:**

	Class 1	Class 2	Class 3
Precision	1.0	1.0	1.0
Recall	1.0	1.0	1.0

Table 1. Confusion Matrix and Analysis for Case 1, Linear Separable Data

### 3.1.2 Non-Linear Data

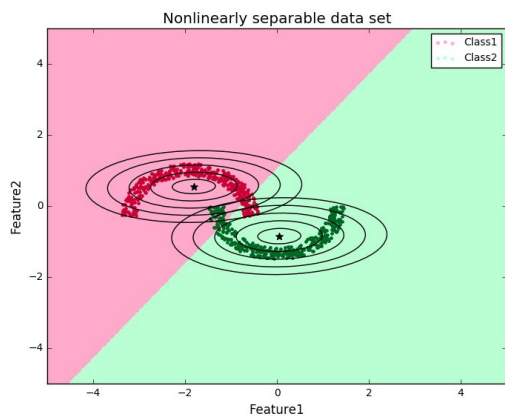


Fig. 2a

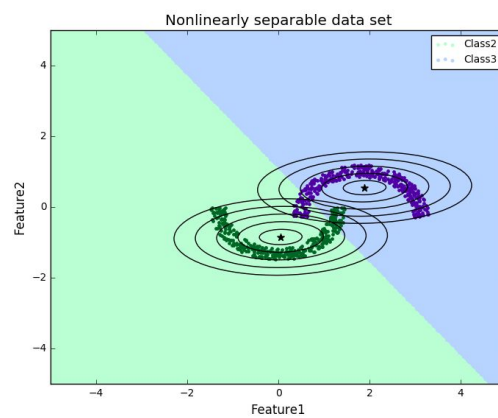


Fig. 2b

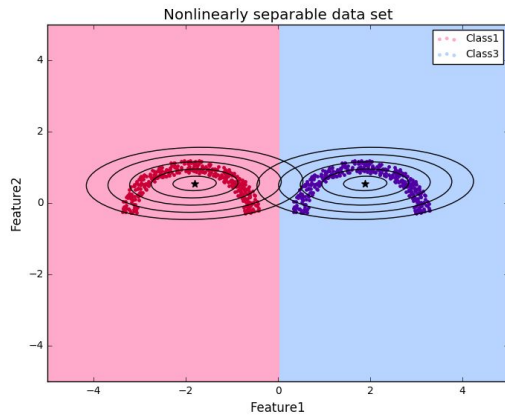


Fig. 2c

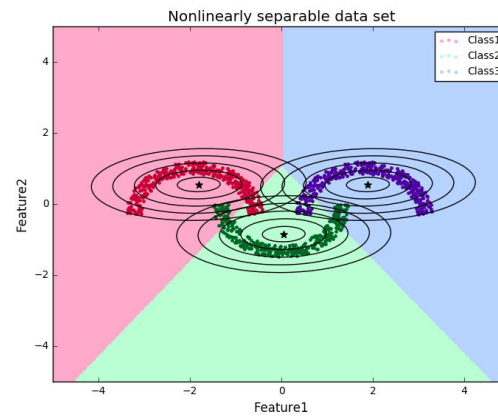


Fig. 2d

Fig2. Decision boundary for Case 1 Non-Linear Separable data depicting between : a) Class 1 and Class 2 b) Class 2 and Class 3 c) Class 1 and Class 3 d) Class 1,2 and 3

**Confusion Matrix:**

	Class 1	Class 2
Class 1	110	15
Class 2	20	105

Accuracy: 0.8600

F Measure: 0.8603

**Analysis:**

	Class 1	Class 2
Precision	0.8461	0.8750
Recall	0.8800	0.8400

**Confusion Matrix:**

	Class 2	Class 3
Class 2	106	19
Class 3	15	110

Accuracy: 0.8640

F Measure: 0.8642

**Analysis:**

	Class 2	Class 3
Precision	0.8760	0.8520
Recall	0.8480	0.8800

**Confusion Matrix:**

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

Accuracy: 1.0

F Measure: 1.0

**Analysis:**

	Class 1	Class 3
Precision	1.0	1.0
Recall	1.0	1.0

**Confusion Matrix:**

	Class 1	Class 2	Class 3
Class 1	110	15	0
Class 2	20	86	19
Class 3	0	15	110

Accuracy: 0.8160

F Measure: 0.8147

**Analysis:**

	Class 1	Class 2	Class 3
Precision	0.8460	0.7410	0.8530
Recall	0.8800	0.6880	0.8800

Table 2. Confusion Matrix and Analysis for Case 1, Non-Linear Separable Data

### 3.1.3 Real World Data

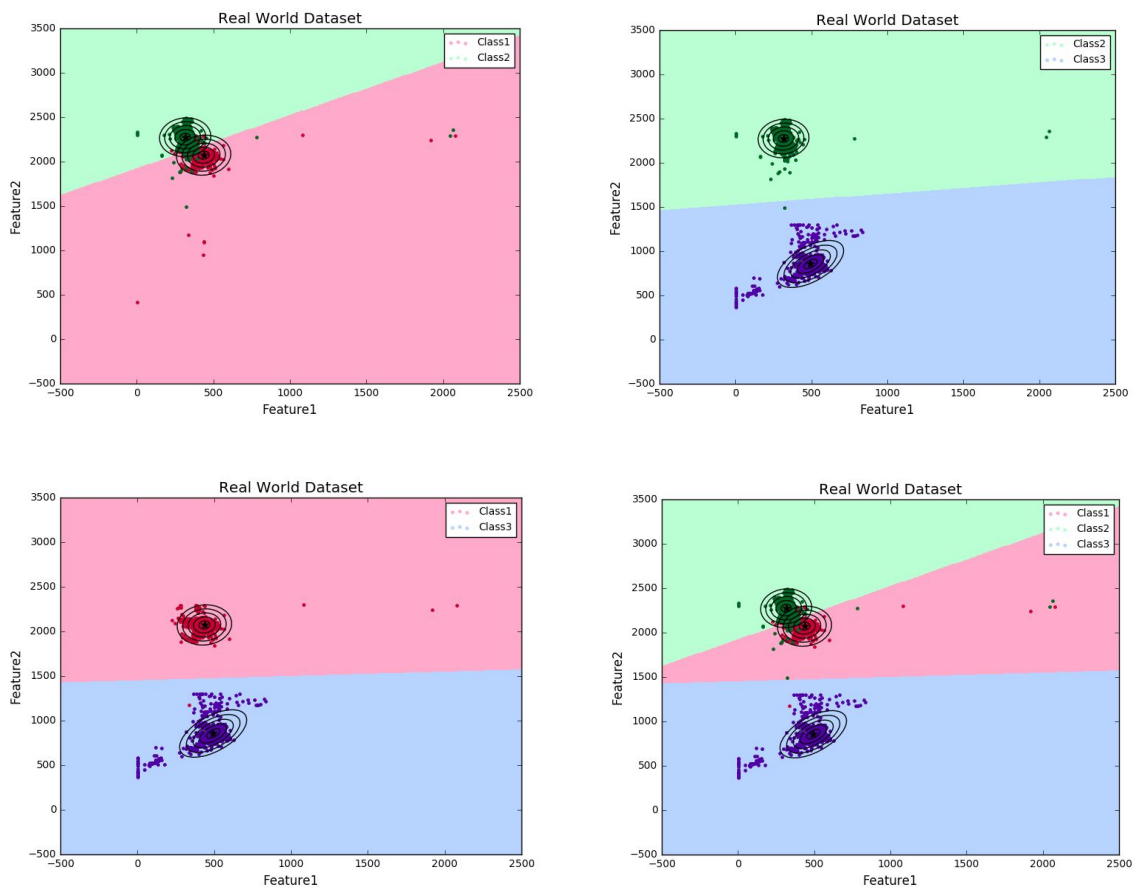


Fig3. Decision boundary for Case 1 Real World data depicting between : a) Class 1 and Class 2 b) Class 2 and Class 3 c) Class 1 and Class 3 d) Class 1,2 and 3



**Confusion Matrix:**

	Class 1	Class 2
Class 1	224	348
Class 2	27	570

Accuracy: 0.6790

F Measure: 0.7120

**Analysis:**

	Class 1	Class 2
Precision	0.8924	0.6210
Recall	0.3920	0.9540

**Confusion Matrix:**

	Class 2	Class 3
Class 1	593	4
Class 2	0	613

Accuracy: 0.9967

F Measure: 0.9967

**Analysis:**

	Class 2	Class 3
Precision	1.0	0.9930
Recall	0.9930	1.0

**Confusion Matrix:**

	Class 1	Class 3
Class 1	554	18
Class 3	0	613

Accuracy: 0.9850

F Measure: 0.9850

**Analysis:**

	Class 1	Class 3
Precision	1.0	0.9714
Recall	0.9685	1.0

**Confusion Matrix:**

	Class 1	Class 2	Class 3
Class 1	206	348	18
Class 2	24	570	3
Class 3	0	0	613

Accuracy: 0.7794

F Measure: 0.7042

**Analysis:**

	Class 1	Class 2	Class 3
Precision	0.8956	0.6209	0.9668
Recall	0.3601	0.9547	1.0

Table 3. Confusion Matrix and Analysis for Case 1, Real World Data

### 3.2 Case 2: $\Sigma_i = \Sigma$

#### 3.2.1 Linear Separable Data

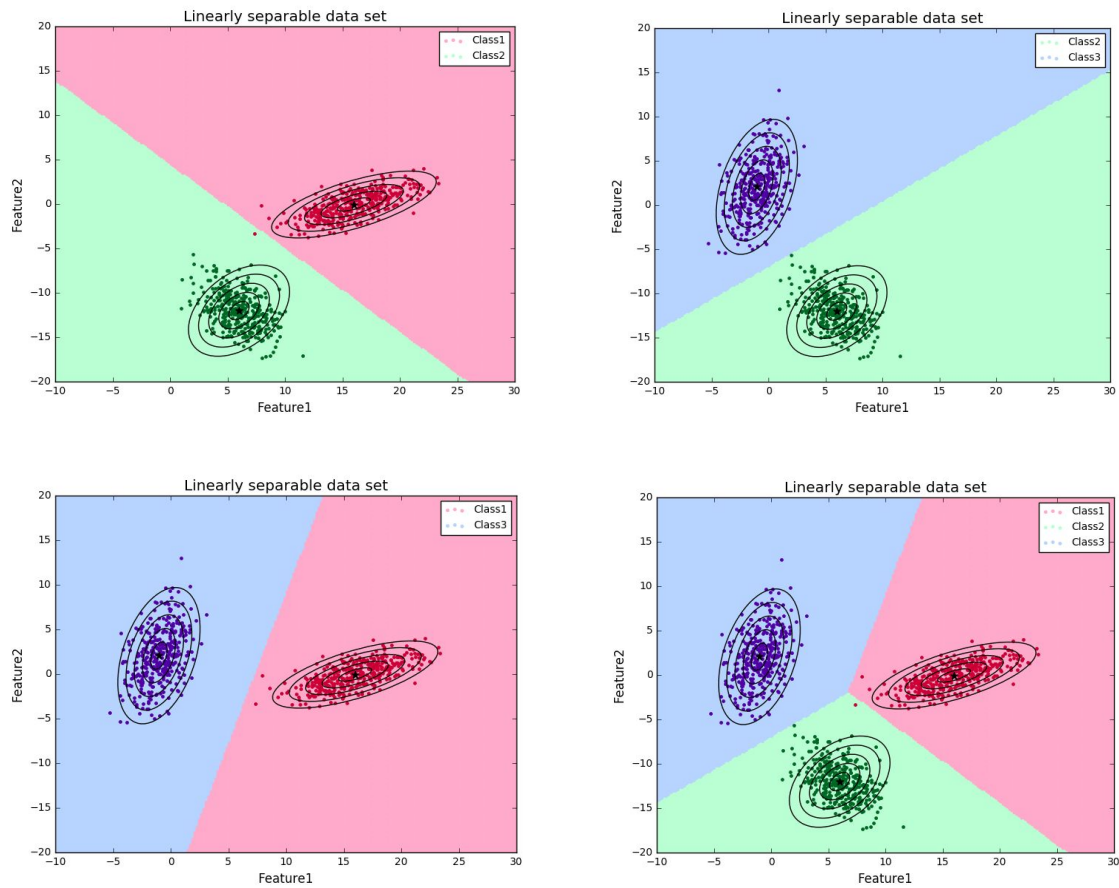


Fig4. Decision boundary for Case 2 Linear Separable Data depicting between : a) Class 1 and Class 2 b) Class 2 and Class 3 c) Class 1 and Class 3 d) Class 1,2 and 3

**Confusion Matrix:**

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

Accuracy: 1.0

F Measure: 1.0

**Analysis:**

	Class 1	Class 2
Precision	1.0	1.0
Recall	1.0	1.0

**Confusion Matrix:**

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

Accuracy: 1.0

F Measure: 1.0

**Analysis:**

	Class 1	Class 2
Precision	1.0	1.0
Recall	1.0	1.0

**Confusion Matrix:**

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

Accuracy: 1.0

F Measure: 1.0

**Analysis:**

	Class 1	Class 2
Precision	1.0	1.0
Recall	1.0	1.0

**Confusion Matrix:**

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

Accuracy: 1.0

F Measure: 1.0

**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 4. Confusion Matrix and Analysis for Case 2, Linear Separable Data

### 3.2.2 Non-Linear Data

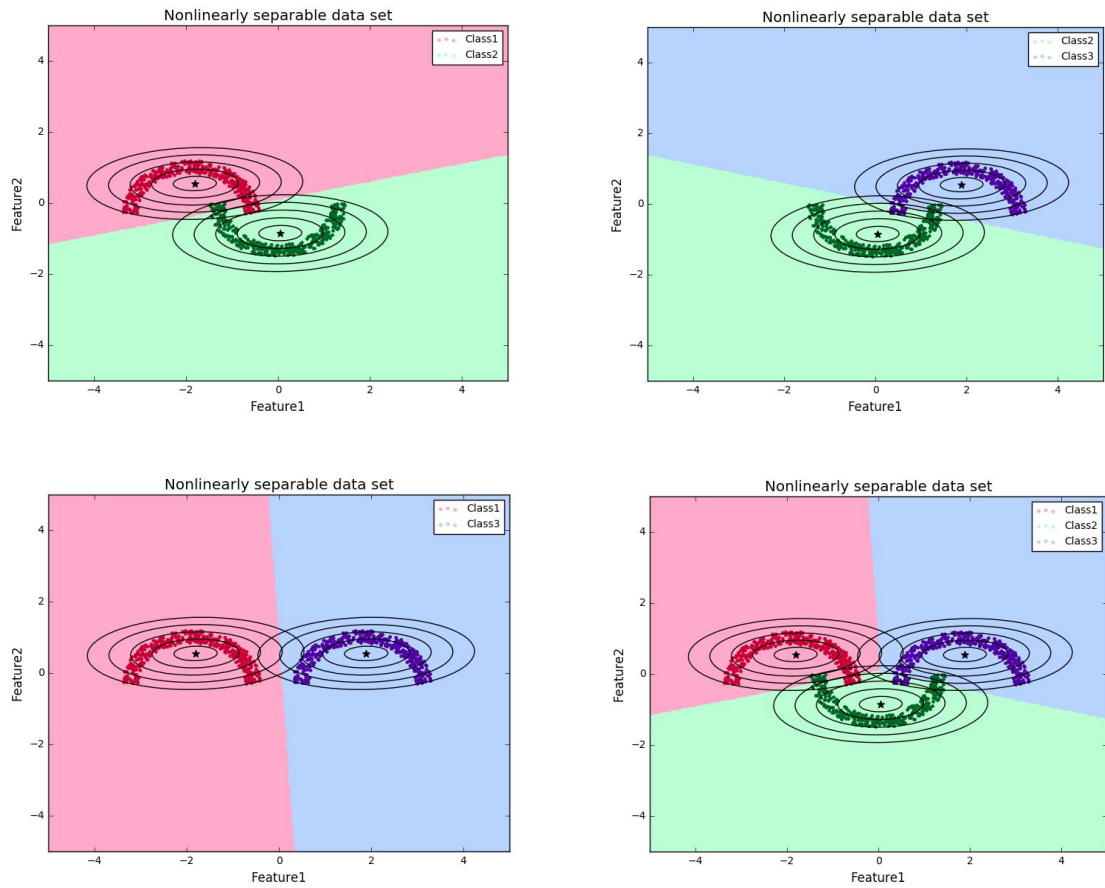


Fig5. Decision boundary for Case 2 Non-Linear Separable depicting between : a) Class 1 and Class 2 b) Class 2 and Class 3 c) Class 3 and Class 1 d) Class 1,2 and 3

**Confusion Matrix:**

	Class 1	Class 2
Class 1	115	10
Class 2	12	113

Accuracy: 0.9120

F Measure: 0.9120

**Analysis:**

	Class 1	Class 2
Precision	0.9055	0.9190
Recall	0.9200	0.9040

**Confusion Matrix:**

	Class 2	Class 3
Class 2	121	4
Class 3	4	121

Accuracy: 0.9680

F Measure: 0.9680

**Analysis:**

	Class 2	Class 3
Precision	0.9680	0.9680
Recall	0.9680	0.9680

**Confusion Matrix:**

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

Accuracy: 1.0

F Measure: 1.0

**Analysis:**

	Class 1	Class 3
Precision	1.0	1.0
Recall	1.0	1.0

**Confusion Matrix:**

	Class 1	Class 2	Class 3
Class 1	115	10	0
Class 2	12	109	4
Class 3	0	4	121

Accuracy: 0.9200

F Measure: 0.9199

**Analysis:**

	Class 1	Class 2	Class 3
Precision	0.9055	0.8861	0.9680
Recall	0.9200	0.8720	0.9680

Table 5. Confusion Matrix and Analysis for Case 2, Non- Linear Separable Data

### 3.2.3 Real World Data

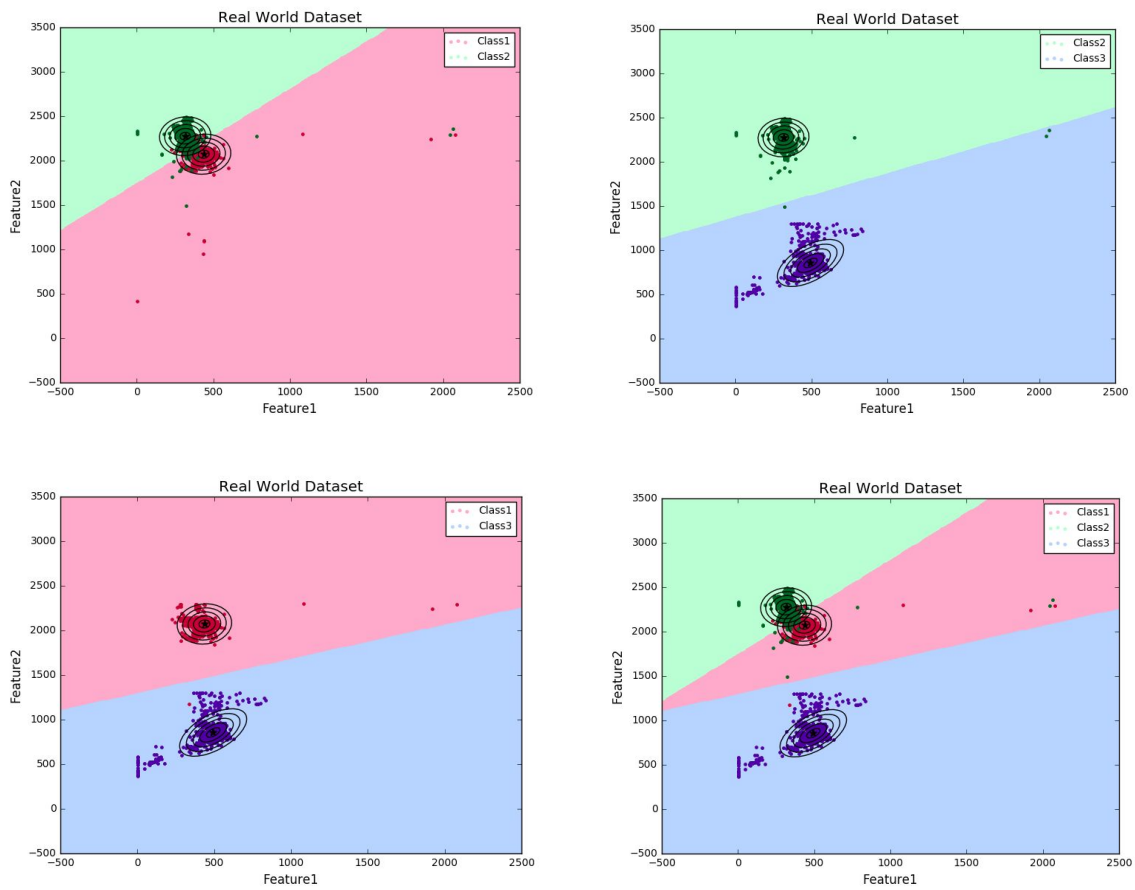


Fig6. Decision boundary for Case 2 Real World data depicting between : a) Class 1 and Class 2 b) Class 2 and Class 3 c) Class 3 and Class 1 d) Class 1,2 and 3

**Confusion Matrix:**

	Class 1	Class 2
Class 1	227	345
Class 2	26	571

Accuracy: 0.6826

F Measure: 0.7160

**Analysis:**

	Class 1	Class 2
Precision	0.8970	0.6233
Recall	0.3970	0.9560

**Confusion Matrix:**

	Class 2	Class 3
Class 2	586	11
Class 3	0	613

Accuracy: 0.9900

F Measure: 0.9900

**Analysis:**

	Class 2	Class 3
Precision	1.0	0.9820
Recall	0.9810	1.0

**Confusion Matrix:**

	Class 1	Class 3
Class 1	555	17
Class 3	0	613

Accuracy: 0.9860

F Measure: 0.9860

**Analysis:**

	Class 1	Class 3
Precision	1.0	0.9730
Recall	0.9700	1.0

**Confusion Matrix:**

	Class 1	Class 2	Class 3
Class 1	210	345	17
Class 2	23	571	3
Class 3	0	0	613

Accuracy: 0.7820

F Measure: 0.8070

**Analysis:**

	Class 1	Class 2	Class 3
Precision	0.9010	0.6230	1.0
Recall	0.3670	0.9560	1.0

Table 6. Confusion Matrix and Analysis for Case 2, Real World Data

### 3.3 Case 3: $\Sigma_i$ is a diagonal matrix

#### 3.3.1 Linear Separable Data

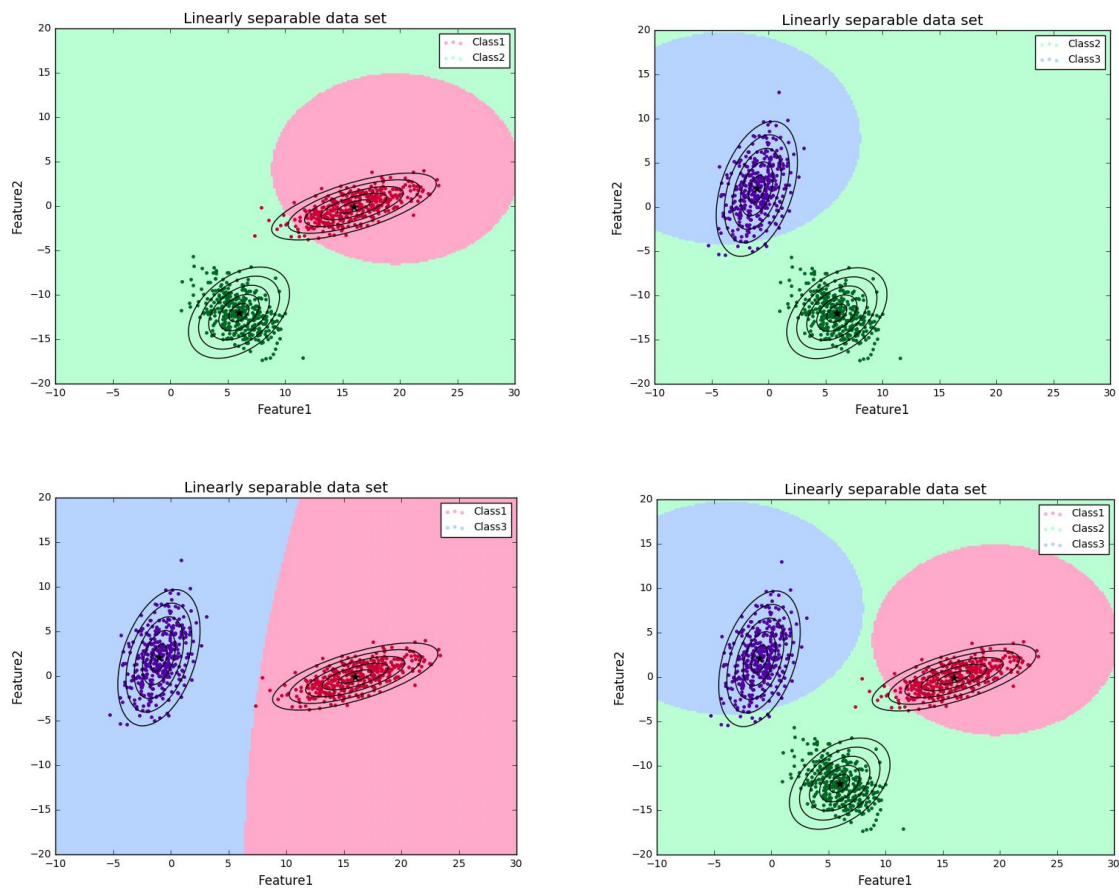


Fig7. Decision boundary for Case 3 Linear Separable data depicting between : a) Class 1 and Class 2 b) Class 2 and Class 3 c) Class 1 and Class 3 d) Class 1,2 and 3



**Confusion Matrix:**

	Class 1	Class 2
Class 1	119	6
Class 2	0	125

Accuracy: 0.9760  
F Measure: 0.9760

**Analysis:**

	Class 1	Class 2
Precision	1.0	0.9540
Recall	0.9520	1.0

**Confusion Matrix:**

	Class 2	Class 3
Class 2	125	0
Class 3	5	120

Accuracy: 0.9800  
F Measure: 0.9800

**Analysis:**

	Class 2	Class 3
Precision	0.9610	1.0
Recall	1.0	0.9600

**Confusion Matrix:**

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

Accuracy: 1.0  
F Measure: 1.0

**Analysis:**

	Class 1	Class 3
Precision	1.0	1.0
Recall	1.0	1.0

**Confusion Matrix:**

	Class 1	Class 2	Class 3
Class 1	119	6	0
Class 2	0	125	0
Class 3	0	5	120

Accuracy: 0.9700  
F Measure: 0.9650

**Analysis:**

	Class 1	Class 2	Class 3
Precision	1.0	0.9190	0.9600
Recall	0.9520	1.0	0.9600

Table 7. Confusion Matrix and Analysis for Case 3, Linear Separable Data

### 3.3.2 Non-Linear Data

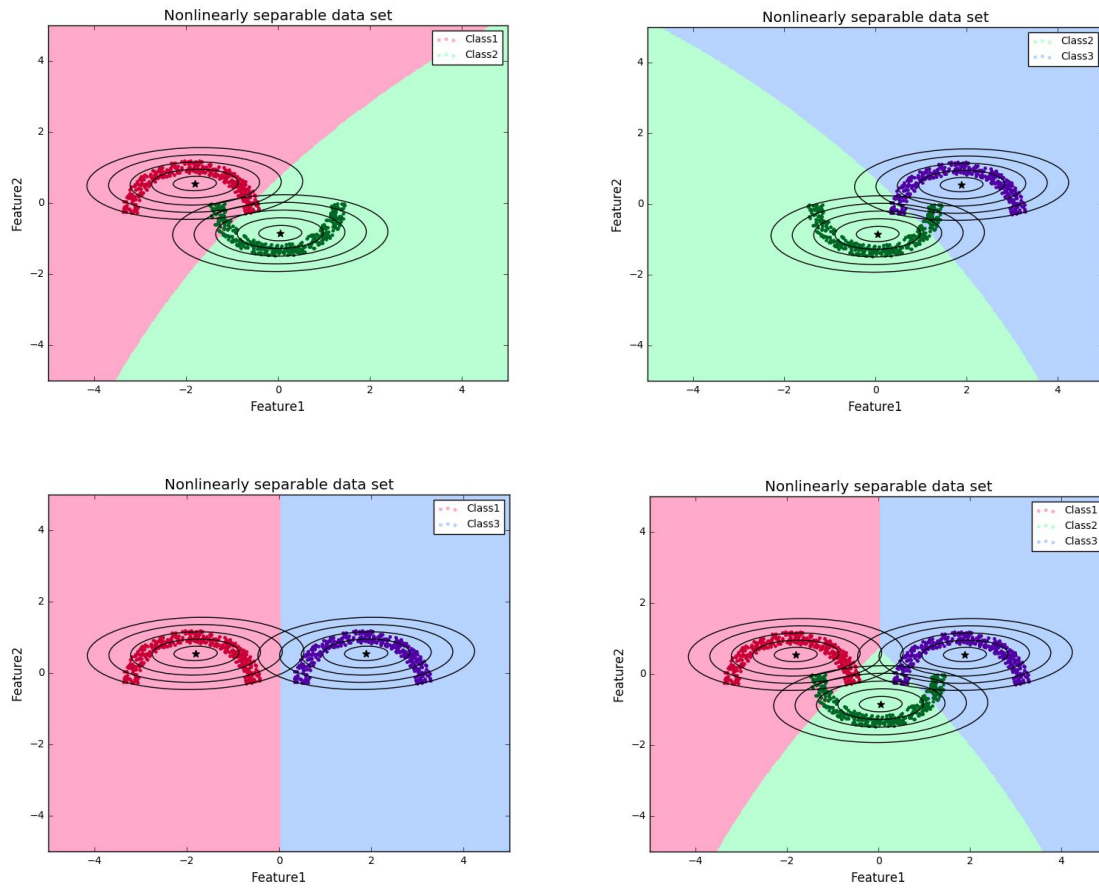


Fig8. Decision boundary for Case 3 Non-Linear Separable data depicting between : a) Class 1 and Class 2 b) Class 2 and Class 3 c) Class 3 and Class 1 d) Class 1,2 and 3

**Confusion Matrix:**

	Class 1	Class 2
Class 1	118	7
Class 2	26	99

Accuracy: 0.8680

F Measure: 0.8720

**Analysis:**

	Class 1	Class 2
Precision	0.8190	0.9330
Recall	0.9440	0.7920

**Confusion Matrix:**

	Class 2	Class3
Class 2	103	22
Class 3	7	118

Accuracy: 0.8840

F Measure: 0.8870

**Analysis:**

	Class 2	Class 3
Precision	0.9360	0.8430
Recall	0.8240	0.9440

**Confusion Matrix:**

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

Accuracy: 1.0

F Measure: 1.0

**Analysis:**

	Class 1	Class 3
Precision	1.0	1.0
Recall	1.0	1.0

**Confusion Matrix:**

	Class 1	Class 2	Class 3
Class 1	118	7	0
Class 2	26	77	22
Class 3	0	7	118

Accuracy: 0.8350

F Measure: 0.8520

**Analysis:**

	Class 1	Class 2	Class 3
Precision	0.8190	0.8460	0.9440
Recall	0.9440	0.6160	0.9440

Table 8. Confusion Matrix and Analysis for Case 3, Non-Linear Separable Data

### 3.3.3 Real World Data

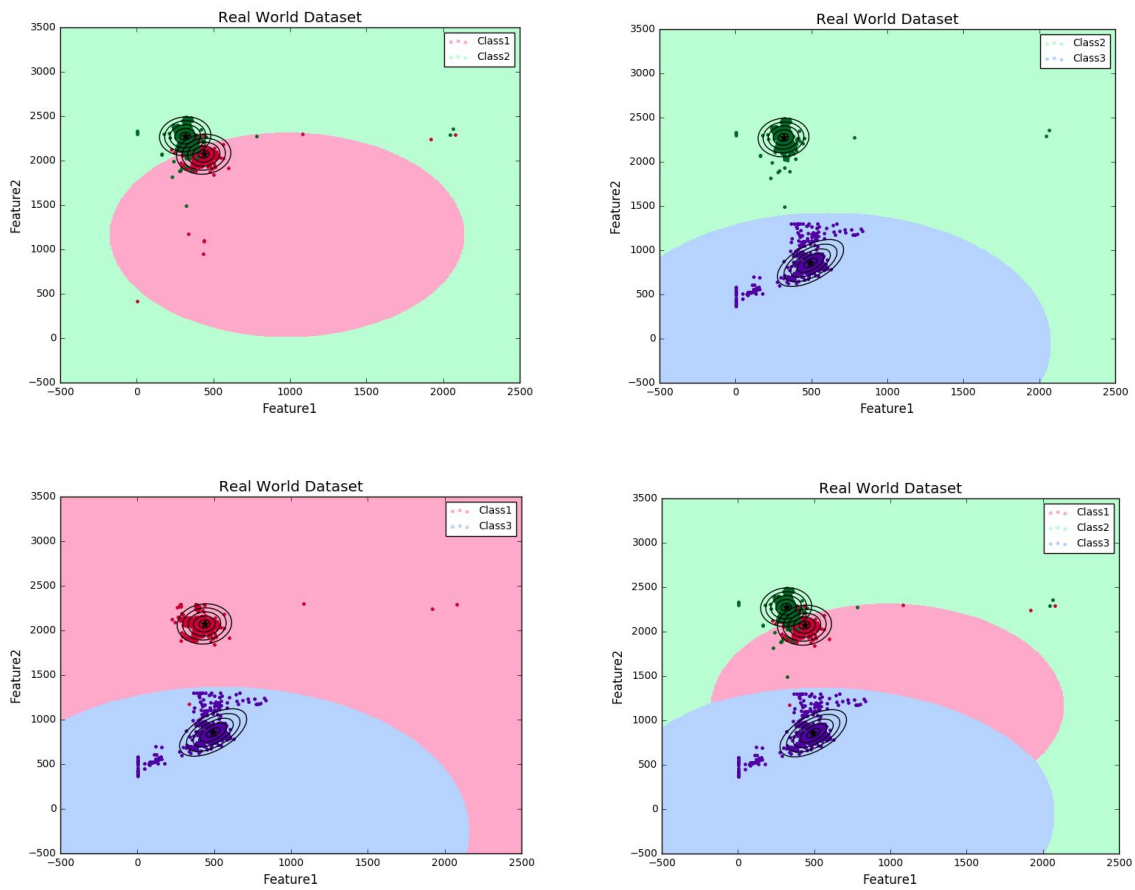


Fig9. Decision boundary for Case 3 Real World data depicting between : a) Class 1 and Class 2 b) Class 2 and Class 3 c) Class 3 and Class 1 d) Class 1,2 and 3

**Confusion Matrix:**

	Class 1	Class 2
Class 1	215	357
Class 2	19	578

Accuracy: 0.6783

F Measure: 0.7170

**Analysis:**

	Class 1	Class 2
Precision	0.9188	0.6181
Recall	0.3758	0.9681

**Confusion Matrix:**

	Class 2	Class 3
Class 2	594	3
Class 3	1	612

Accuracy: 0.9966

F Measure: 0.9966

**Analysis:**

	Class 2	Class 3
Precision	0.9983	0.9951
Recall	0.9949	0.9983

**Confusion Matrix:**

	Class 1	Class 3
Class 1	555	17
Class 3	0	613

Accuracy: 0.9856

F Measure: 0.9858

**Analysis:**

	Class 1	Class 3
Precision	1.0	0.9730
Recall	0.9702	1.0

**Confusion Matrix:**

	Class 1	Class 2	Class 3
Class 1	205	352	15
Class 2	17	577	3
Class 3	1	0	612

Accuracy: 0.7820

F Measure: 0.8090

**Analysis:**

	Class 1	Class 2	Class 3
Precision	0.919	0.621	0.998
Recall	0.358	0.966	0.998

Table 9. Confusion Matrix and Analysis for Case 3, Real World Data

### 3.4 Case 4: $\Sigma_i$ is unique

#### 3.4.1 Linear Separable Data

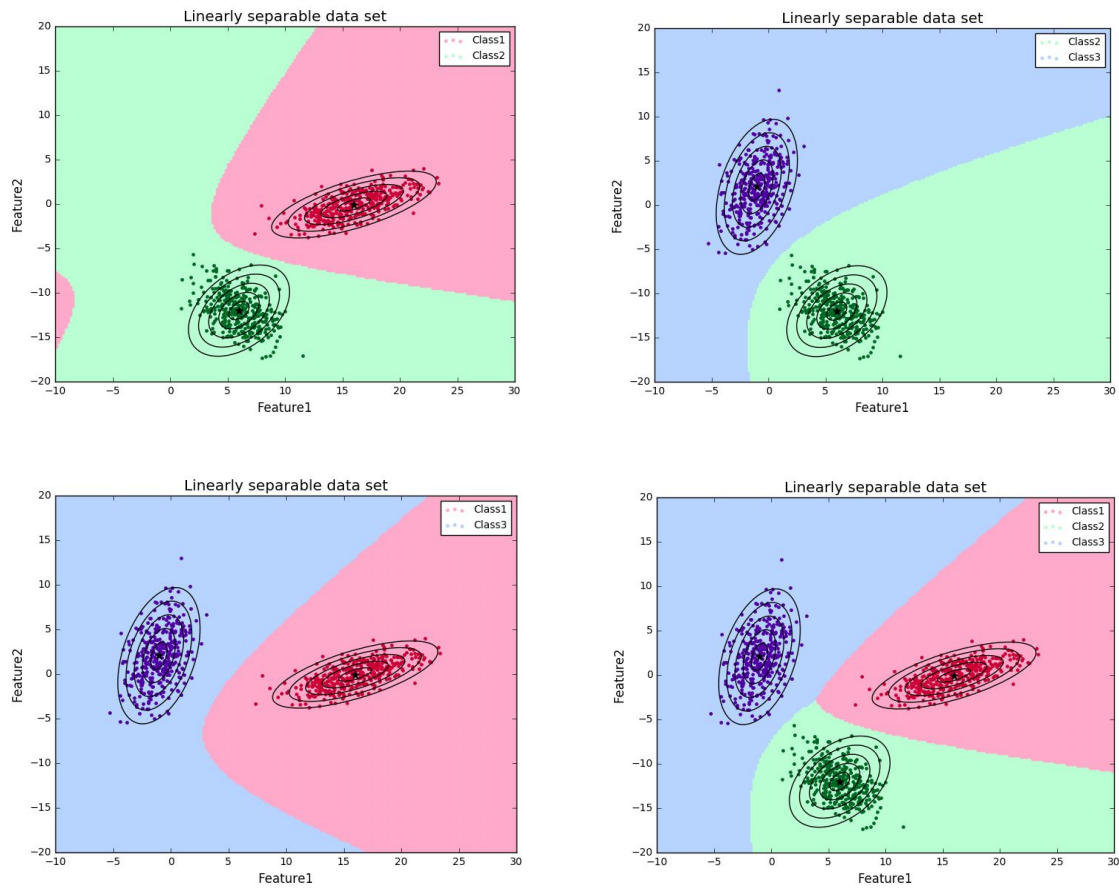


Fig10. Decision boundary for Case 4 Linear Separable data depicting between : a) Class 1 and Class 2 b) Class 2 and Class 3 c) Class 1 and Class 3 d) Class 1,2 and 3

**Confusion Matrix:**

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

Accuracy: 1.0

F Measure: 1.0

**Analysis:**

	Class 1	Class 2
Precision	1.0	1.0
Recall	1.0	1.0

**Confusion Matrix:**

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

Accuracy: 1.0

F Measure: 1.0

**Analysis:**

	Class 2	Class 3
Precision	1.0	1.0
Recall	1.0	1.0

**Confusion Matrix:**

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

Accuracy: 1.0

F Measure: 1.0

**Analysis:**

	Class 1	Class 3
Precision	1.0	1.0
Recall	1.0	1.0

**Confusion Matrix:**

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

Accuracy: 1.0

F Measure: 1.0

**Analysis:**

	Class 1	Class 2	Class 3
Precision	1.0	1.0	1.0
Recall	1.0	1.0	1.0

Table 10. Confusion Matrix and Analysis for Case 4, Linear Separable Data

### 3.4.2 Non-Linear Data

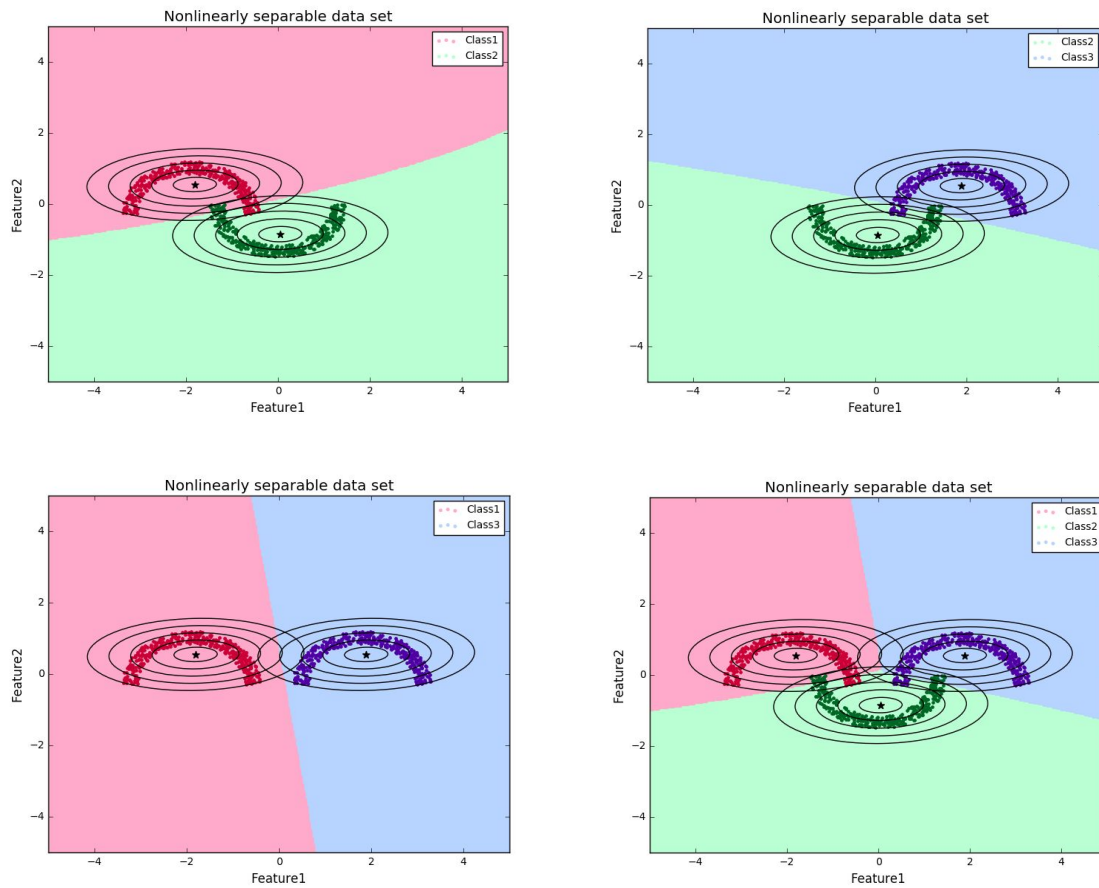


Fig11. Decision boundary for Case 4 Non-Linear Separable data depicting between : a) Class 1 and Class 2 b) Class 2 and Class 3 c) Class 3 and Class 1 d) Class 1,2 and 3



**Confusion Matrix:**

	Class 1	Class 2
Class 1	114	11
Class 2	12	113

Accuracy: 0.9080

F Measure: 0.9080

**Analysis:**

	Class 1	Class 2
Precision	0.9047	0.9112
Recall	0.9120	0.9040

**Confusion Matrix:**

	Class 2	Class 3
Class 2	103	22
Class 3	7	118

Accuracy: 0.8840

F Measure: 0.8867

**Analysis:**

	Class 2	Class 3
Precision	0.9363	0.8428
Recall	0.8240	0.9440

**Confusion Matrix:**

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

Accuracy: 1.0

F Measure: 1.0

**Analysis:**

	Class 1	Class 3
Precision	1.0	1.0
Recall	1.0	1.0

**Confusion Matrix:**

	Class 1	Class 2	Class 3
Class 1	114	11	0
Class 2	12	110	3
Class 3	0	7	118

Accuracy: 0.9120

F Measure: 0.8889

**Analysis:**

	Class 1	Class 2	Class 3
Precision	0.9047	0.8593	0.9752
Recall	0.9120	0.8800	0.9440

Table 11. Confusion Matrix and Analysis for Case 4, Non-Linear Separable Data

### 3.4.3 Real World Data

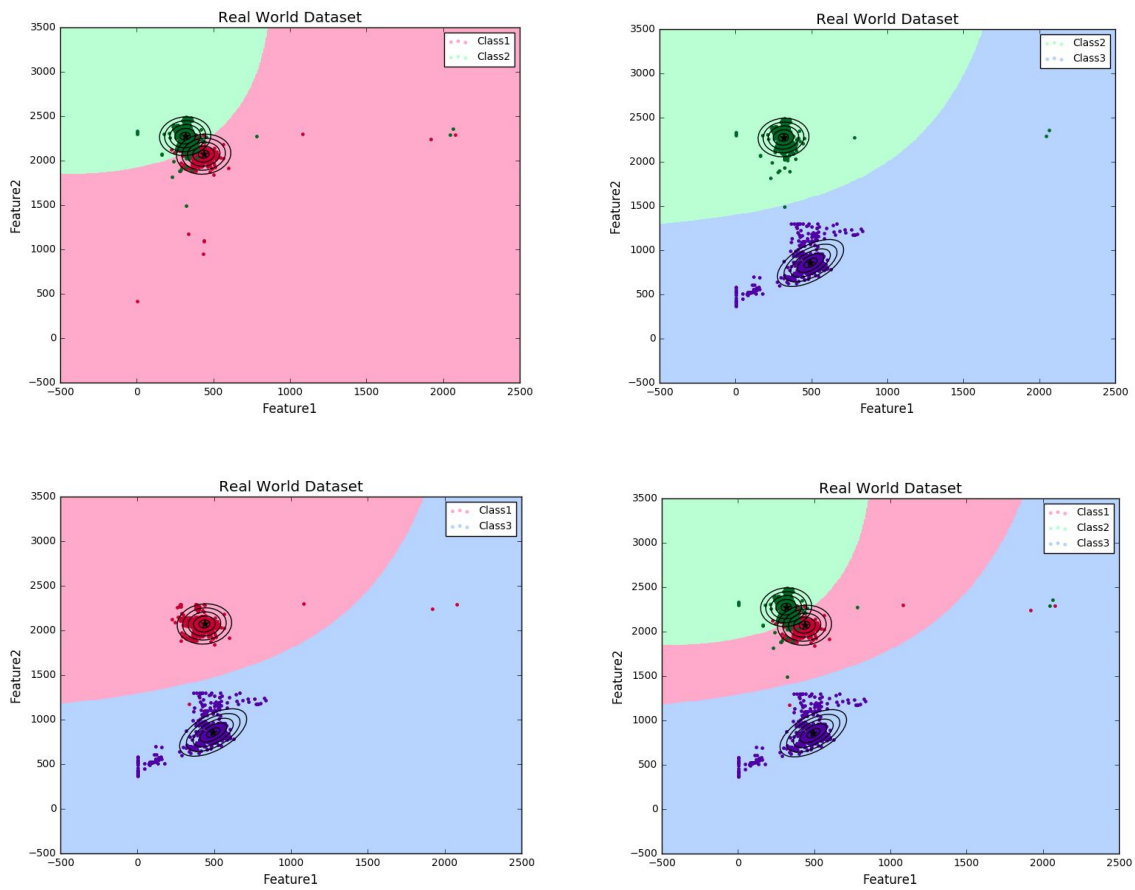


Fig12. Decision boundary for Case 4 Real World data depicting between : a) Class 1 and Class 2 b) Class 2 and Class 3 c) Class 3 and Class 1 d) Class 1,2 and 3

**Confusion Matrix:**

	Class 1	Class 2
Class 1	227	345
Class 2	26	571

Accuracy: 0.6826

F Measure: 0.7160

**Analysis:**

	Class 1	Class 2
Precision	0.8972	0.6233
Recall	0.3968	0.9564

**Confusion Matrix:**

	Class 2	Class 3
Class 2	581	16
Class 3	0	613

Accuracy: 0.9839

F Measure: 0.9869

**Analysis:**

	Class 2	Class 3
Precision	1.0	0.9745
Recall	0.9731	1.0

**Confusion Matrix:**

	Class 1	Class 3
Class 1	553	19
Class 3	0	613

Accuracy: 0.9839

F Measure: 0.9841

**Analysis:**

	Class 1	Class 3
Precision	1.0	0.9699
Recall	0.9667	1.0

**Confusion Matrix:**

	Class 1	Class 2	Class 3
Class 1	208	345	19
Class 2	11	571	15
Class 3	0	0	613

Accuracy: 0.7811

F Measure: 0.7177

**Analysis:**

	Class 1	Class 2	Class 3
Precision	0.9497	0.6233	0.9474
Recall	0.3636	0.9564	1.0

Table 12. Confusion Matrix and Analysis for Case 4, Real World Data

## 4. Conclusion

1. On the given data, Bayes Classifier works well but only for linearly separable data but was not so good with non-linearly separable data.
2. As the data is scattered and randomly distributed in real world data, that's why the accuracies came out to be not very good.
3. As from the theory the decision boundary successfully came out to be linear in first two cases whereas in the other two cases it was hyper-quadratic.
4. The accuracies came out to be the least for real world data in some of the cases as the data was completely scattered and randomly distributed.