Pattern Recognition

Bayes classifier

Group: 10

September 13, 2018

Team members:

1.Randheer kumar (B16139)

2.Dhrubodeep Basumatary (B16017)

3.Dilip Kumar Chauhan (B16018)

Contents:

Objective

To Classify the following dataset on the basis of Bayes Classifier:

- 2D dataset-1 , (Artificial)
  1. Linearly separable dataset-1a
  2. Non-linearly separable dataset-1b
- 2D dataset-2 , Real world
- Plot decision region of all pair of combination of classes and all the three classes together
- Contour plot of all pair of combination of classes and all the three classes together
- Calculate Accuracy, Precision,Recall, mean Precision,mean Recall,F-measure, mean F-measure and Confusion Matrix

Procedure

- Data for each class is partitioned into two text files 75% for training and 25% for testing.
- It is assumed that dataset for each class follow Gaussian-distribution
- For case1: We assumed covariance matrix Σ is same for all classes, and we forcefully make it's off diagonal terms 0 for further calculations. Σ is taken as a mean covariance matrix.
- For case 2: We assumed covariance matrix Σ is same for all classes, Σ is taken as a mean covariance matrix.
- For case 3: Σ is diagonal matrix and different for all classes. we forcefully make it's off diagonal terms 0 for further calculations.
- For case 4: Σ is different for all classes.
- Based on above assumptions, the discriminant function (g_i (x)) was calculated for each class and decision region & Contour were plotted.

Observations:

3.1 Linearly separable data set :

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

Confusion Matrix

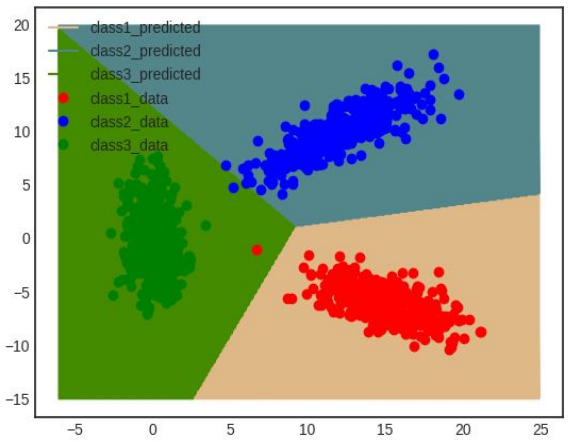
	Class1	**Class2**	**Class3**
Class1	125	0	0
Class2	0	125	0
Class3	0	0	125

Analysis

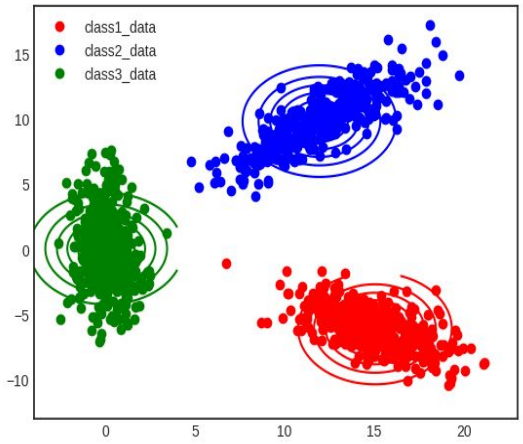
	Class1	**Class2**	**Class3**	**Mean**
Precision	1	1	1	1
Recall	1	1	1	1
F-measure	1	1	1	1

Accuracy : 100%

Class_1 vs Class_2 vs Class_3

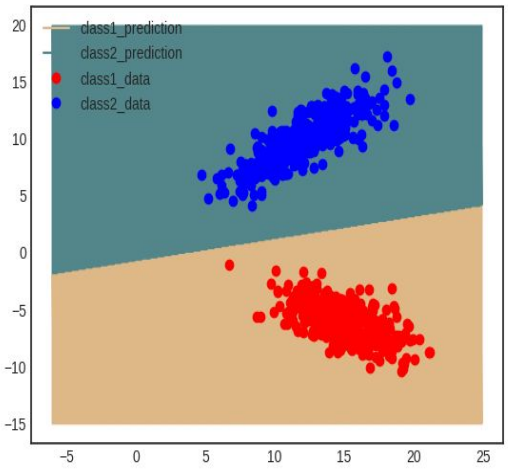


Decision Region

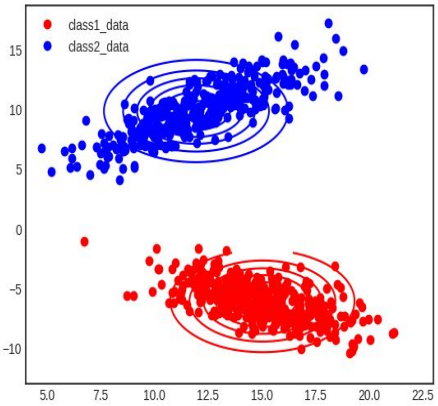


Contour Plot

Class_1 vs Class_2

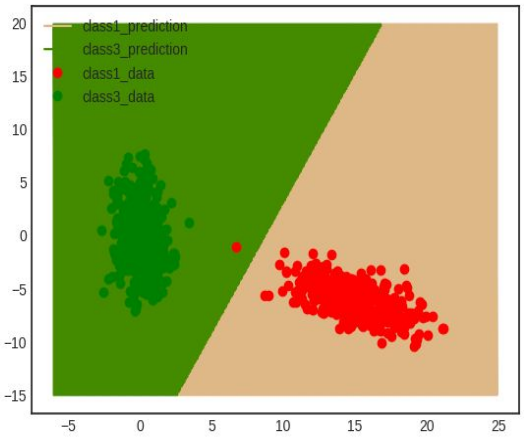


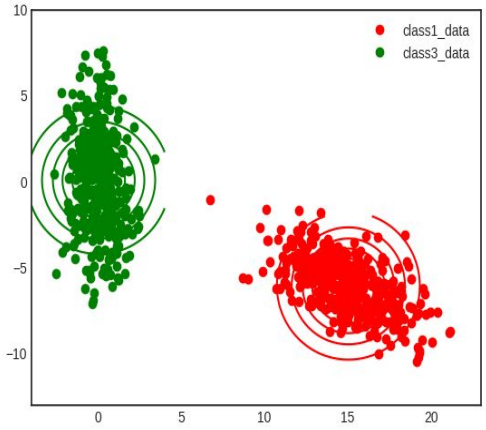
Decision Region



Contour Plot

Class_1 vs Class_3

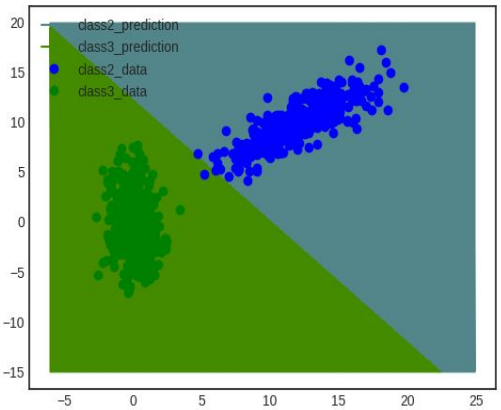


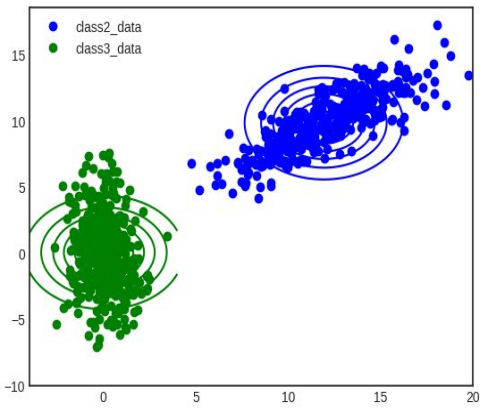


`

Decision region **Contour plot**

Class_2 vs Class_3





Decision region **Contour plot**

3.1.2 Case 2: Σ_i =Σ

Confusion Matrix

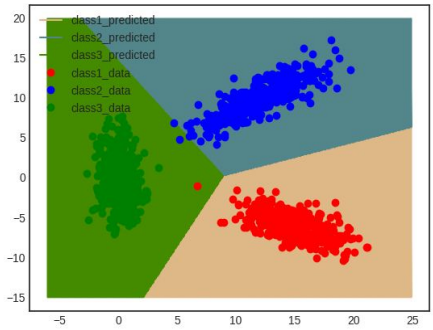
	Class1	**Class2**	**Class3**
Class1	125	0	0
Class2	0	123	2
Class3	0	0	125

Analysis

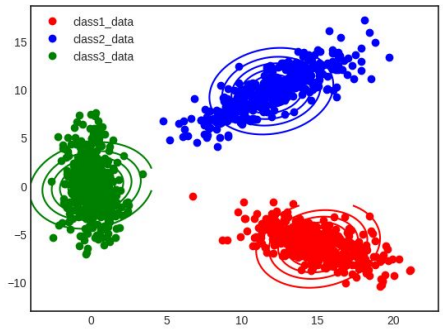
	Class1	**Class2**	**Class3**	**Mean**
Precision	1	1	0.9842	0.9947
Recall	1	0.984	1	0.9947
F-measure	1	0.99193	0.9920	0.9947

Accuracy : 99.4%

Class_1 vs Class_2 vs Class_3

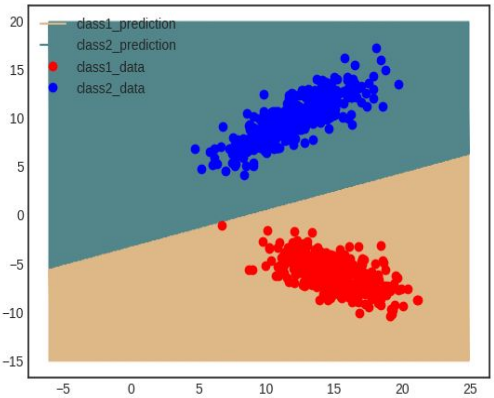


Decision region

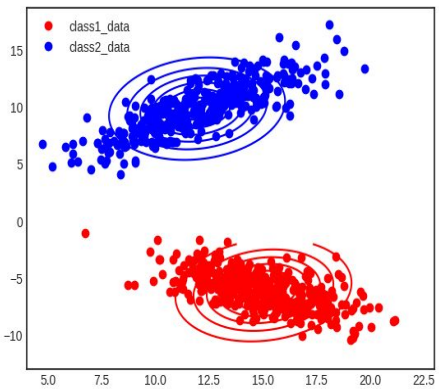


Contour plot

Class_1 vs Class_2

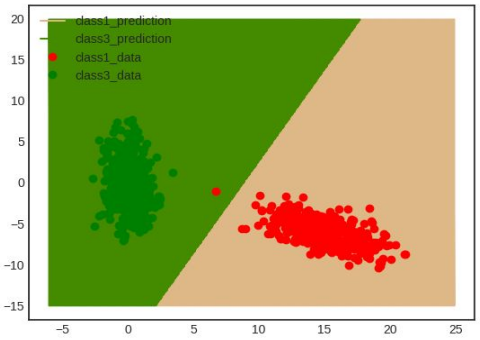


Decision region

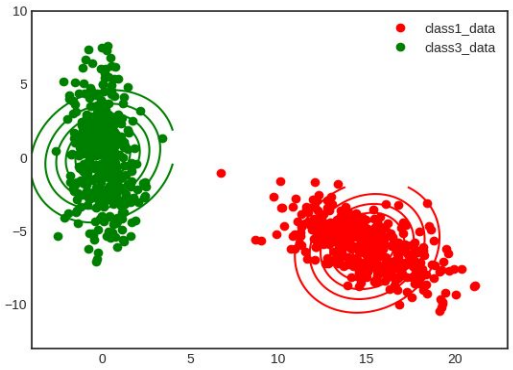


Contour plot

Class_1 vs Class_3

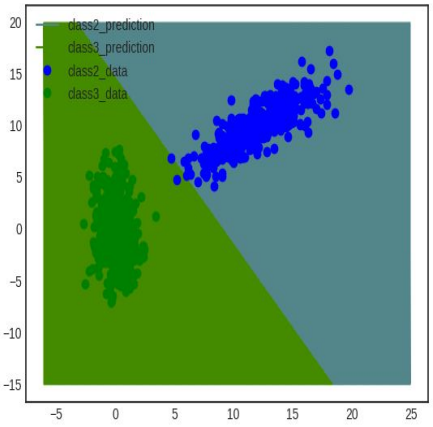


Decision region

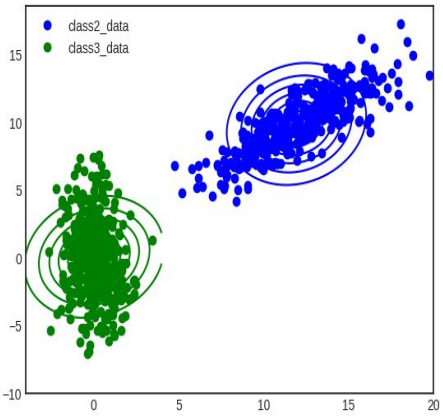


Contour plot

Class_2 vs Class_3



Decision region



Contour plot

3.1.3 Case 3: Σ_i is diagonal matrix

Confusion Matrix

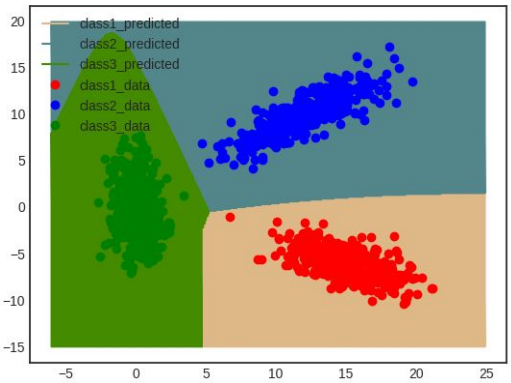
	Class1	**Class2**	**Class3**
class1	125	0	0
class2	0	125	0
class3	0	0	125

Analysis

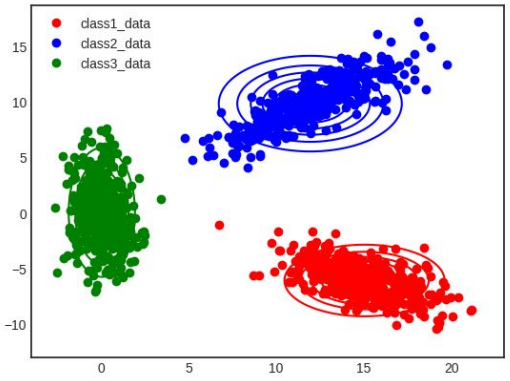
	Class1	**Class2**	**Class3**	**Mean**
Precision	1	1	1	1
Recall	1	1	1	1
F-measure	1	1	1	1

Accuracy : 100%

Class_1 vs Class_2 vs Class_3

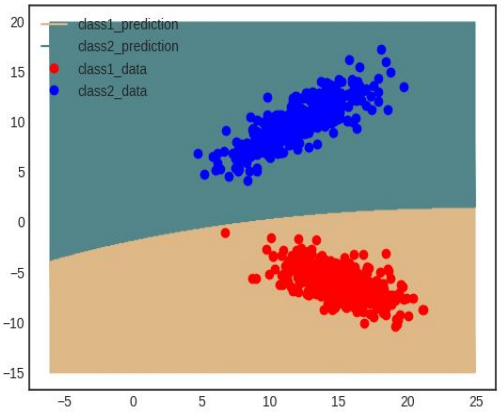


Decision region

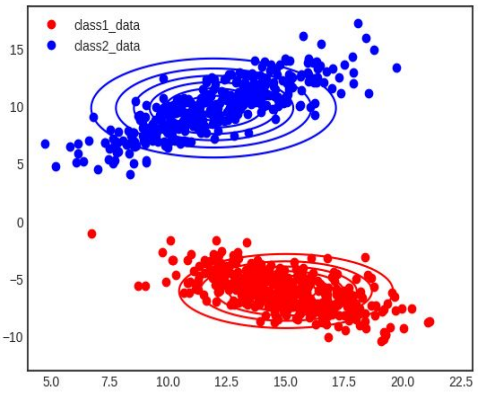


Contour plot

Class_1 vs Class_2

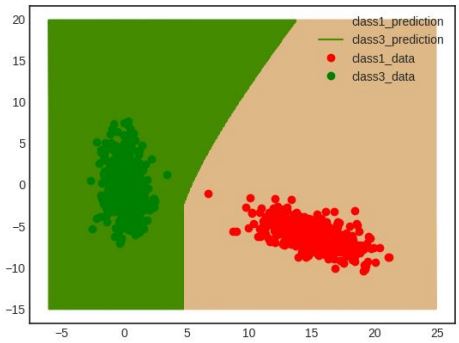


Decision region

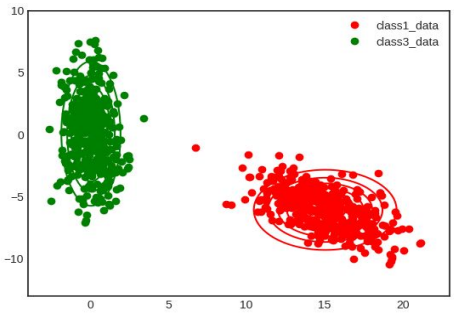


Contour plot

Class_1 vs Class_3

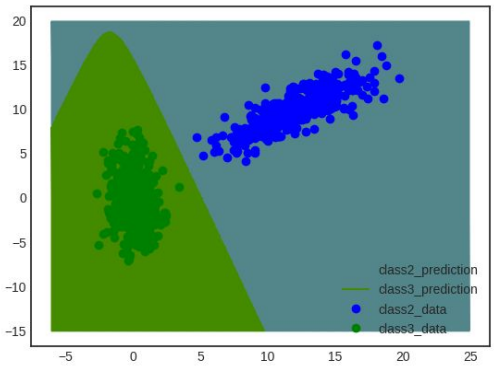


Decision region

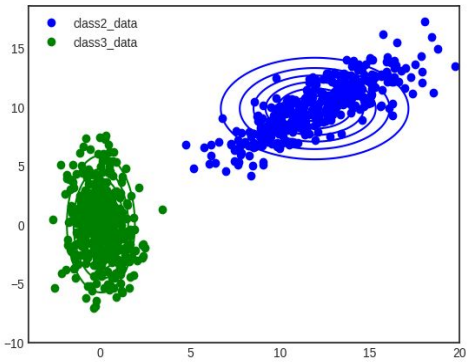


Contour plot

Class_2 vs Class_3



Decision region



Contour plot

3.1.4 Case 4 : Σ_i is unique

Confusion Matrix for data set 1a,

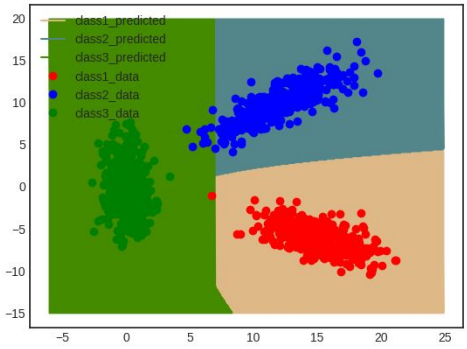
	Class1	**class2**	**class3**
class1	125	0	0
class2	0	125	0
class3	0	0	125

Analysis

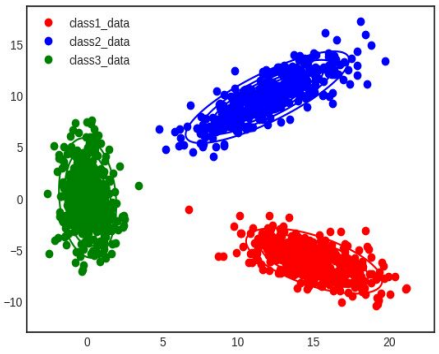
	Class1	**Class2**	**Class3**	**Mean**
Precision	1	1	1	1
Recall	1	1	1	1
F-measure	1	1	1	1

Accuracy : 100%

Class_1 vs Class_2 vs Class_3

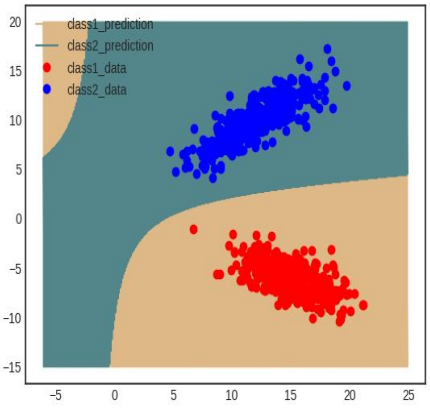


Decision region

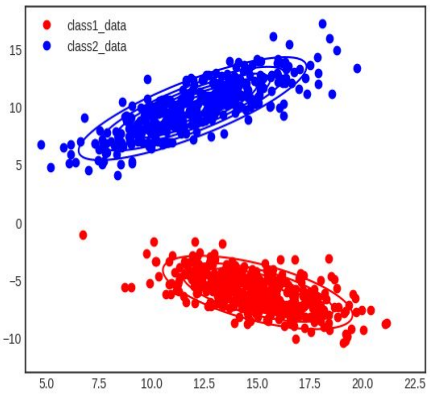


Contour plot

Class_1 vs Class_2

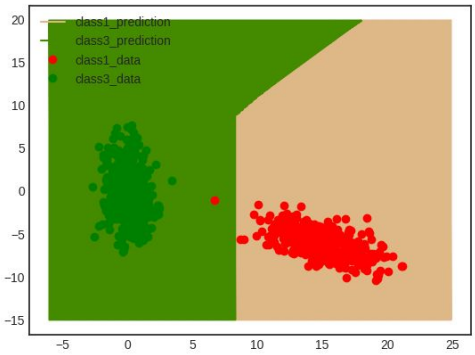


Decision region

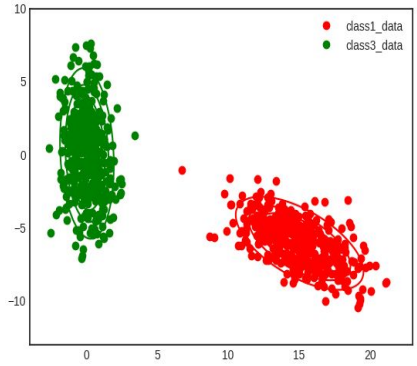


Contour plot

Class_1 vs Class_3

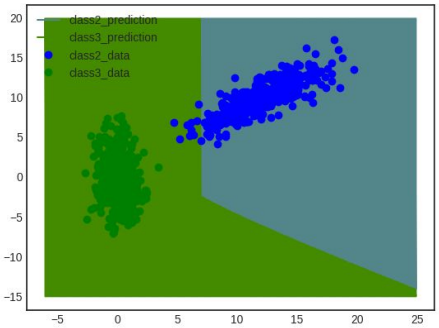


Decision region

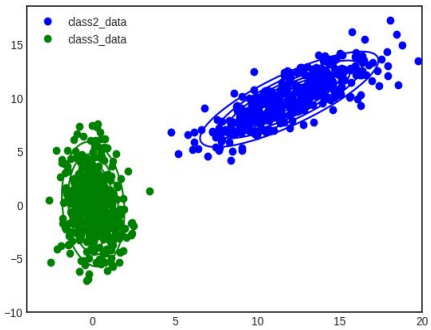


Contour plot

Class_2 vs Class_3



Decision region



Contour plot

3.1.5. Inference:

- For case_1: It can be seen that the decision region is linear and the corresponding contours of every class is close to circular .It works fine as the data is linearly separable
- For case_2: It can be seen that the decision region is linear and the corresponding contours of every class is elliptical .It works fine as the data is linearly separable

- For case_3: It can be inferred from the plot that the decision region is non-linear and the contours of every class is elliptical & its major and minor axes are parallel to coordinate axes
- For case_4: It can be seen that the decision region is non- linear and the contours of different classes are elliptical and aligned in different directions

3.2 Nonlinearly separable data set:

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

Confusion Matrix

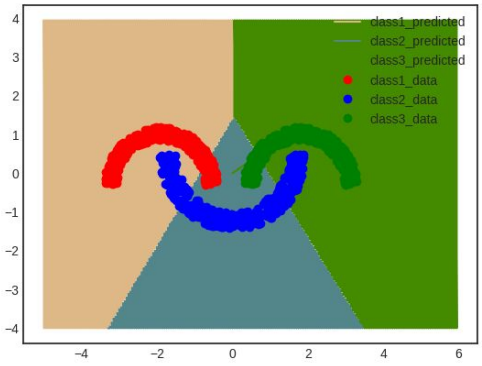
	Class1	**class2**	**class3**
class1	106	19	0
class2	34	65	25
class3	0	17	108

Analysis

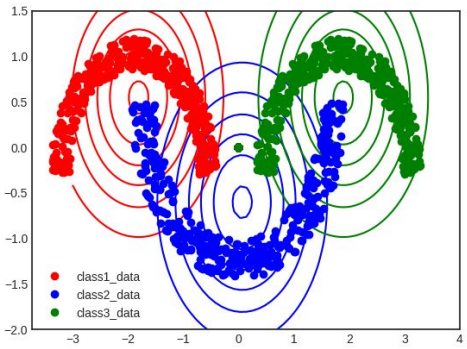
	Class1	**Class2**	**Class3**	**Mean**
Precision	0.7571	0.6435	0.8120	0.7374
Recall	0.848	0.5241	0.864	0.7454
F-measure	0.7999	0.5776	0.8365	0.7414

Accuracy: 74.59%

Class_1 vs Class_2 vs Class_3

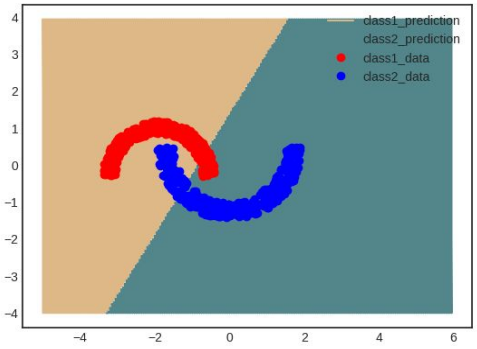


Decision region

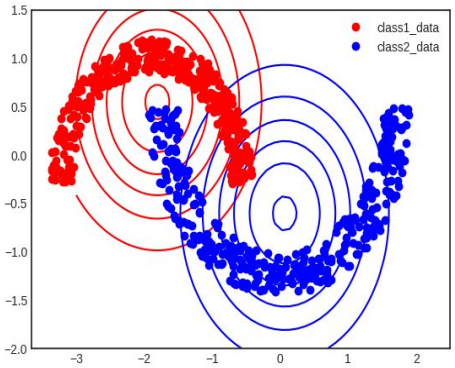


Contour plot

Class_1 vs Class_2

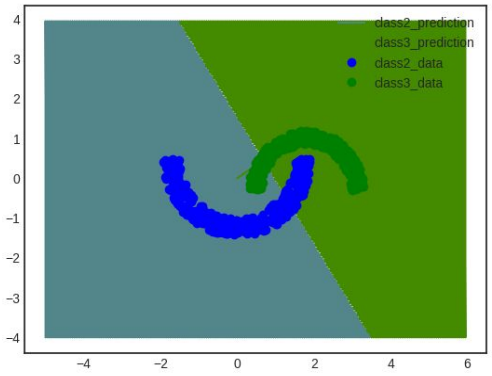


Decision Region

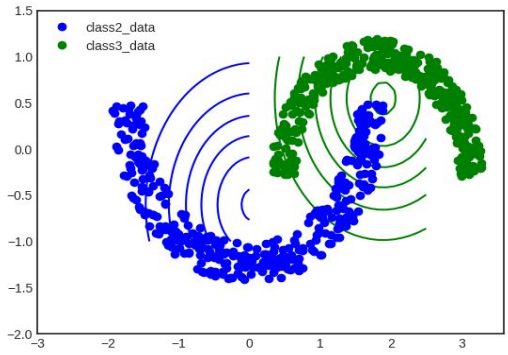


Contour plot

Class_1 vs Class_3

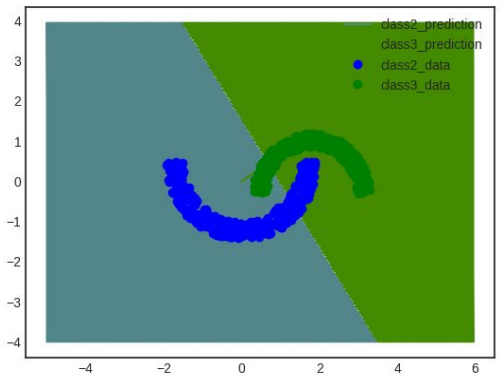


Decision region

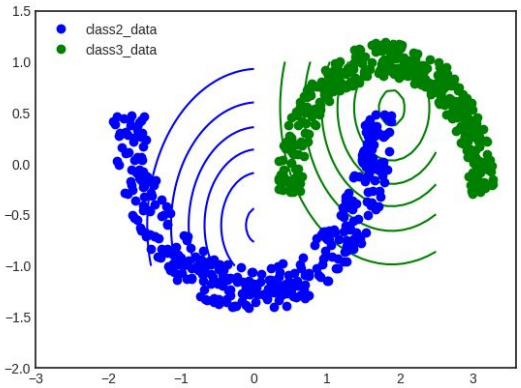


Contour plot

Class_2 vs Class_3



Decision region



Contour plot

3.2.2 Case 2: Σ_i =Σ

Confusion Matrix

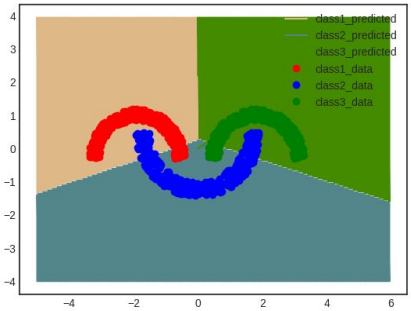
	Class1	**Class2**	**Class3**
Class1	113	12	0
Class2	21	83	21
Class3	0	8	117

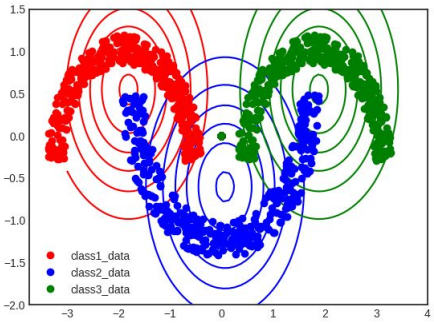
Analysis

	Class1	**Class2**	**Class3**	**Mean**
Precision	0.8432	0.8058	0.8666	0.8358
Recall	0.904	0.664	0.936	0.8347
F-measure	0.8751	0.7280	0.9002	0.8352

Accuracy : 83.4%

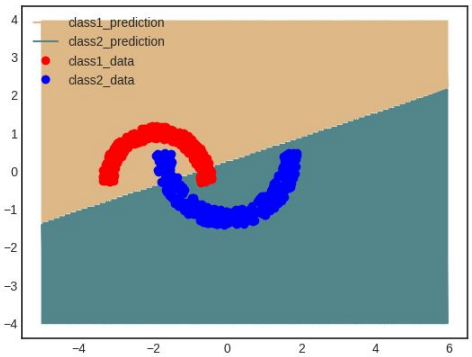
Class_1 vs Class_2 vs Class_3

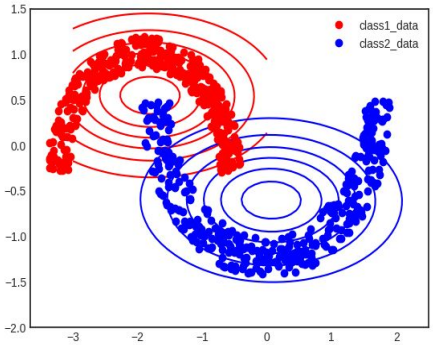




Decision region **Contour plot**

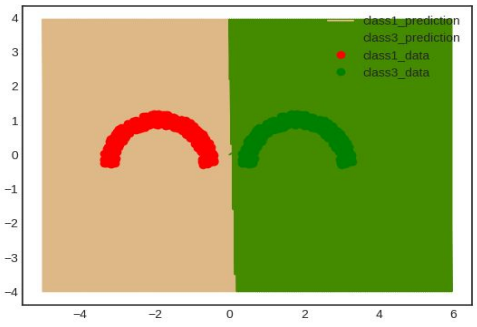
Class_1 vs Class_2



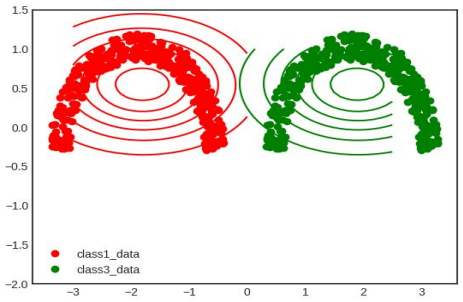


Decision region **Contour plot**

Class_1 vs Class_3

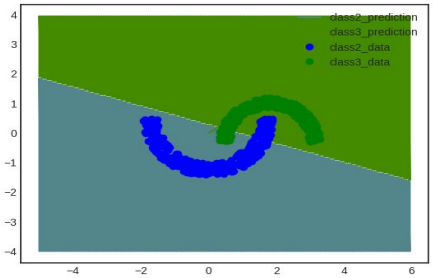


Decision region

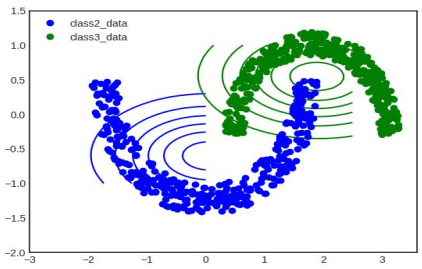


Contour plot

Class_2 vs Class_3



Decision region



Contour plot

3.2.3 Case 3: Σ_i is diagonal matrix

Confusion Matrix

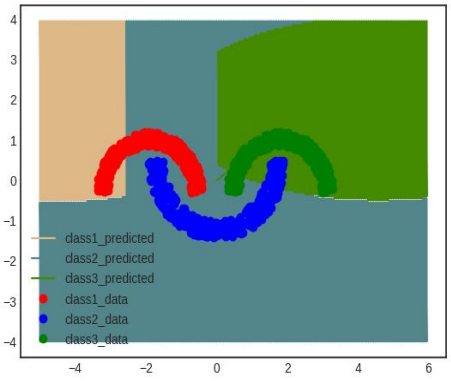
	Class1	**class2**	**class3**
Class1	112	13	0
Class2	21	84	20
Class3	0	8	117

Analysis

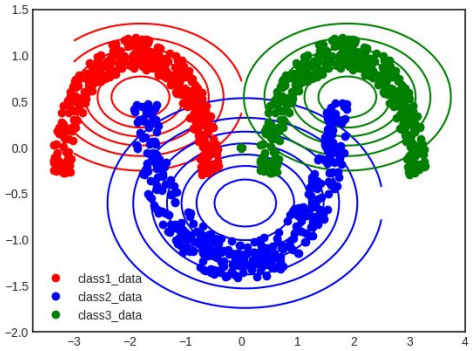
	Class1	**Class2**	**Class3**	**mean**
Precision	0.8421	0.800	0.8540	0.8320
Recall	0.896	0.672	0.936	0.8347
F-measure	0.8682	0.625	0.8931	0.8307

Accuracy : 83.46%

Class_1 vs Class 2 vs Class_3

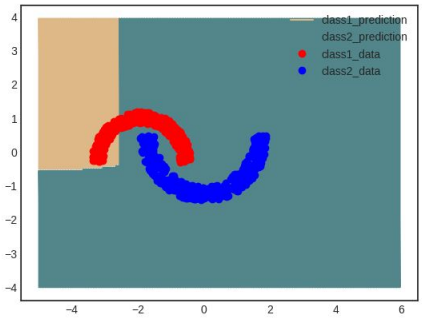


Decision region

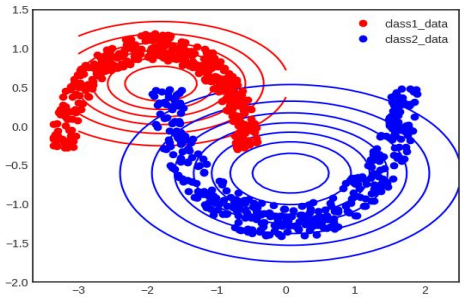


Contour plot

Class_1 vs Class_2

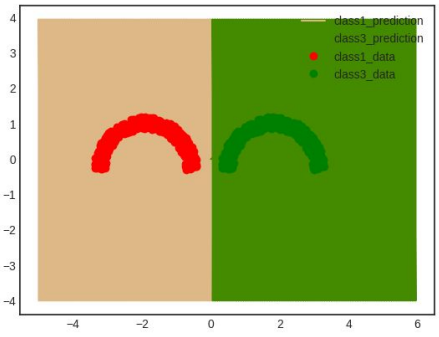


Decision region

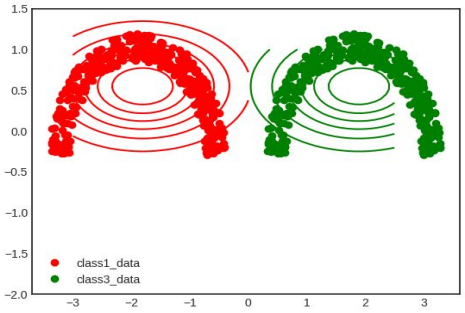


Contour plot

Class_1 vs Class_3

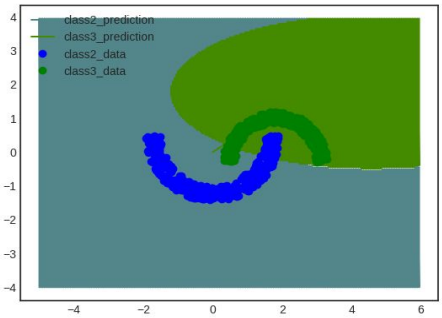


Decision region

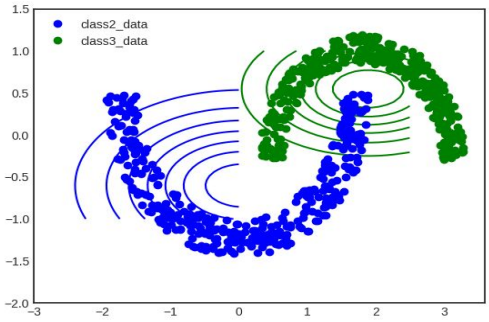


Contour plot

Class_2 vs Class_3



Decision region



Contour plot

3.2.4 Case 4: Σ_i is unique

Confusion Matrix

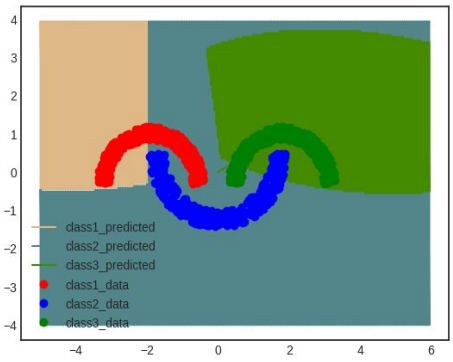
	Class1	**class2**	**class3**
class1	113	12	0
class2	21	84	20
class3	0	9	116

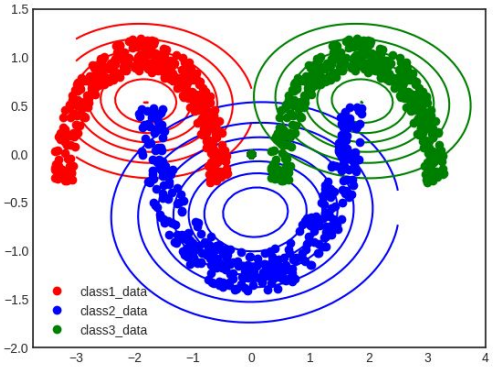
Analysis

	Class1	**Class2**	**Class3**	**Mean**
Precision	0.8432	0.800	0.8529	0.8320
Recall	0.904	0.672	0.928	0.8346
F-measure	0.8731	0.7304	0.888	0.8333

Accuracy : 83.46%

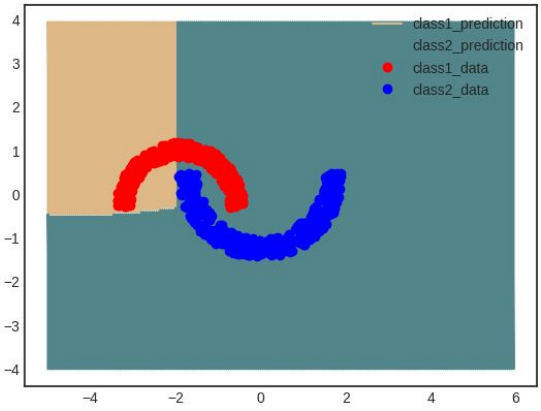
Class_1 vs Class_2 vs Class_3

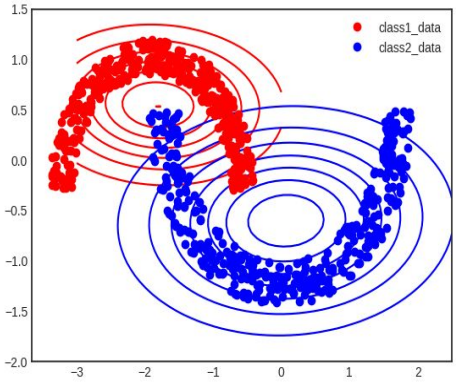




Decision region **Contour plot**

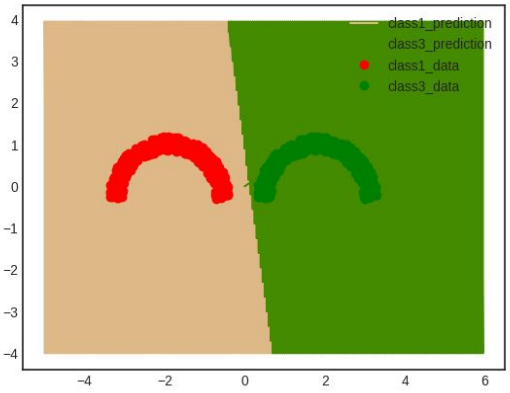
Class_1 vs Class_2

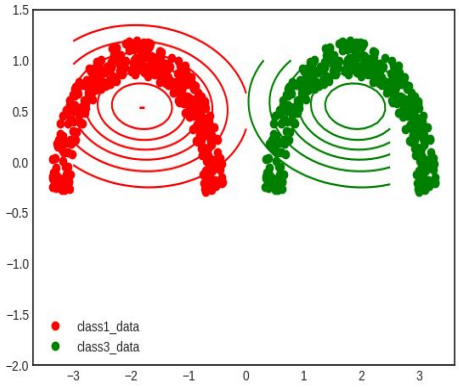




Decision region **Contour plot**

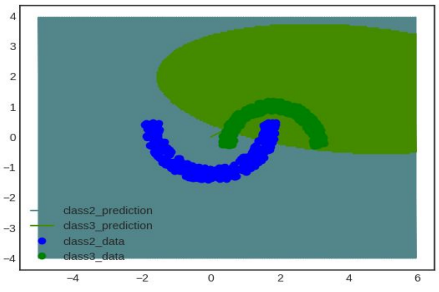
Class_1 vs Class_3

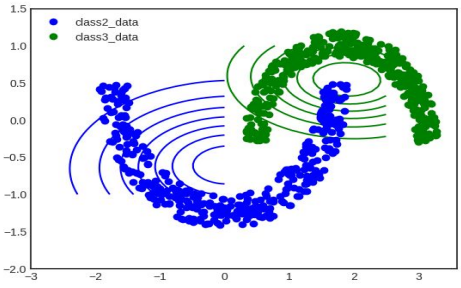




Decision region **Contour plot**

Class_2 vs Class_3





Decision region **Contour plot**

3.2.5.Inference

- For case_1: It can be seen that the decision region is linear and the corresponding contours of every class is close to circular .It doesn't work fine as the data is non-linear as the accuracy(74.59%) is very low
- For case_2: It can be seen that the decision region is linear and the corresponding contours of every class is elliptical .Here even with the linear boundary good accuracy (83.4%) is achieved
- For case_3: It can be inferred from the plots that the decision region is non-linear and the contours of every class is elliptical & its major and minor axes are parallel to coordinate axes.It gives quite well accuracy
- For case_4: It can be seen from the plots that the decision region is non- linear and the contours of different classes are elliptical and aligned in different directions .Good accuracy is achieved in this case.

3.3 Real world data set:

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

Confusion Matrix

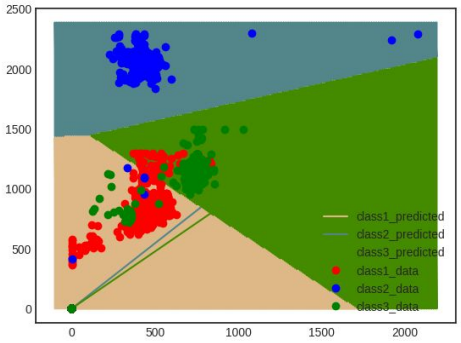
	Class1	class2	class3
class1	570	0	44
class2	16	553	4
class3	28	1	512

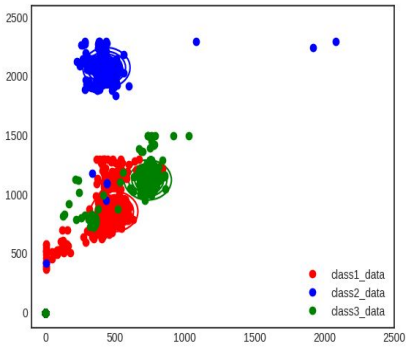
Analysis

	class1	**class2**	**class3**	**Mean**
Precision	0.9283	0.9981	0.9142	0.9468
Recall	0.9283	0.9509	0.9463	0.9418
F-measure	0.93887	0.9738	0.9154	0.9468

Accuracy : 94.61%

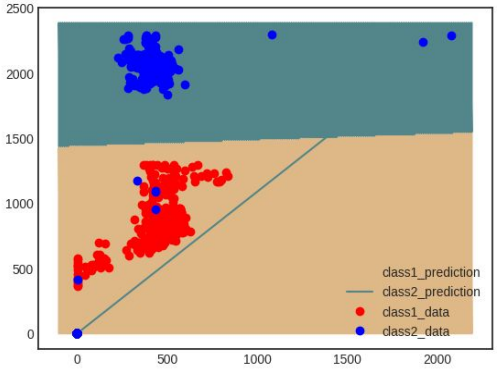
Class_1 vs Class_2 vs Class_3

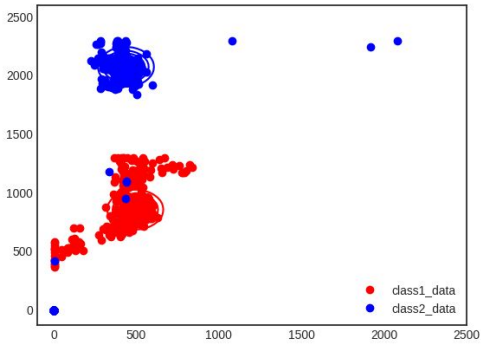




Decision region **Contour plot**

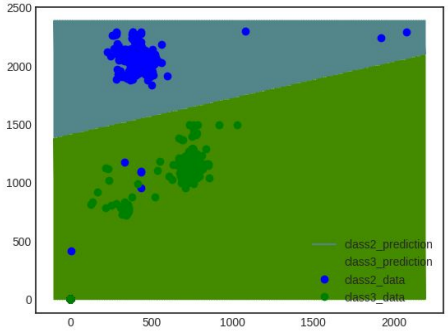
Class_1 vs Class_2

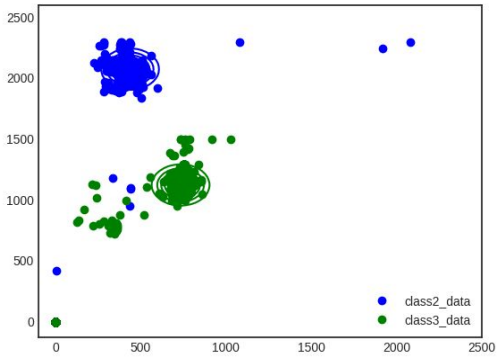




Decision region **Contour plot**

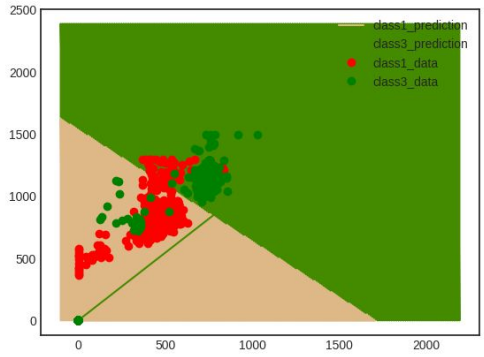
Class_2 vs Class_3

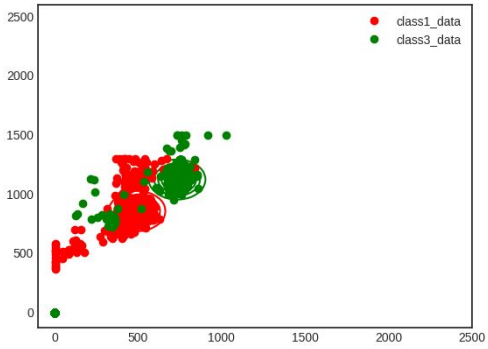




Decision region **Contour plot**

Class_1 vs Class_3





Decision region **Contour plot**

3.3.2 Case 2: Σ_i =Σ

Confusion Matrix

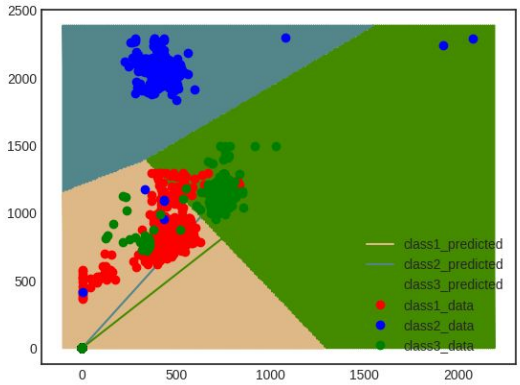
	Class1	**class2**	**class3**
class1	583	0	31
class2	16	553	4
class3	28	1	512

Analysis

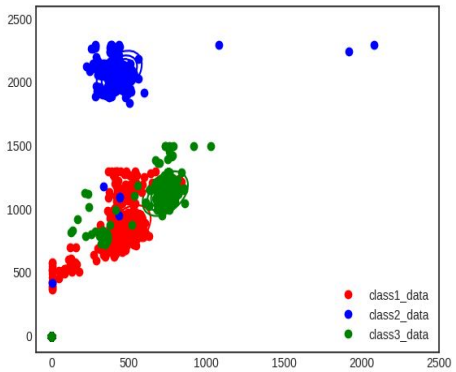
	class1	**class2**	**class3**	**Mean**
Precision	0.929	0.9981	0.9360	0.9543
Recall	0.9495	0.9650	0.9463	0.9536
F-measure	0.941	0.9812	0.94112	0.9539

Accuracy 95.37%

Class_1 vs Class_2 vs Class_3

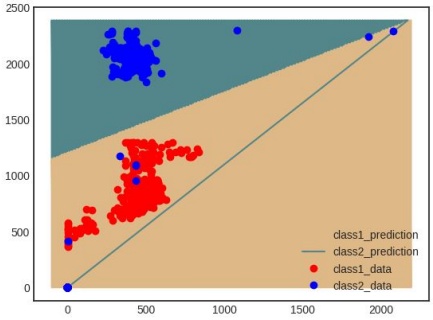


Decision region

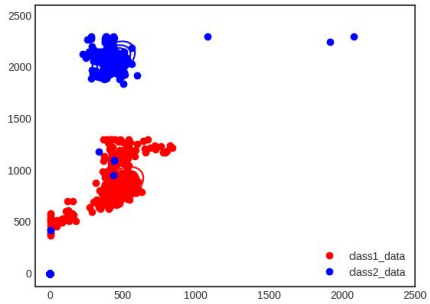


Contour plot

Class_1 vs Class_2

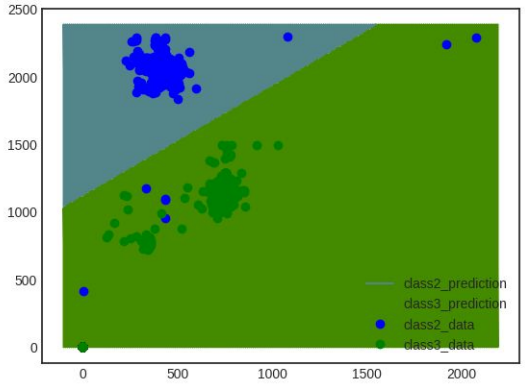


Decision region

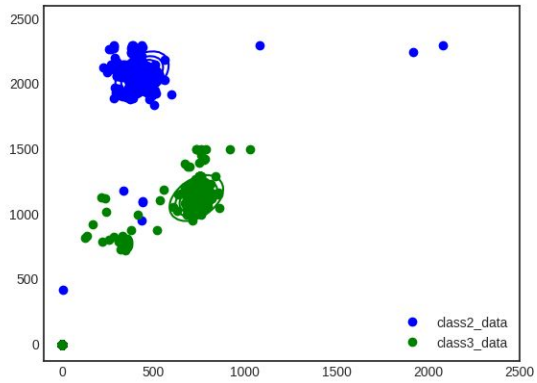


Contour plot

Class_2 vs Class_3

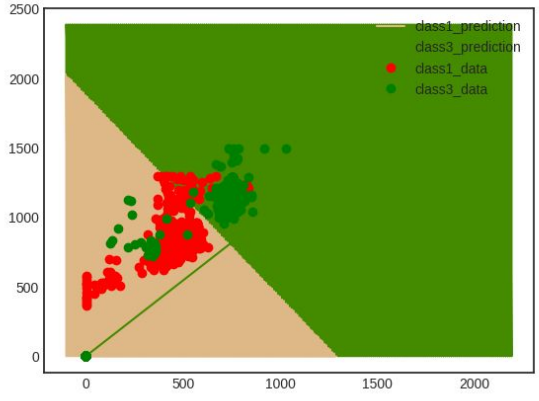


Decision region

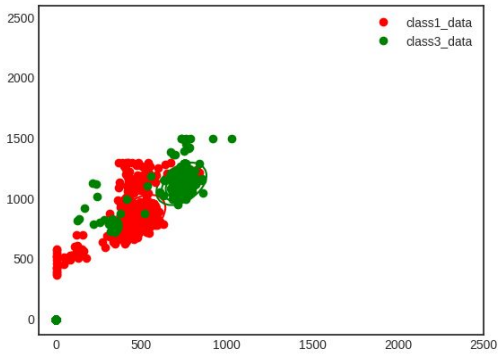


Contour plot

Class_1 vs Class_3



Decision region



Contour plot

3.3.3 Case 3: Σ_i is diagonal matrix

Confusion Matrix

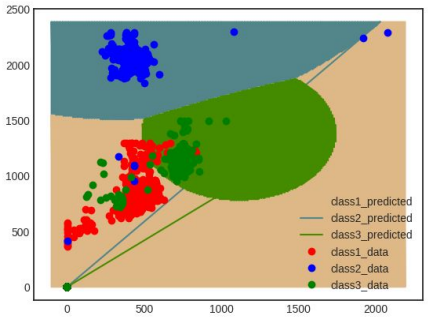
	Class1	**class2**	**class3**
class1	596	0	18
class2	19	554	0
class3	31	0	510

Analysis

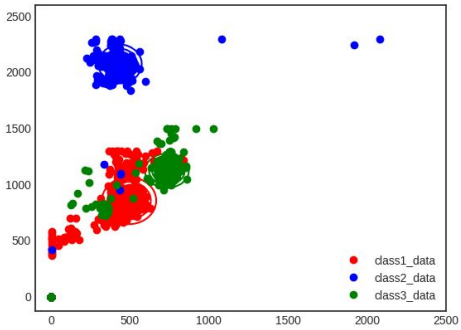
	class1	**class2**	**class3**	**Mean**
Precision	0.9226	1	0.9659	0.9628
Recall	0.9707	0.9668	0.9426	0.96
F-measure	0.9460	0.9831	0.9541	0.9613

Accuracy =96.06%

Class_1 vs Class_2 vs Class_3

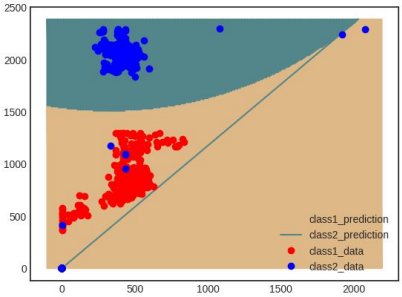


Decision region

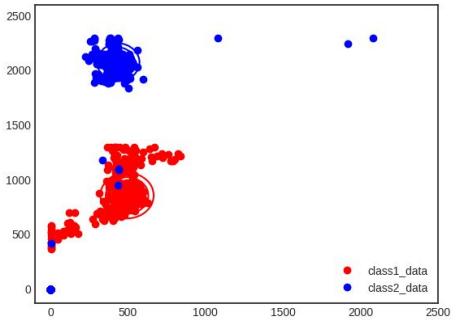


Contour plot

Class_1 vs Class_2

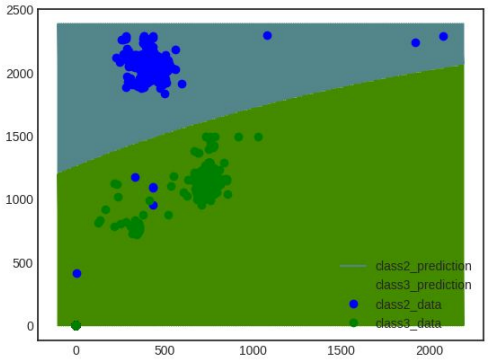


Decision region

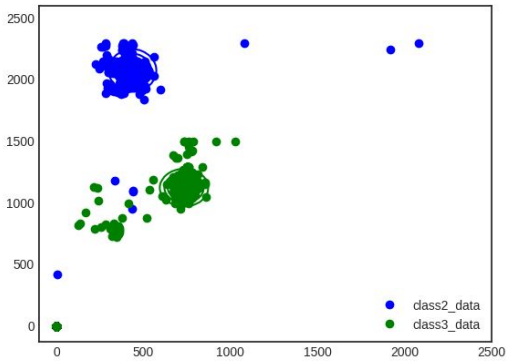


Contour plot

Class_2 vs Class_3

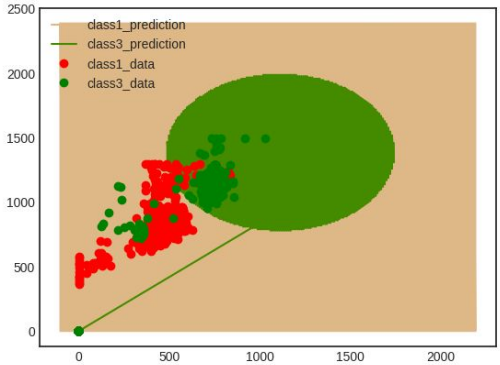


Decision region

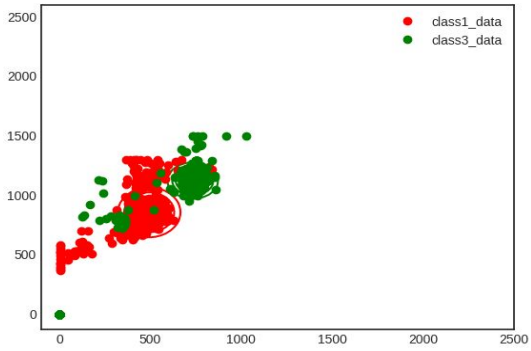


Contour plot

Class_1 vs Class_3



Decision region



Contour plot

3.3.4 Case 4: Σ_i is unique

Confusion Matrix

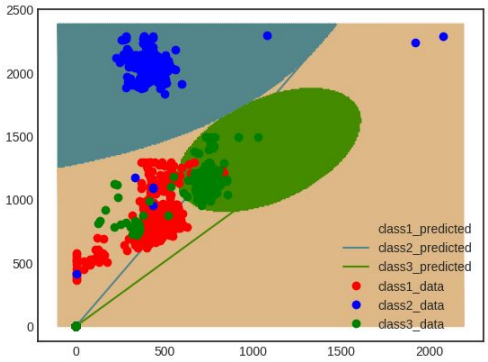
	Class1	**class2**	**class3**
class1	609	0	5
class2	19	554	0
class3	31	1	509

Analysis

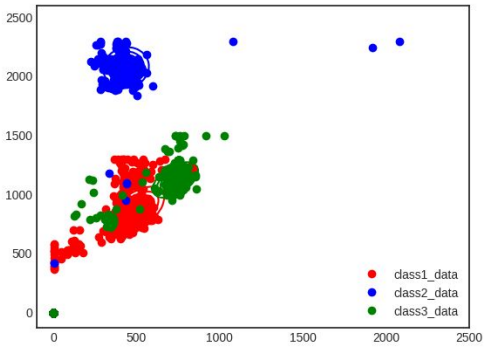
	Class1	**class2**	**class3**	**Mean**
Precision	0.9241	0.9982	0.9903	0.9708
Recall	0.9918	0.9668	0.9408	0.9664
F-measure	0.9568	0.9822	0.9649	0.9685

Accuracy 96.75%

Class_1 vs Class_2 vs Class_3

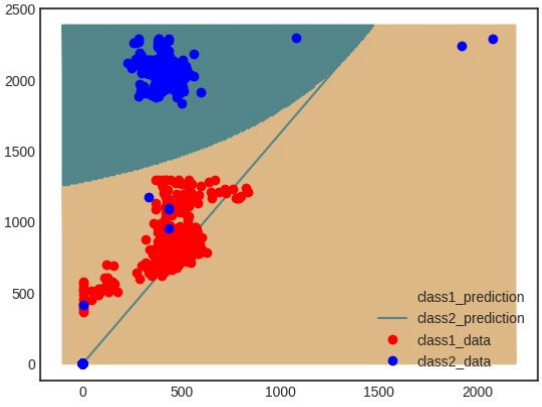


Decision region

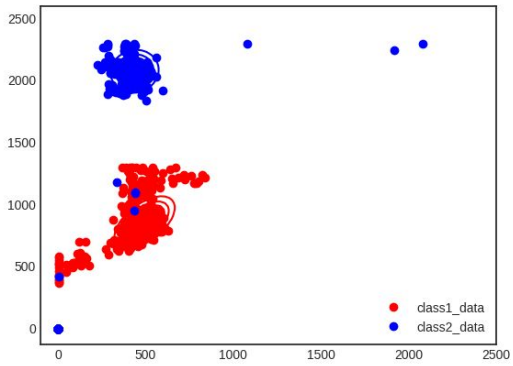


Contour plot

Class_1 vs Class_2

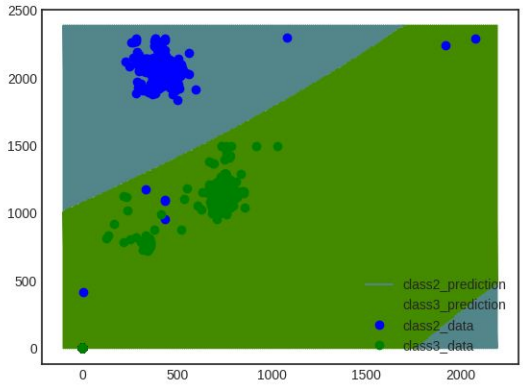


Decision region

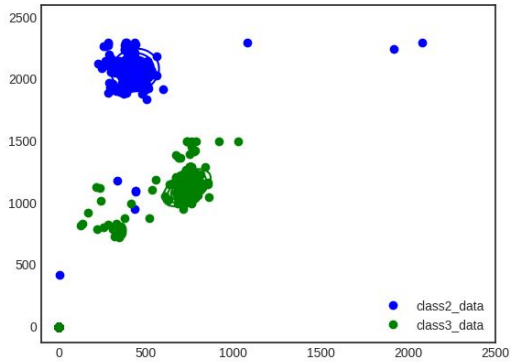


Contour plot

Class_2 vs Class_3

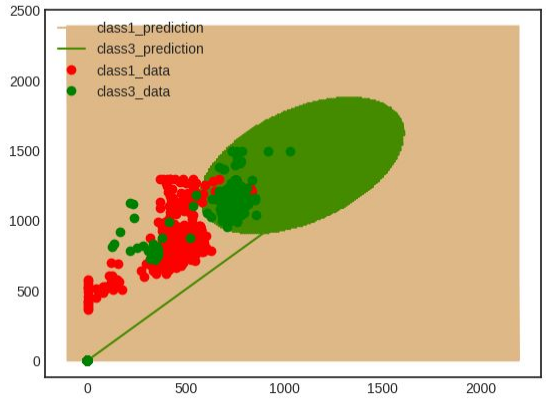


Decision region

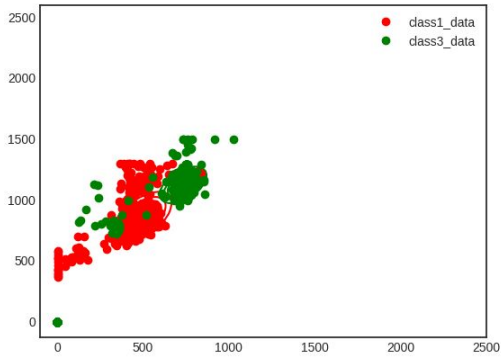


Contour plot

Class_1 vs Class_3



Decision region



Contour plot

3.3.5 Inference:

- For case_1: It can be seen that the decision region is linear and the corresponding contours of every class is close to circular(not exactly)but theoretically its should be circular.In this case accuracy is lowest among.
- For case_2: It can be seen that the decision region is linear and the corresponding contours of every class is elliptical .
- For case_3: It can be inferred from the plots that the decision region is non-linear and the contours of every class is elliptical & its major and minor axes are parallel to coordinate axes.It gives better accuracy than previous two cases.
- For case_4: It can be seen from the plots hat the decision region is non- linear and the contours of different classes are elliptical and aligned in different directions.In this case the accuracy is highest.
- As the data is overlapping so any decision boundary can't give the best accuracy.

4.conclusion

- Bayes Classifier gives the higher accuracy for linear data but low accuracy for non-linear data and overlapping data
- The nature of decision boundary is dependent on the covariance matrices as in case_1 & case_2 decision boundary is linear and for case_3 & casr_4 it is quadratic.The same result we have got theoretically .
- Since we have forcefully modified the covariance matrix in case_1 & case_2 and case_3 So we didn't get the 100% desired result.