Pattern Recognition (CS-669)

Assignment 1

Bayes Classifier

Suryavanshi Virendrasingh (B16037)
Abhishek Pal (B15203)
Nemani Sri Hari (B15224)
Group - 13

September 16, 2018

1 Objective

1. To build a Bayes Classifier and classify the following datasets -
   - 2D Dataset - 1 (Artificial)
     - Linearly Separable Dataset
     - Non Linearly Separable Dataset
   - 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.) Data for each class is partitioned into 75% for training and 25% for testing.

2.) The data set for each class is assumed to come from Gaussian distribution.

3.) In Case-1 ($\Sigma = \sigma^2 I$), mean of the covariance matrix for each class was calculated and it's off-diagonal terms were assumed to be 0 for further calculations.

4.) In Case-2 ($\Sigma_i = \Sigma$ for every class), mean of the covariance matrix for each class was calculated for further calculations.

5.) In Case-3 (Σ_i is a diagonal matrix), the 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 (Σ_i is unique), no assumptions were made for further calculations.

7.) Based on 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 analysis was made.

3 Formulae/Definition of Terms

Confusion Matrix is defined as the matrix of the actual class vs predicted class.

Confusion Matrix

	Actual C1	**Actual C2**	**Actual C3**
Predicted C1	Actual C1 & Predicted C1	Actual C2 & Predicted C1	Actual C3 & Predicted C1
Predicted C2	Actual C1 & Predicted C2	Actual C2 & Predicted C2	Actual C3 & Predicted C2
Predicted C3	Actual C1 & Predicted C3	Actual C2 & Predicted C3	Actual C3 & Predicted C3

Classification Accuracy = (Number of samples correctly classified/Total number of samples)*100 = ((TP+TN)/(TP+TN+FP+FN)) * 100

Precision for one class = Number of true and predicted class 1 cases/Number of predicted class 1 cases = TP/(TP +FP)

Mean precision = (precision 1+precision 2+ precision 3)/3

Recall for one class = number of true and predicted class 1 cases/Number of true class 1 cases = TP/(TP+FN)

Mean recall = (recall 1+recall 2+recall 3)/3

F-measure for one class = (precision * recall)/((precision + recall)/2)

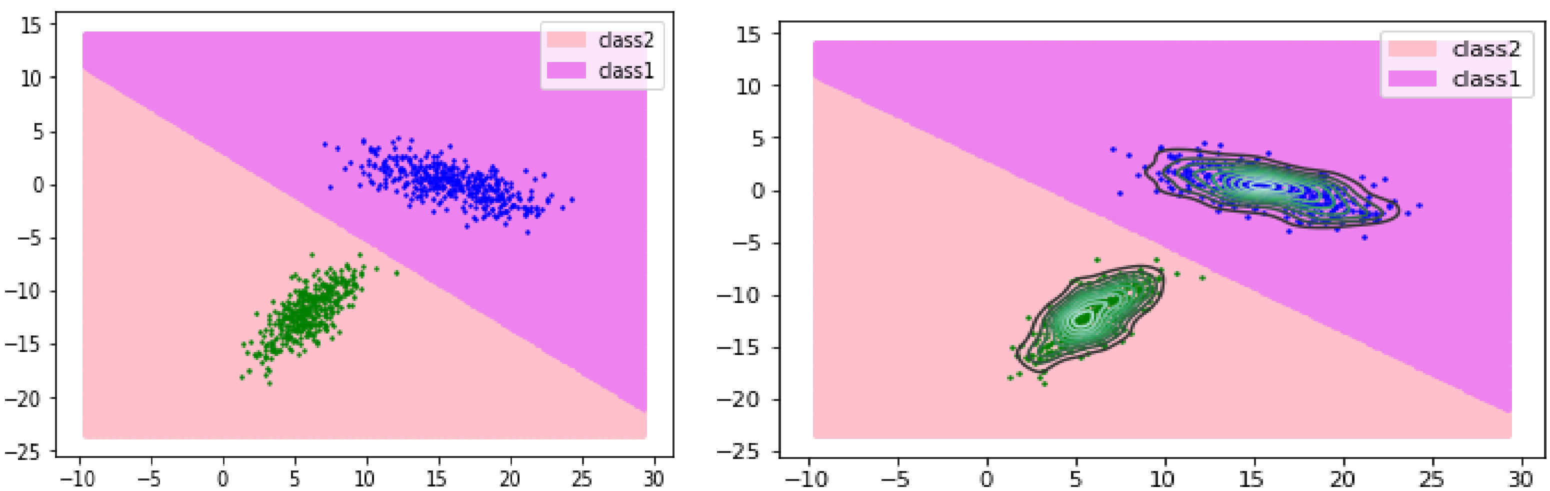
Mean F-measure = (F-measure 1+F-measure 2+ F-measure 3)/3

4 Observations

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

4.1.1 Linear Data

Class 1 vs Class 2



Accuracy = 100%

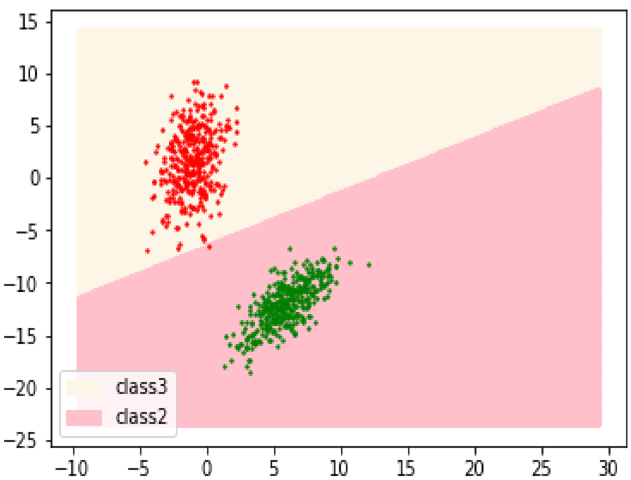
Confusion Matrix

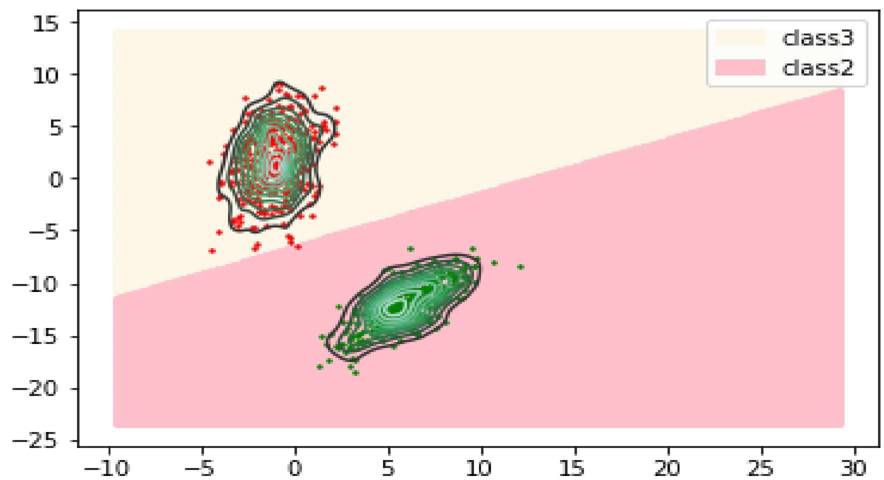
Actual/ Predicted	**Class 1**	**Class 2**
Class 1	125	0
Class 2	0	125

Analysis

	Class 1	**Class 2**
Precision	1.000	1.000
Recall	1.000	1.000
F-measure	1.000	1.000

Class 2 vs Class 3





Accuracy = 100%

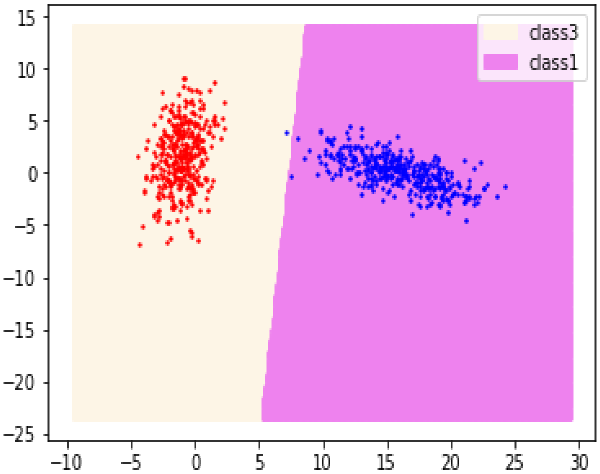
Confusion Matrix

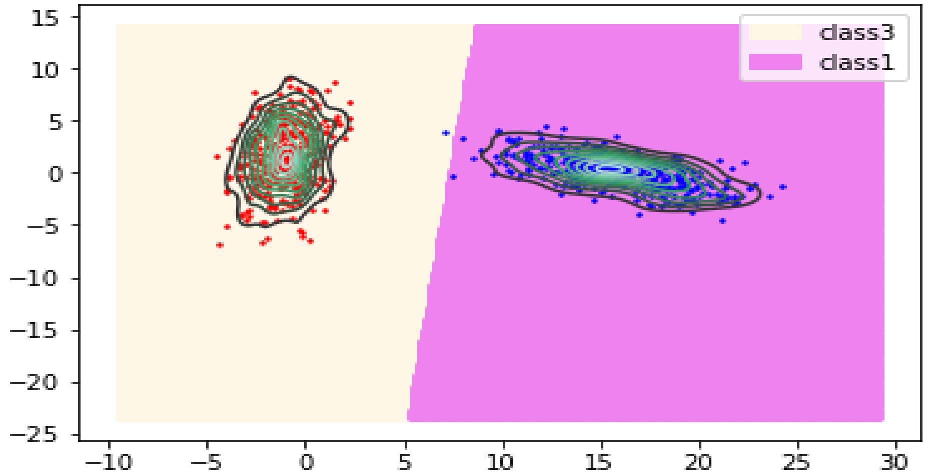
Actual/ Predicted	**Class 2**	**Class 3**
Class 2	125	0
Class 3	0	125

Analysis

	Class 2	**Class 3**
Precision	1.000	1.000
Recall	1.000	1.000
F-measure	1.000	1.000

Class 1 vs Class 3





Accuracy = 99.6%

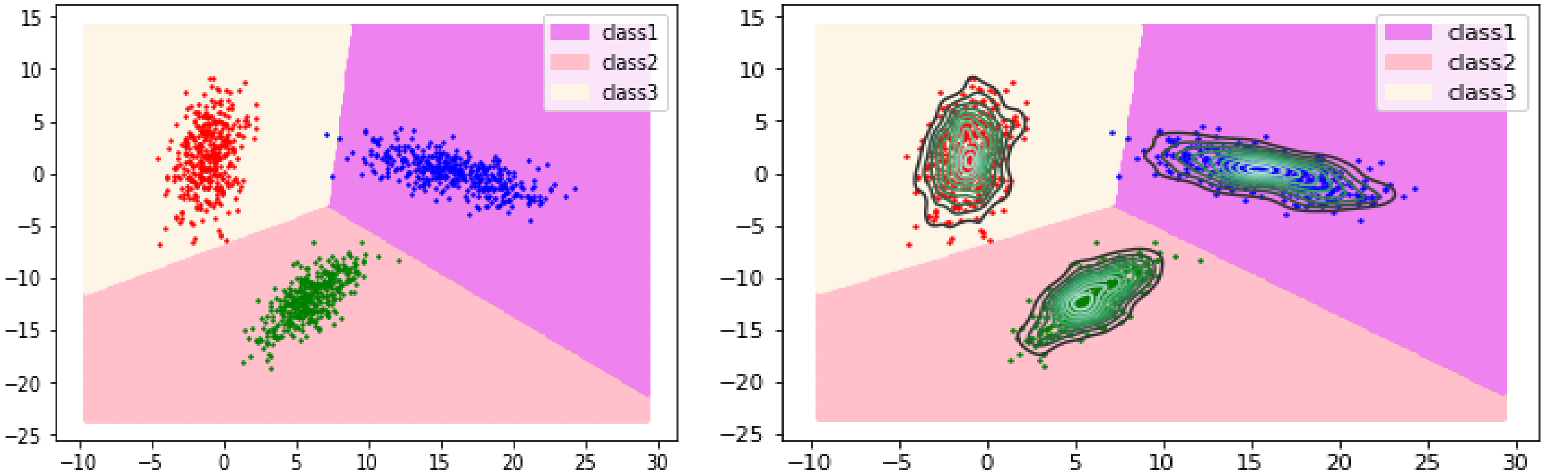
Confusion Matrix

Actual/ Predicted	**Class 1**	**Class 3**
Class 1	124	0
Class 3	0	125

Analysis

	Class 1	**Class 3**
Precision	1.000	99.206
Recall	0.992	1.000
F-measure	0.995	0.996

Class 1 vs Class 2 vs class 3



Accuracy = 99.73%

Confusion Matrix

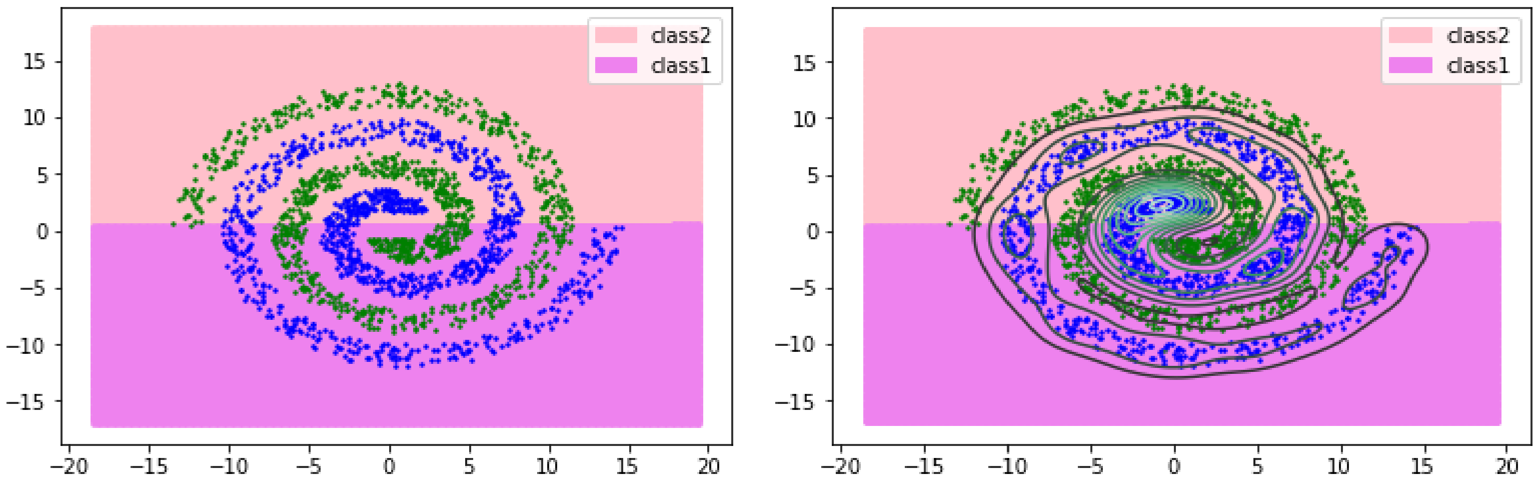
Actual/Predicted	Class 1	Class 2	Class 3
Class 1	124	0	0
Class 2	0	125	0
Class 3	0	0	125

Analysis

	Class 1	Class 2	Class 3
Precision	1.000	1.000	0.992
Recall	0.992	1.000	1.000
F-measure	0.995	1.000	0.996

4.1.2 Non-Linear Data

Class 1 vs Class 2



Accuracy = 55.67%

Confusion Matrix

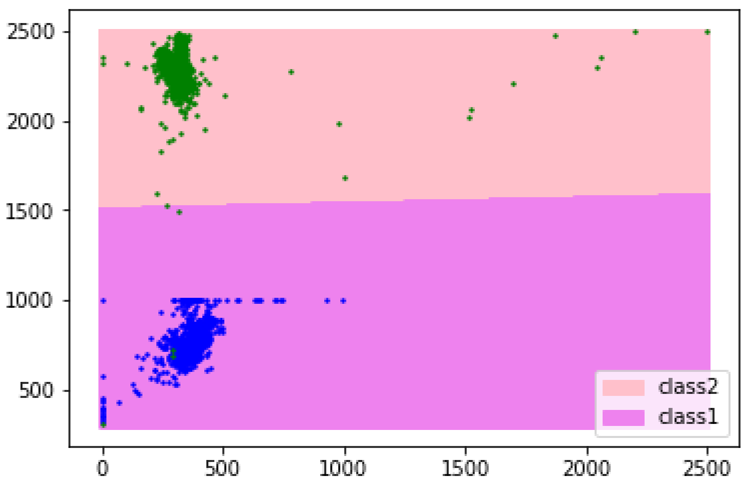
Actual/ Predicted	**Class 1**	**Class 2**
Class 1	188	138
Class 2	151	175

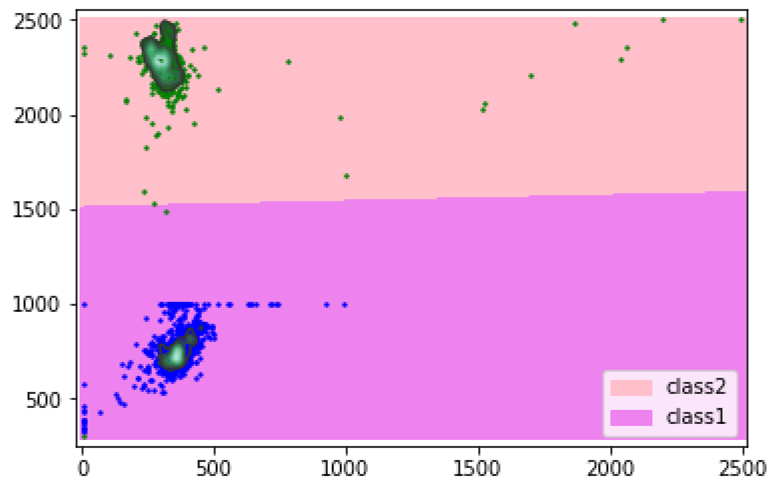
Analysis

	Class 1	**Class 2**
Precision	0.554	0.559
Recall	0.576	0.536
F-measure	0.565	0.547

4.1.3 Real World Data

Class 1 vs Class 2





Accuracy = 100%

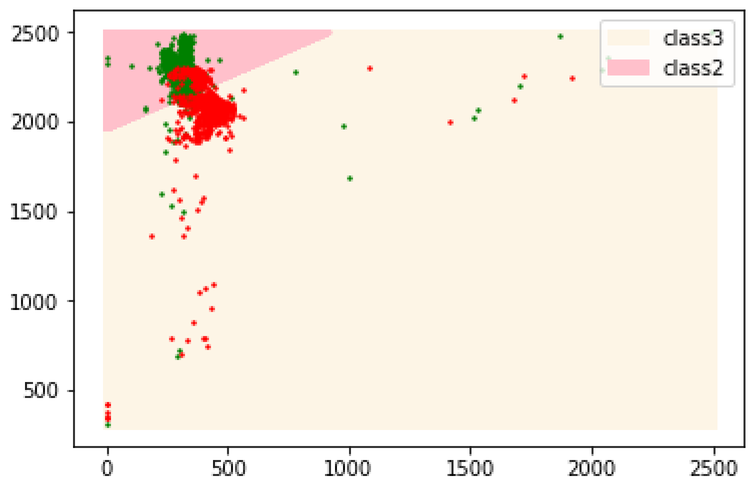
Confusion Matrix

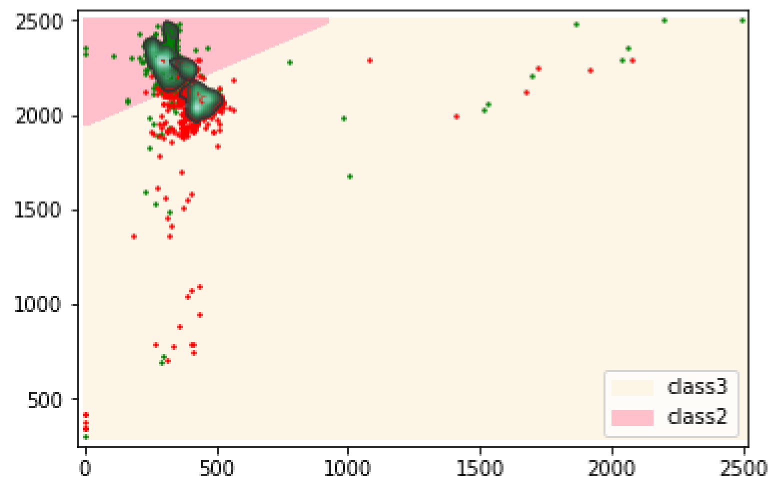
Actual/ Predicted	**Class 1**	**Class 2**
Class 1	622	0
Class 2	0	597

Analysis

	Class 1	**Class 2**
Precision	1.000	1.000
Recall	1.000	1.000
F-measure	1.000	1.000

Class 2 vs Class 3





Accuracy = 86.92%

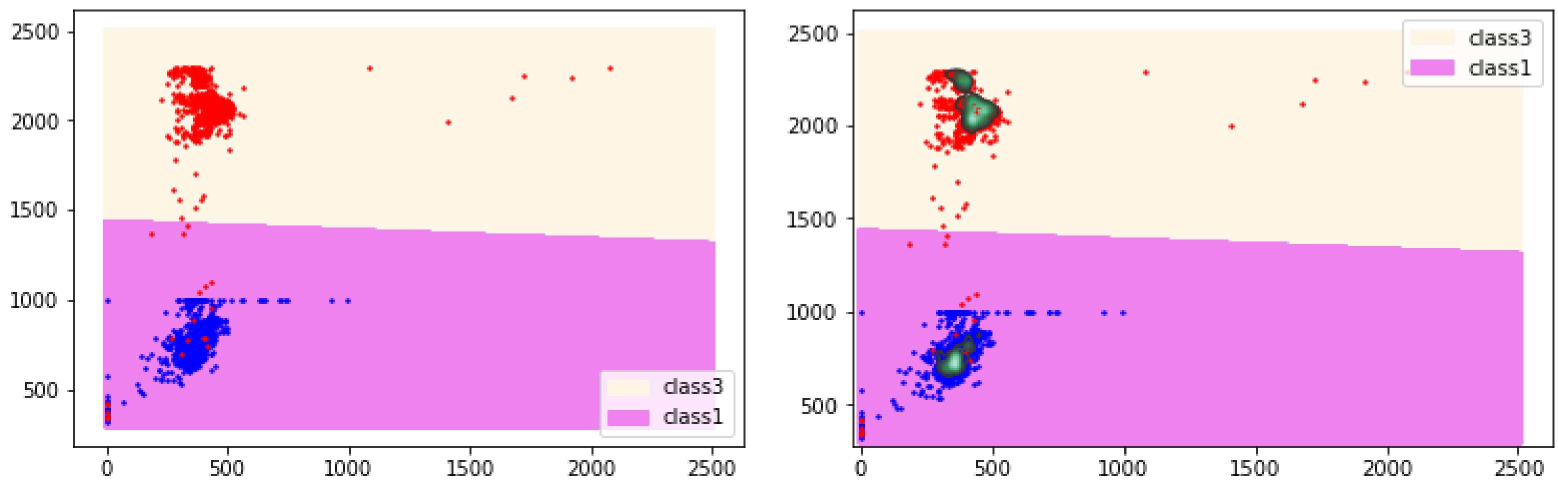
Confusion Matrix

Actual/ Predicted	**Class 2**	**Class 3**
Class 2	557	40
Class 3	113	460

Analysis

	Class 2	**Class 3**
Precision	0.831	0.920
Recall	0.932	0.802
F-measure	0.879	0.857

Class 1 vs Class 3



Accuracy = 99.74%

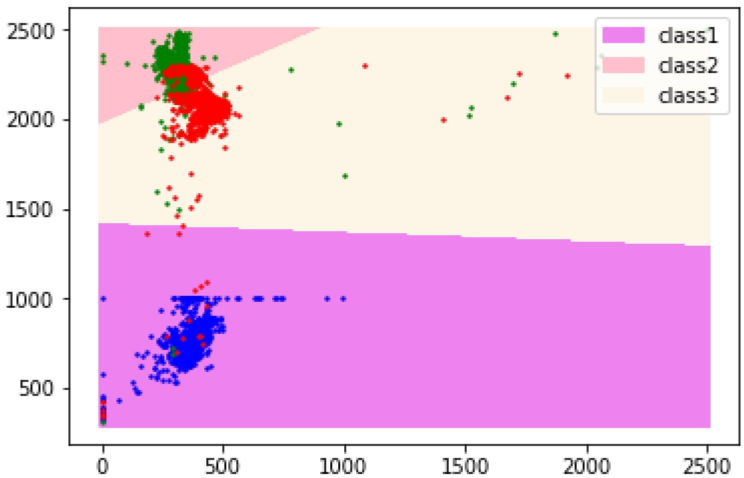
Confusion Matrix

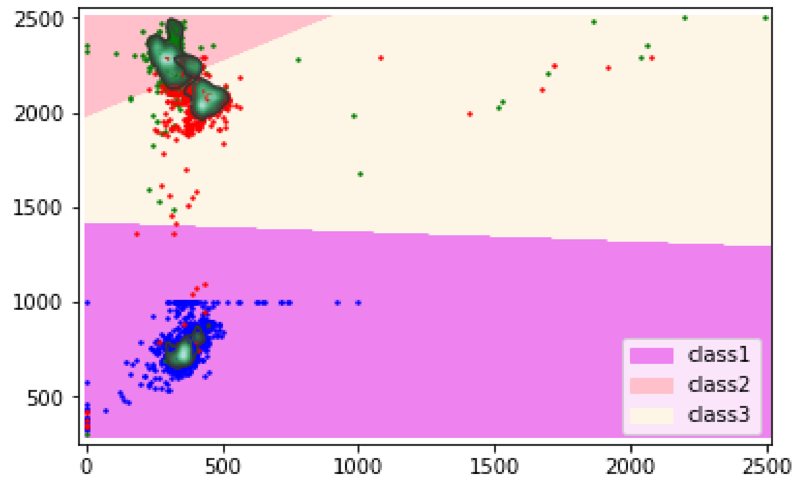
Actual/ Predicted	**Class 1**	**Class 3**
Class 1	622	0
Class 3	3	570

Analysis

	Class 1	**Class 3**
Precision	99.520	1.000
Recall	1.000	99.476
F-measure	99.759	99.737

Class 1 vs Class 2 vs class 3





Accuracy = 99.832%

Confusion Matrix

Actual/ Predicted	**Class 1**	**Class 2**	**Class 3**
Class 1	622	0	0
Class 2	0	557	40
Class 3	3	113	457

Analysis

	Class 1	**Class 2**	**Class 3**
Precision	0.995	0.831	0.919
Recall	1	0.932	0.797
F-measure	0.997	0.879	0.854

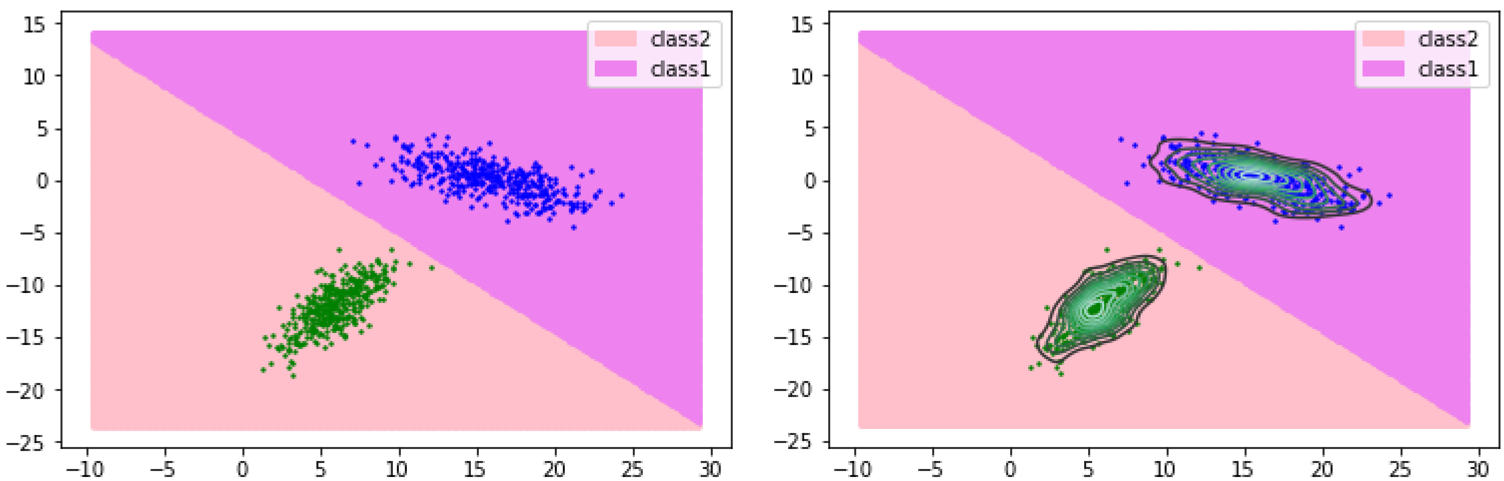
4.1.4 Inference

1. In this case we assume that our features are statistically independent which is why we consider all the classes have same covariance matrix which is diagonal.
2. We get the covariance matrix by taking average of all the diagonal elements of all classes. $\Sigma_i = \sigma^2 I$
3. This method works best when the data is linearly separable.
4. The decision boundary formed will be perpendicular to the line joining the means and will pass through the point on the line joining the mean. The point of intersection will be nearer to the class with more prior. In our case here as the prior is equal the decision boundary passes through the midpoint of means.
5. This can be inferred from the first and third case where data is almost linearly separable the accuracy is 100% and in the second case where data is not linearly separable the accuracy is 55.7%.

4.2 Case 2 - $\Sigma_i = \Sigma$

4.2.1 Linear Data

Class 1 vs Class 2



Accuracy = 100%

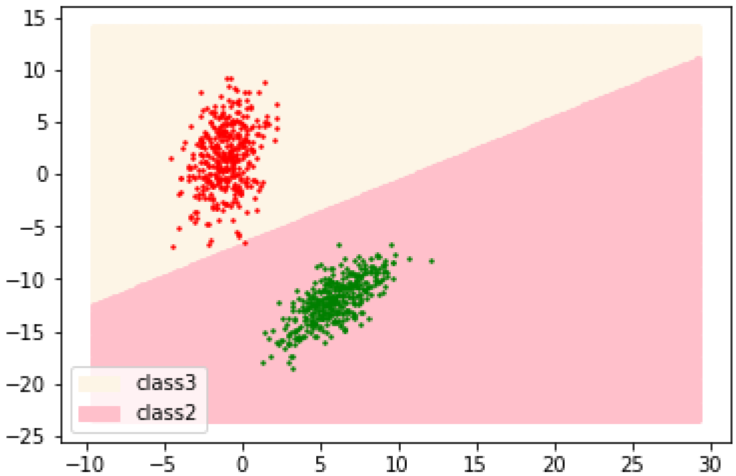
Confusion Matrix

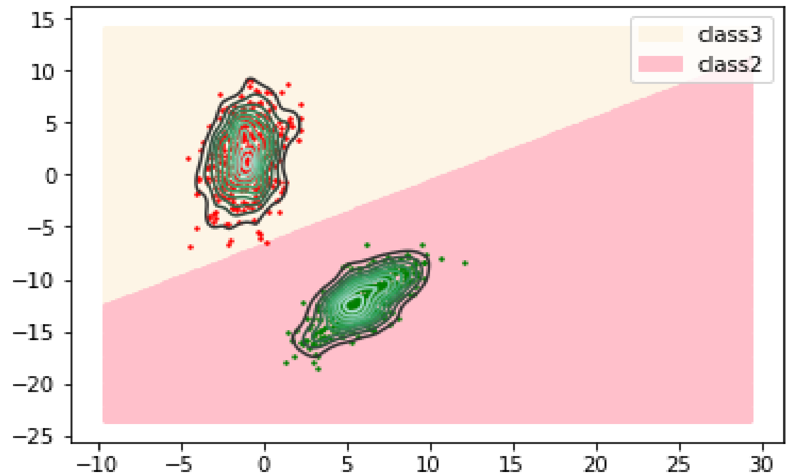
Actual/ Predicted	**Class 1**	**Class 2**
Class 1	125	0
Class 2	0	125

Analysis

	Class 1	**Class 2**
Precision	1.000	1.000
Recall	1.000	1.000
F-measure	1.000	1.000

Class 2 vs class 3





Accuracy = 100%

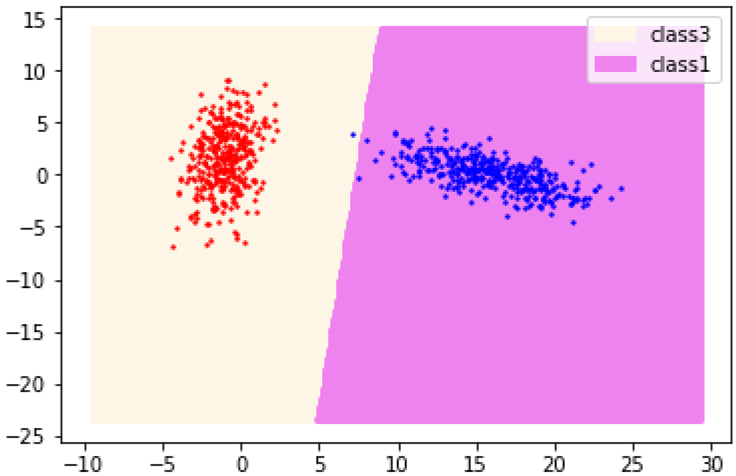
Confusion Matrix

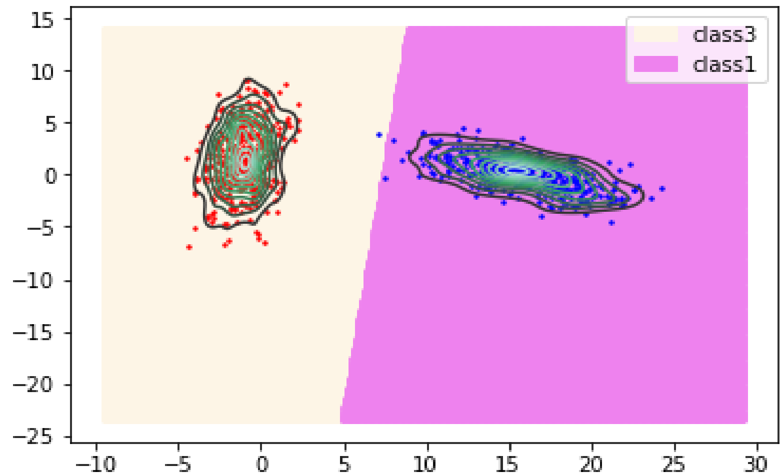
Actual/ Predicted	**Class 2**	**Class 3**
Class 2	125	0
Class 3	0	125

Analysis

	Class 2	**Class 3**
Precision	1.000	1.000
Recall	1.000	1.000
F-measure	1.000	1.000

Class 1 vs Class3





Accuracy = 99.60%

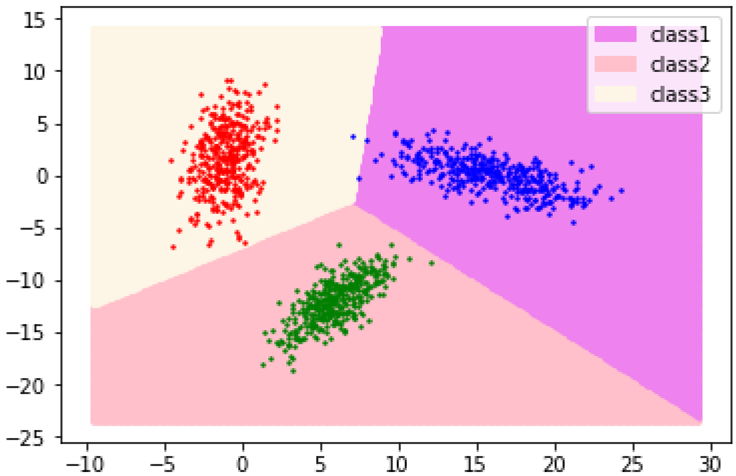
Confusion Matrix

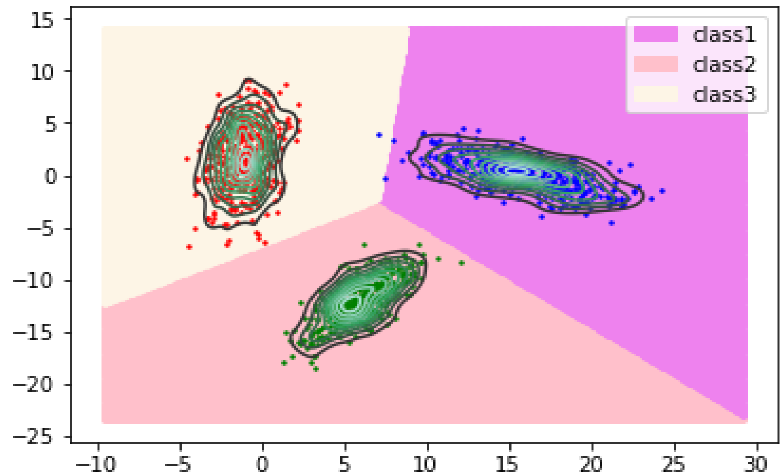
Actual/ Predicted	**Class 1**	**Class 3**
Class 1	124	1
Class 3	0	125

Analysis

	Class 1	**Class 3**
Precision	1.000	0.992
Recall	0.992	1.000
F-measure	0.995	0.996

Class 1 vs Class 2 vs class 3





Accuracy = 99.73%

Confusion Matrix

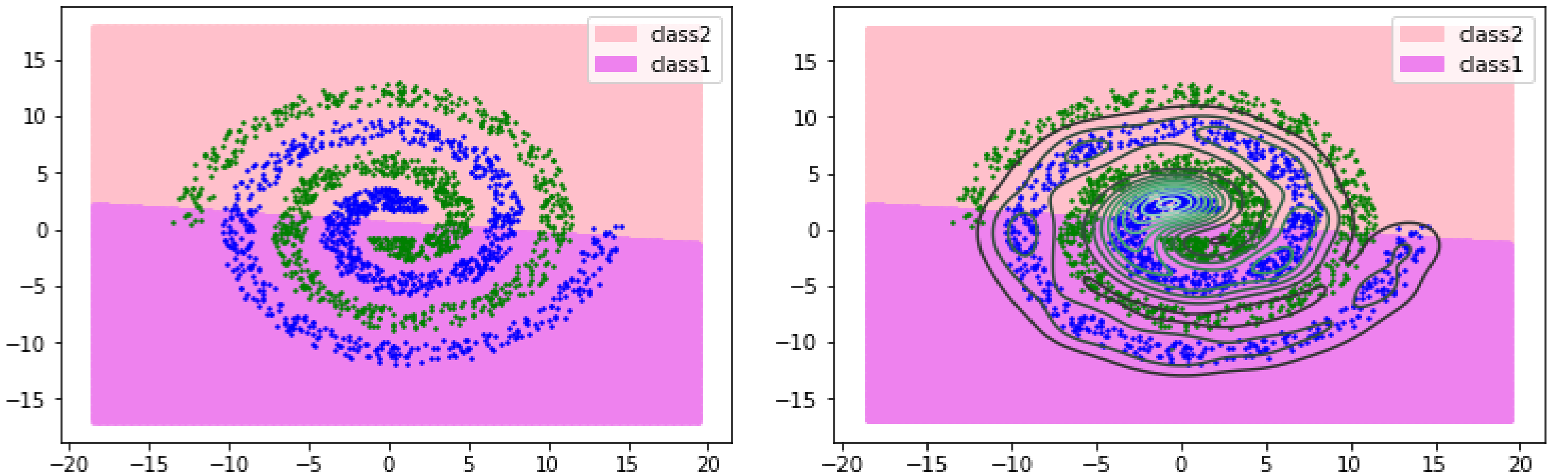
Actual/ Predicted	Class 1	Class 2	Class 3
Class 1	124	0	0
Class 2	0	125	0
Class 3	0	0	125

Analysis

	Class 1	Class 2	Class 3
Precision	1.000	1.000	0.992
Recall	0.992	1.000	1.000
F-measure	0.995	1.000	0.996

4.2.2 Non-Linear Data

Class 1 vs Class 2



Accuracy = 55.67%

Confusion Matrix

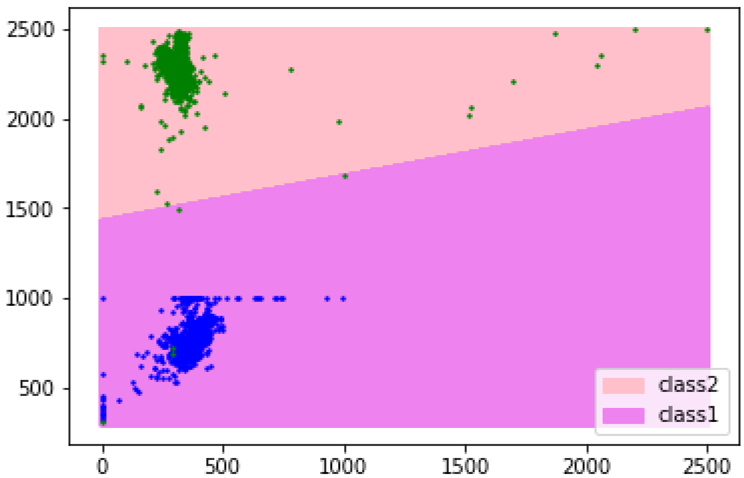
Actual/ Predicted	**Class 1**	**Class 2**
Class 1	188	138
Class 2	151	175

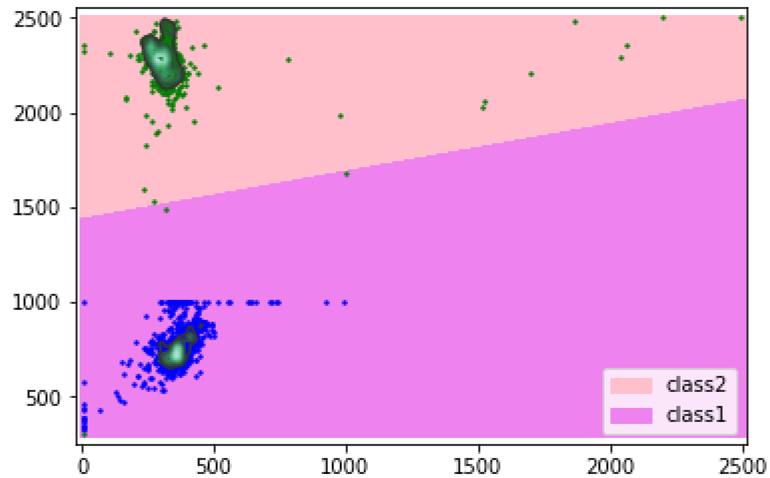
Analysis

	Class 1	**Class 2**
Precision	0.554	0.559
Recall	0.576	0.536
F-measure	0.565	0.547

4.2.3 Real World Data

Class 1 vs Class 2





Accuracy = 100%

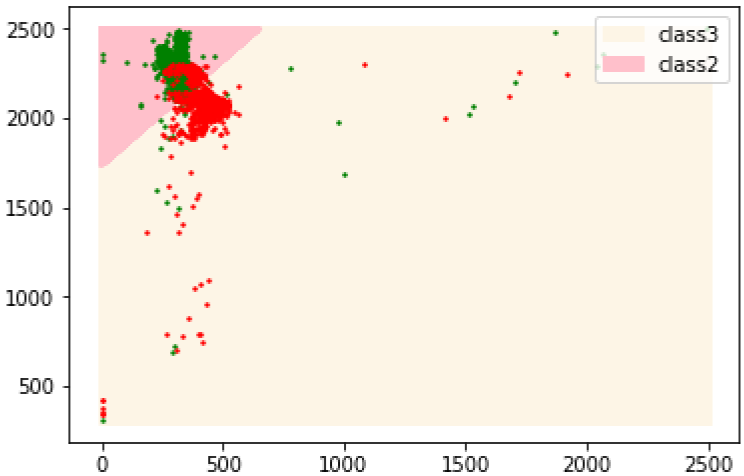
Confusion Matrix

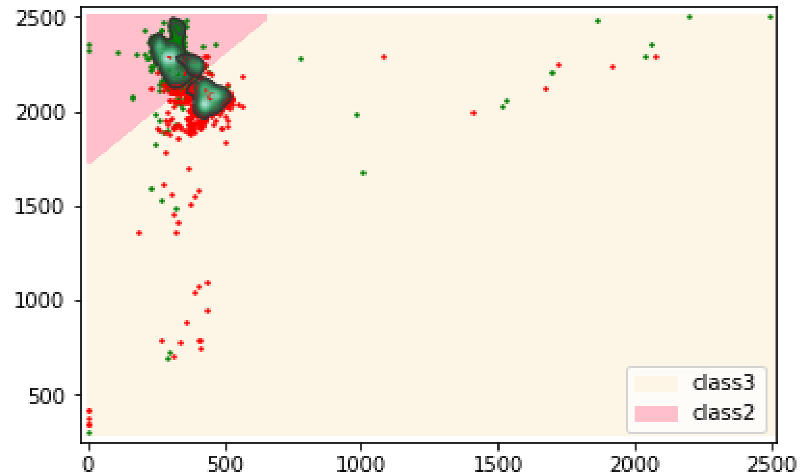
Actual/ Predicted	**Class 1**	**Class 2**
Class 1	622	0
Class 2	0	597

Analysis

	Class 1	**Class 2**
Precision	1.000	1.000
Recall	1.000	1.000
F-measure	1.000	1.000

Class 2 vs class 3





Accuracy = 88.37%

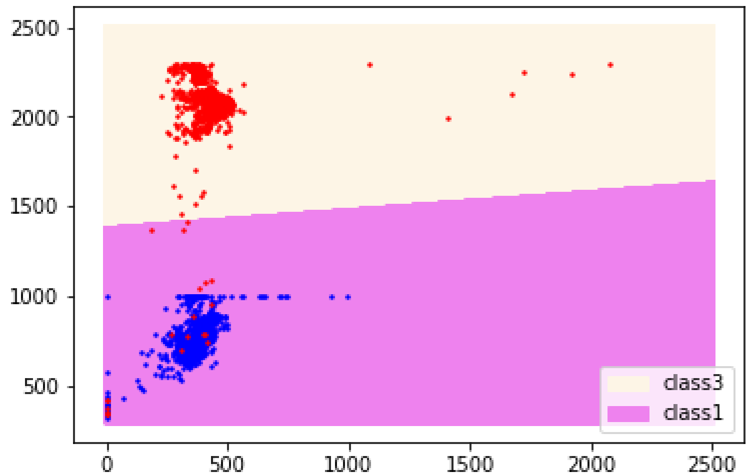
Confusion Matrix

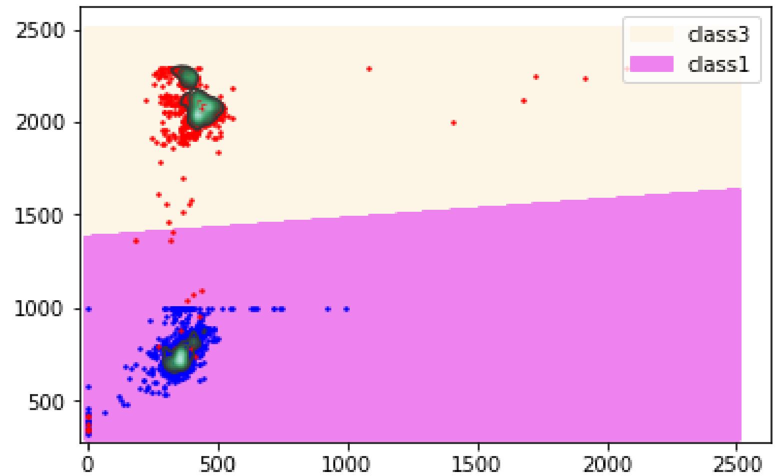
Actual/ Predicted	**Class 2**	**Class 3**
Class 2	572	25
Class 3	111	462

Analysis

	Class 2	**Class 3**
Precision	0.837	0.948
Recall	0.958	0.806
F-measure	0.893	0.871

Class 1 vs class 3





Accuracy = 99.74%

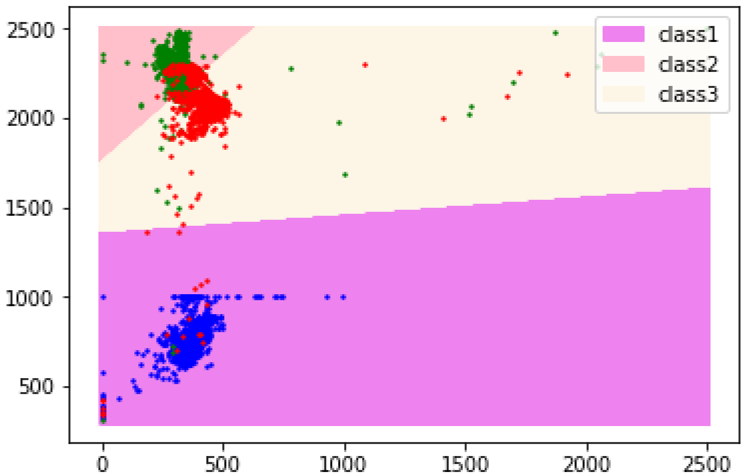
Confusion Matrix

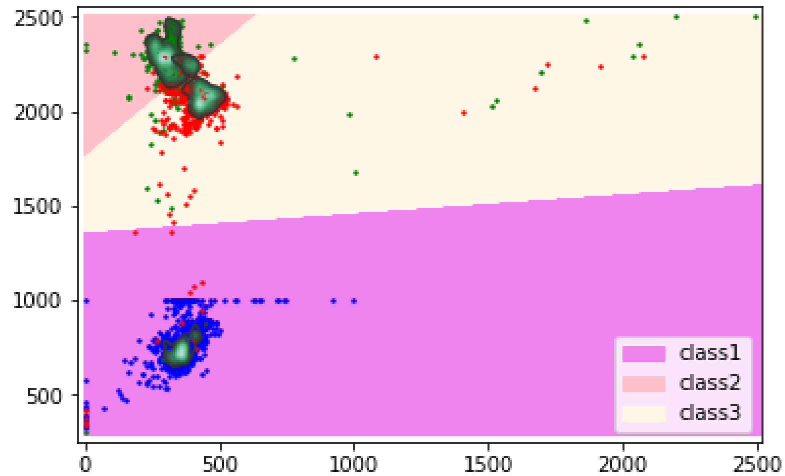
Actual/ Predicted	**Class 1**	**Class 3**
Class 1	622	0
Class 3	3	570

Analysis

	Class 1	**Class 3**
Precision	0.995	1.000
Recall	1.000	0.994
F-measure	0.997	0.997

Class 1 vs Class 2 vs class 3





Accuracy = 99.83%

Confusion Matrix

Actual/ Predicted	Class 1	Class 2	Class 3
Class 1	622	0	0
Class 2	0	572	25
Class 3	3	111	459

Analysis

	Class 1	Class 2	Class 3
Precision	0.995	0.837	0.948
Recall	1.000	0.958	0.801
F-measure	0.997	0.893	0.868

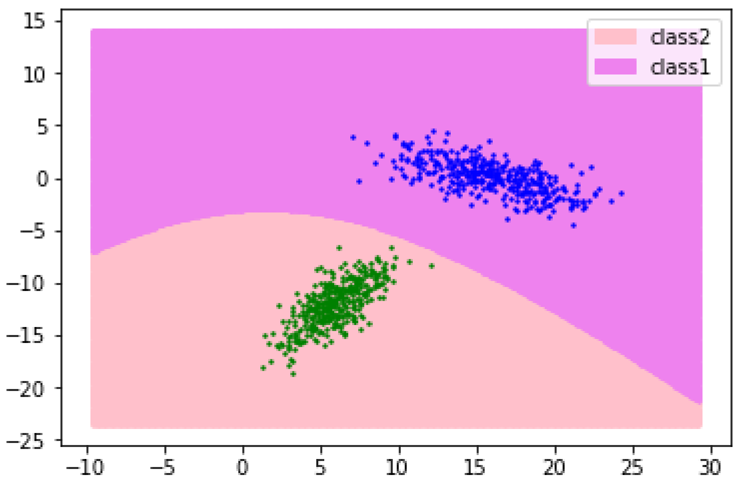
4.2.4 Inference

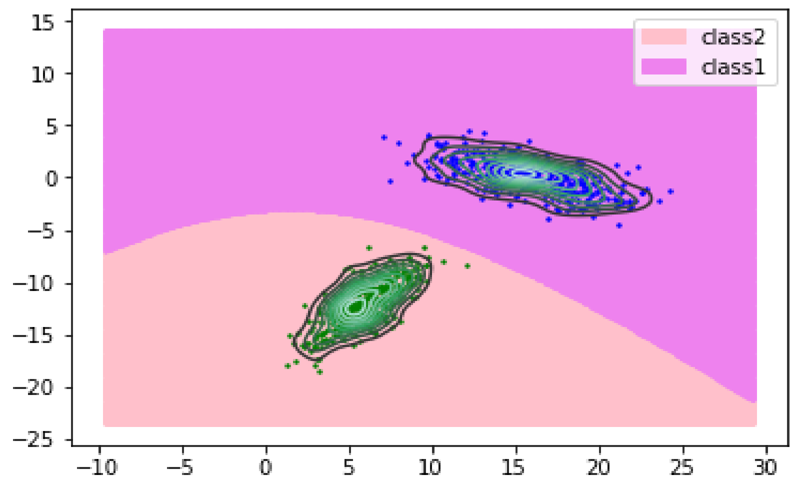
1. In this case we consider all the classes have same covariance matrix which is arbitrary but not diagonal.
2. We get the covariance matrix by taking average of all the covariance matrix elements of all classes. Σ = summation of covariance matrices/total no of classes.
3. This method works best when the data is linearly separable.
4. The decision boundary formed will not be perpendicular but inclined to the line joining the means (angle basing on the inverse of covariance matrix) and will pass through the point on the line joining the mean. The point of intersection will be nearer to the class with more prior. In our case here as the prior is equal the decision boundary passes through the midpoint of means.
5. This can be inferred from the first and third case where data is almost linearly separable the accuracy is 100% and in the second case where data is not linearly separable the accuracy is 55.7% which are almost similar to the case 1.

4.3 Case 3 – Σ_i *is a diagonal matrix*

4.3.1 Linear Data

Class 1 vs Class 2





Accuracy = 100%

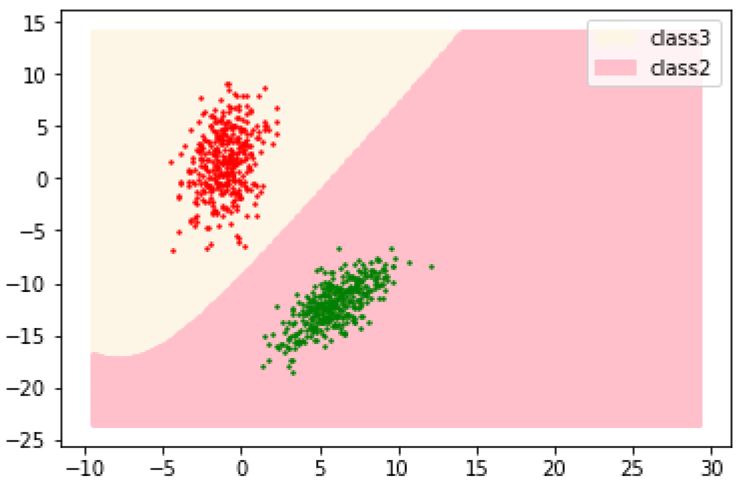
Confusion Matrix

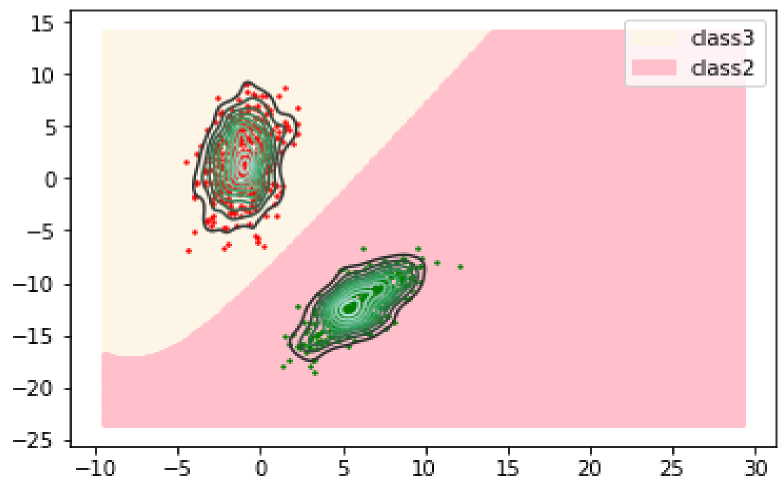
Actual/ Predicted	**Class 1**	**Class 2**
Class 1	125	0
Class 2	0	125

Analysis

	Class 1	**Class 2**
Precision	1.000	1.000
Recall	1.000	1.000
F-measure	1.000	1.000

Class 2 vs class 3





Accuracy = 100%

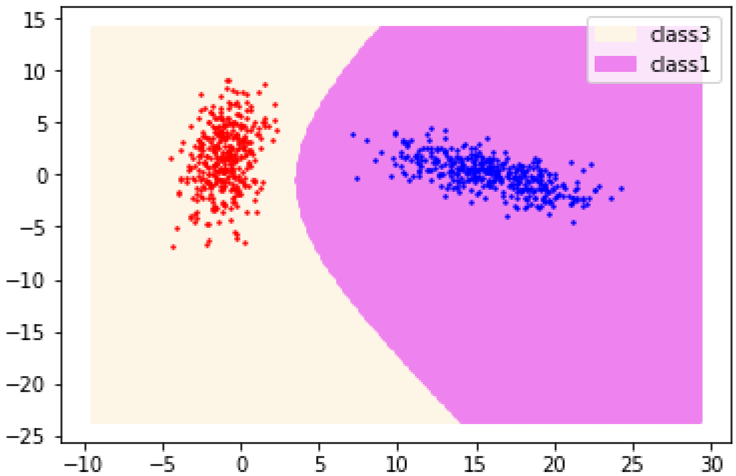
Confusion Matrix

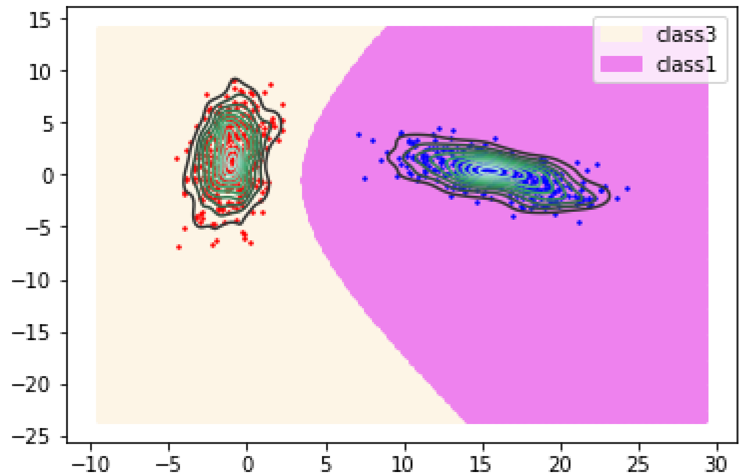
Actual/ Predicted	**Class 2**	**Class 3**
Class 2	125	0
Class 3	0	125

Analysis

	Class 2	**Class 3**
Precision	1.000	1.000
Recall	1.000	1.000
F-measure	1.000	1.000

Class 1 vs Class 3





Accuracy = 100%

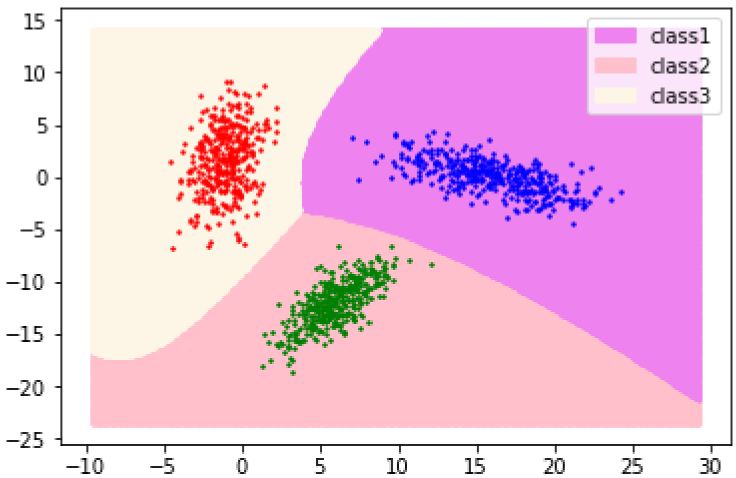
Confusion Matrix

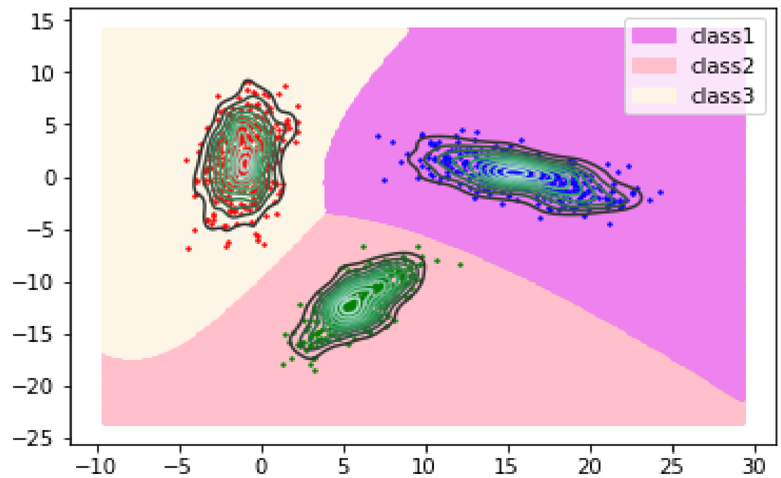
Actual/ Predicted	**Class 1**	**Class 3**
Class 1	125	0
Class 3	0	125

Analysis

	Class 1	**Class 3**
Precision	1.000	1.000
Recall	1.000	1.000
F-measure	1.000	1.000

Class 1 vs Class 2 vs class 3





Accuracy = 100%

Confusion Matrix

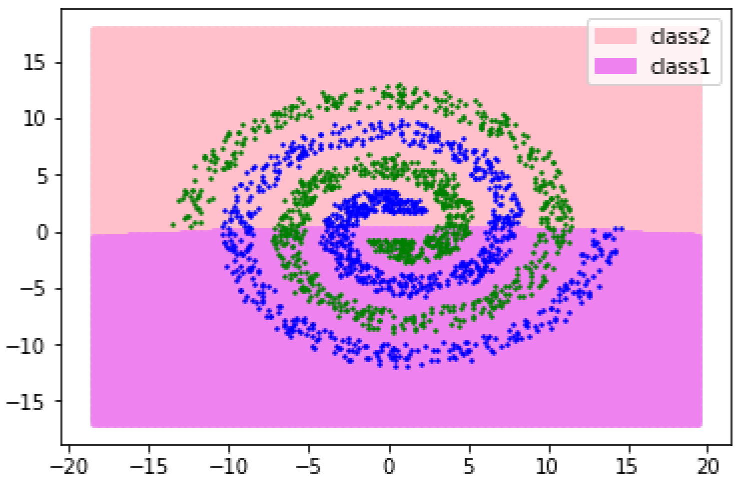
Actual/ Predicted	**Class 1**	**Class 2**	**Class 3**
Class 1	125	0	0
Class 2	0	125	0
Class 3	0	0	125

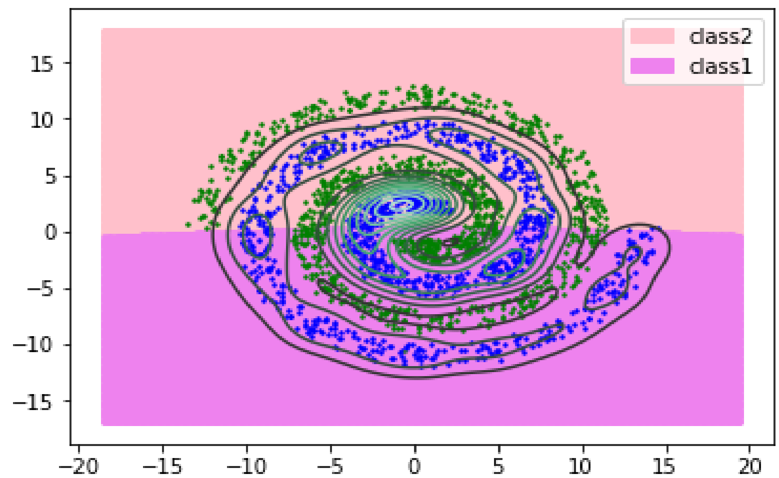
Analysis

	Class 1	**Class 2**	**Class 3**
Precision	1.000	1.000	1.000
Recall	1.000	1.000	1.000
F-measure	1.000	1.000	1.000

4.3.2 Non-Linear Data

Class 1 vs Class 2





Accuracy = 55.21%

Confusion Matrix

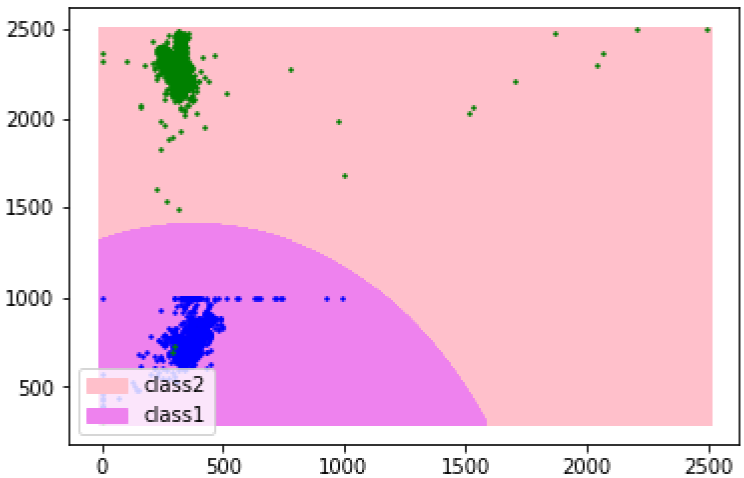
Actual/ Predicted	**Class 1**	**Class 2**
Class 1	180	146
Class 2	144	182

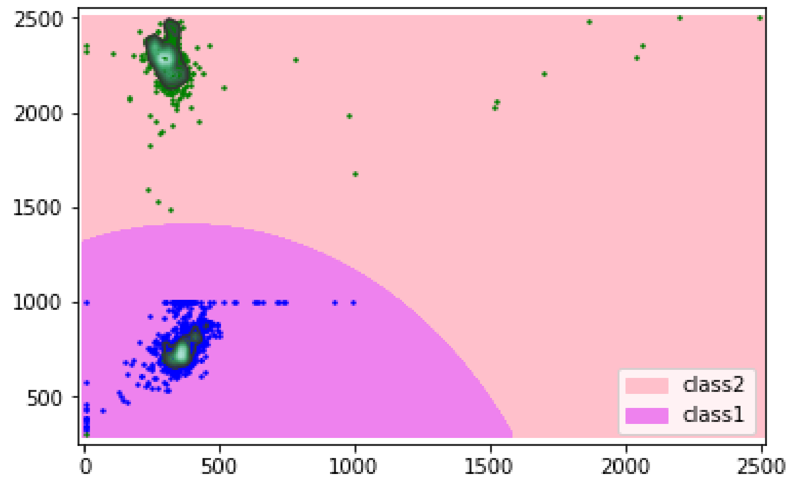
Analysis

	Class 1	**Class 2**
Precision	0.555	0.554
Recall	0.552	0.558
F-measure	0.553	0.556

4.3.3 Real World Data

Class 1 vs Class 2





Accuracy = 99.83%

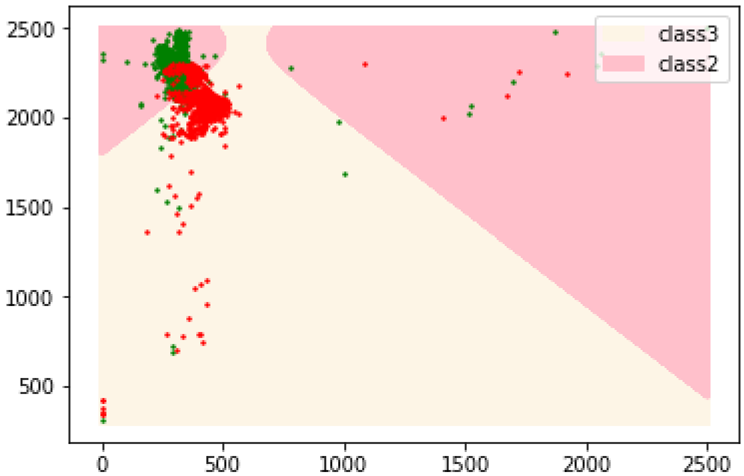
Confusion Matrix

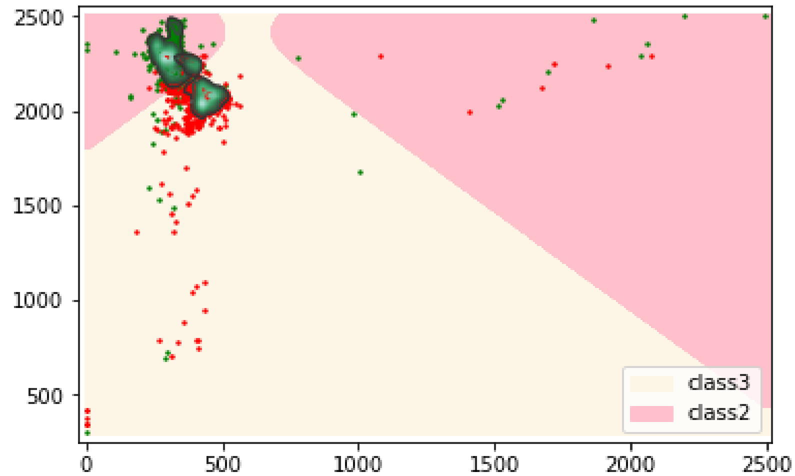
Actual/ Predicted	**Class 1**	**Class 2**
Class 1	622	0
Class 2	0	572

Analysis

	Class 1	**Class 2**
Precision	0.995	0.837
Recall	1.000	0.958
F-measure	0.997	0.893

Class 2 vs class 3





Accuracy = 88.37%

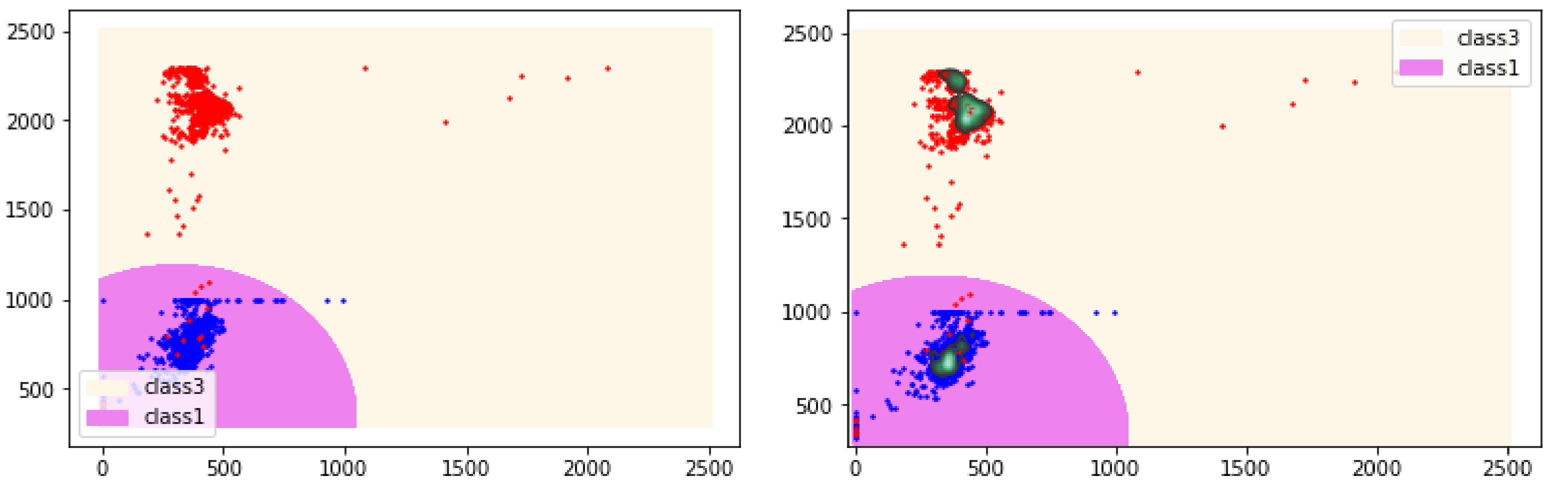
Confusion Matrix

Actual/ Predicted	**Class 2**	**Class 3**
Class 2	572	25
Class 3	111	462

Analysis

	Class 2	**Class 3**
Precision	0.837	0.948
Recall	0.958	0.806
F-measure	0.893	0.871

Class 1 vs Class 3



Accuracy = 99.74%

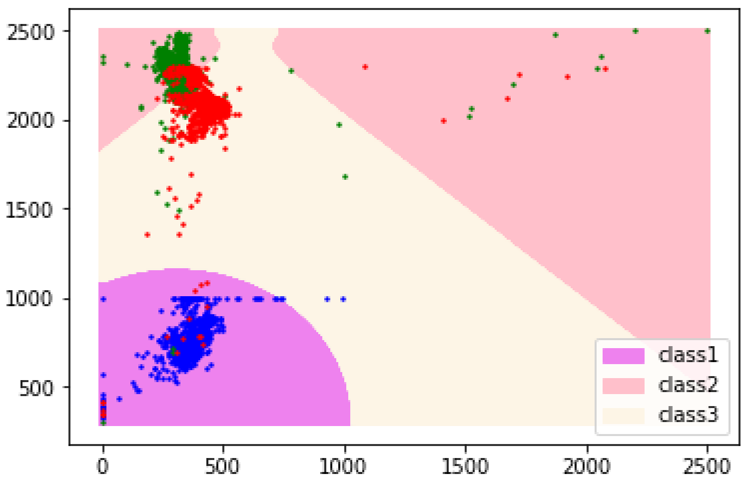
Confusion Matrix

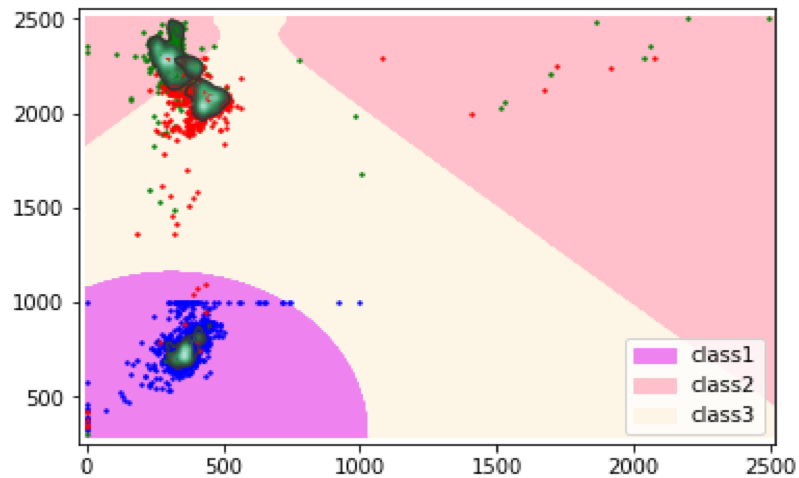
Actual/ Predicted	Class 1	Class 3
Class 1	622	0
Class 3	3	570

Analysis

	Class 1	Class 3
Precision	0.995	1.000
Recall	1.000	0.994
F-measure	0.997	0.997

Class 1 vs Class 2 vs class 3





Accuracy = 99.83%

Confusion Matrix

Actual/ Predicted	**Class 1**	**Class 2**	**Class 3**
Class 1	622	0	0
Class 2	0	572	25
Class 3	3	111	459

Analysis

	Class 1	**Class 2**	**Class 3**
Precision	0.998	0.924	0.922
Recall	1.000	0.958	0.801
F-measure	0.997	0.893	0.868

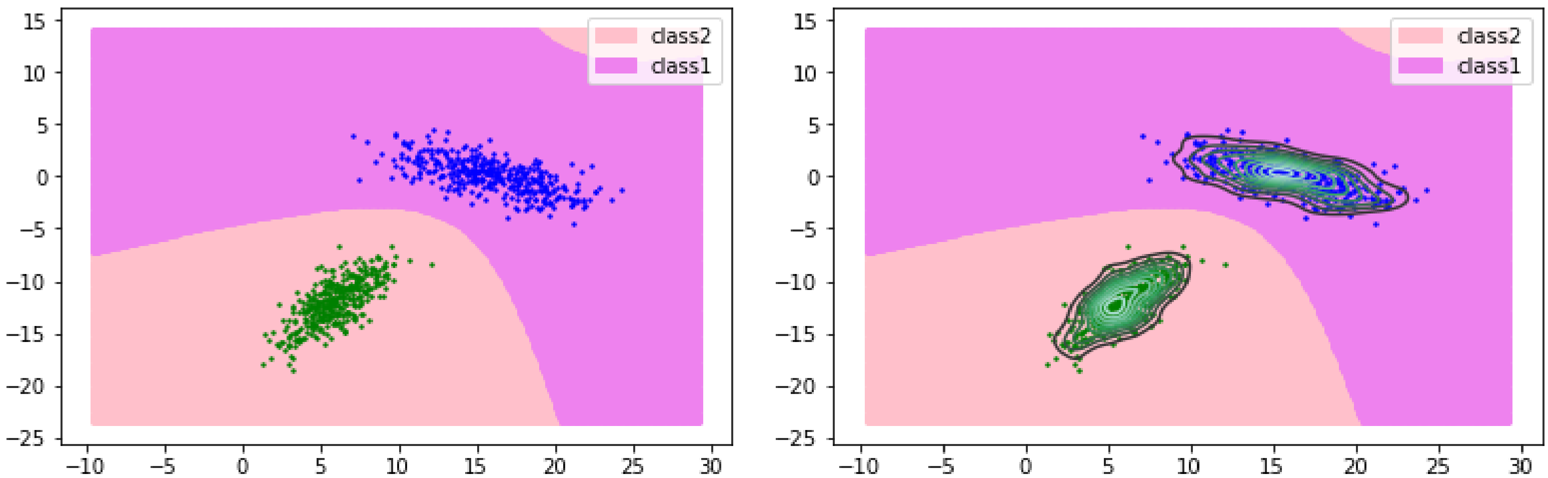
4.3.4 Inference

1. In this case we assume that our features are statistically independent which is why we consider all the classes matrices which are diagonal but are different for different classes.
2. We get the covariance matrix by making non diagonal elements of original covariance matrices zero for all classes.
3. This method works well for both linearly and nonlinearly separable data.
4. The decision boundary formed will be a curve of degree two whose shape can be predicted by observing the covariance matrices, that is the decision boundary curve will have concave surface towards the class having maximum variance in that direction and convex towards the other one.
5. This can be inferred from the first and third case where data is almost linearly separable the accuracy is 100% and in the second case where data is not linearly separable the accuracy is 55.21%.

4.4 Case 4 - Σ_i *is Unique*

4.4.1 Linear Data

Class 1 vs Class 2



Accuracy = 99.83%

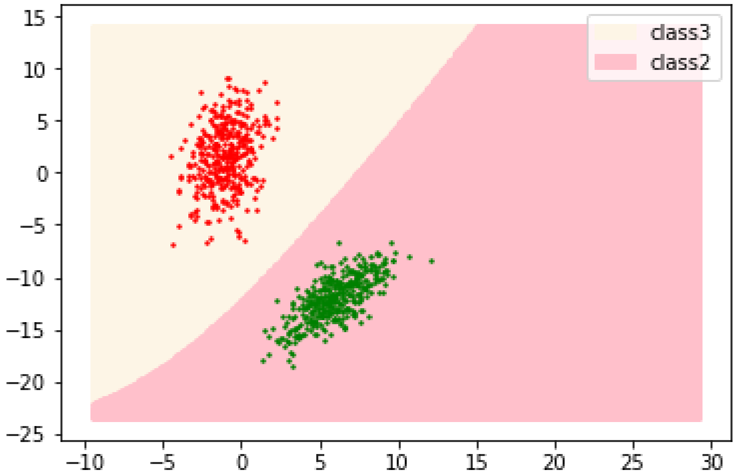
Confusion Matrix

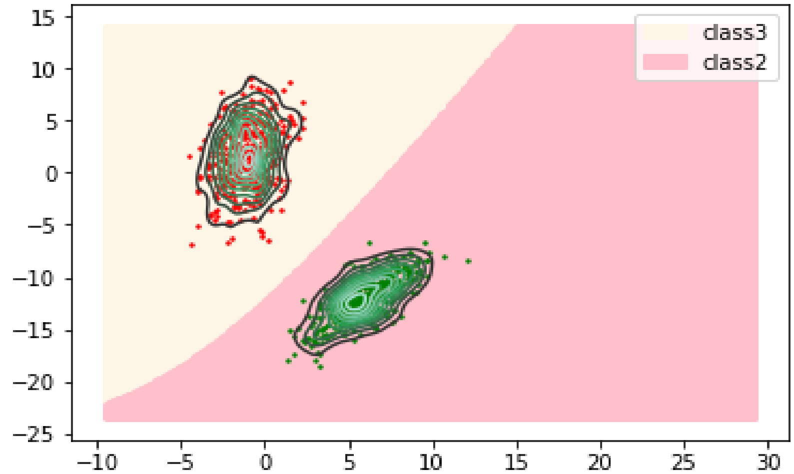
Actual/ Predicted	**Class 1**	**Class 2**
Class 1	622	0
Class 2	0	572

Analysis

	Class 1	**Class 2**
Precision	0.995	0.837
Recall	1.000	0.958
F-measure	0.997	0.893

Class 2 vs Class 3





Accuracy = 100%

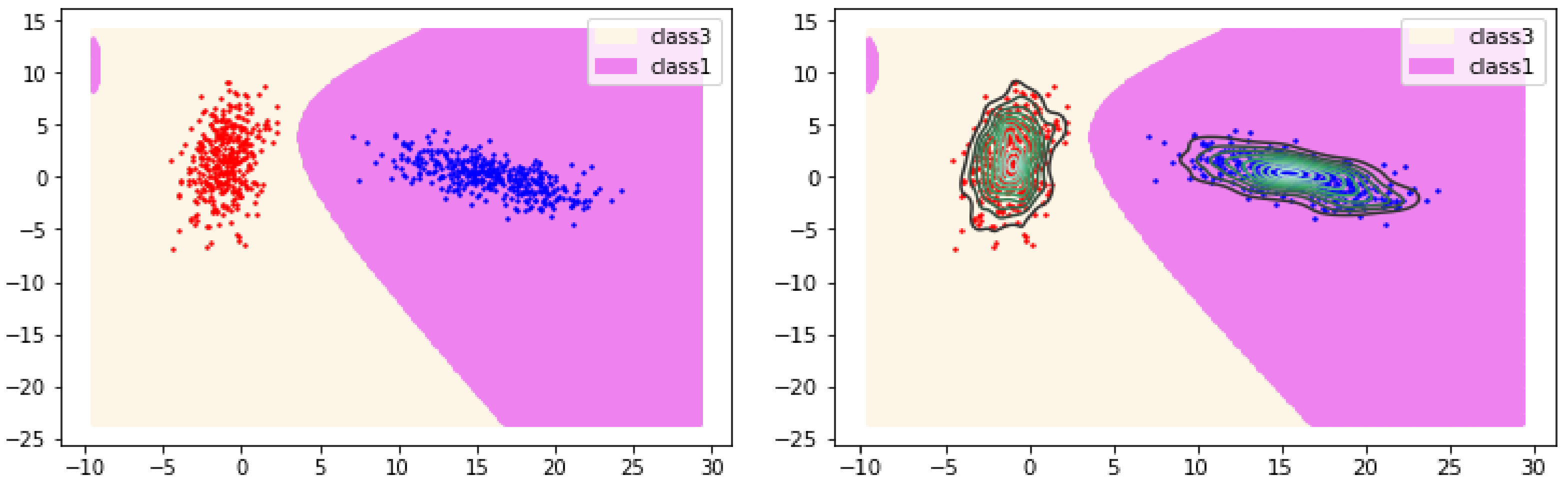
Confusion Matrix

Actual/ Predicted	**Class 2**	**Class 3**
Class 2	125	0
Class 3	0	125

Analysis

	Class 2	**Class 3**
Precision	1.000	1.000
Recall	1.000	1.000
F-measure	1.000	1.000

Class 1 vs Class 3



Accuracy = 100%

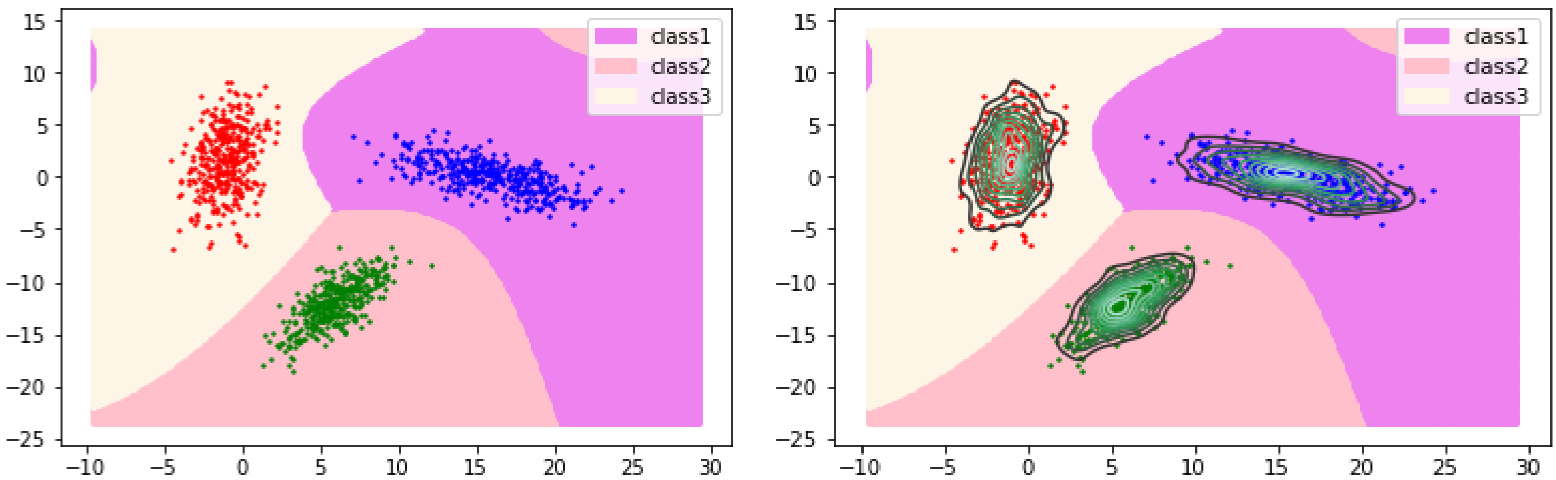
Confusion Matrix

Actual/ Predicted	**Class 1**	**Class 3**
Class 1	125	0
Class 3	0	125

Analysis

	Class 1	**Class 3**
Precision	1.000	1.000
Recall	1.000	1.000
F-measure	1.000	1.000

Class 1 vs Class 2 vs Class 3



Accuracy = 100%

Confusion Matrix

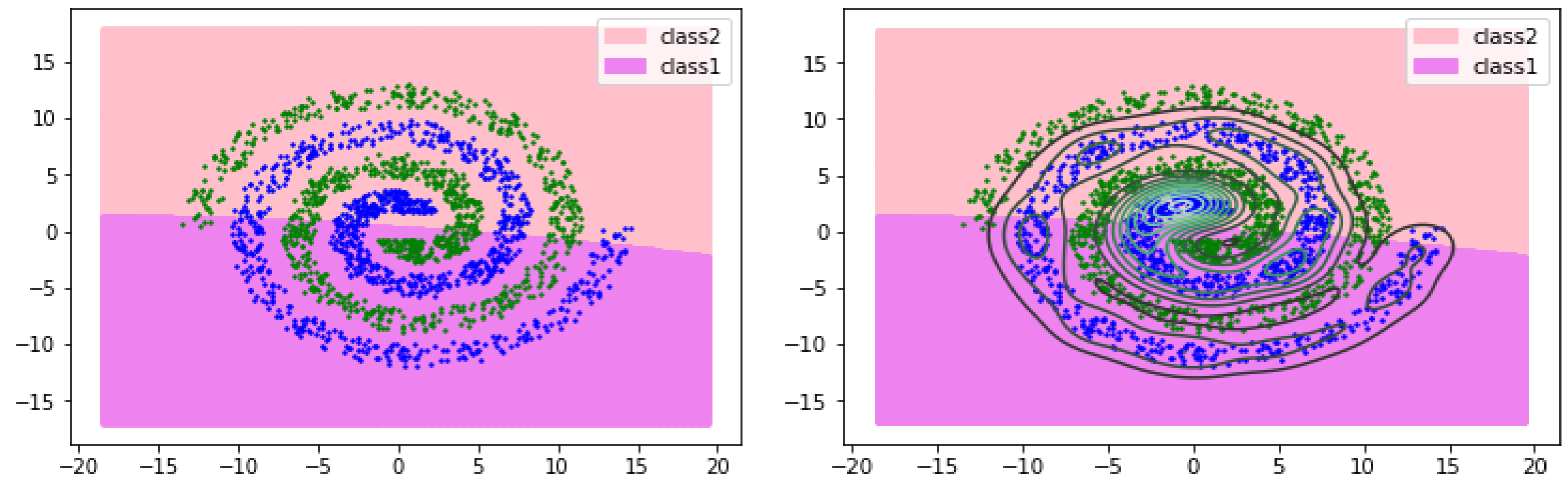
Actual/ Predicted	Class 1	Class 2	Class 3
Class 1	125	0	0
Class 2	0	125	0
Class 3	0	0	125

Analysis

	Class 1	Class 2	Class 3
Precision	1.000	1.000	1.000
Recall	1.000	1.000	1.000
F-measure	1.000	1.000	1.000

4.4.2 Non-Linear Data

Class 1 vs Class 2



Accuracy = 55.52%

Confusion Matrix

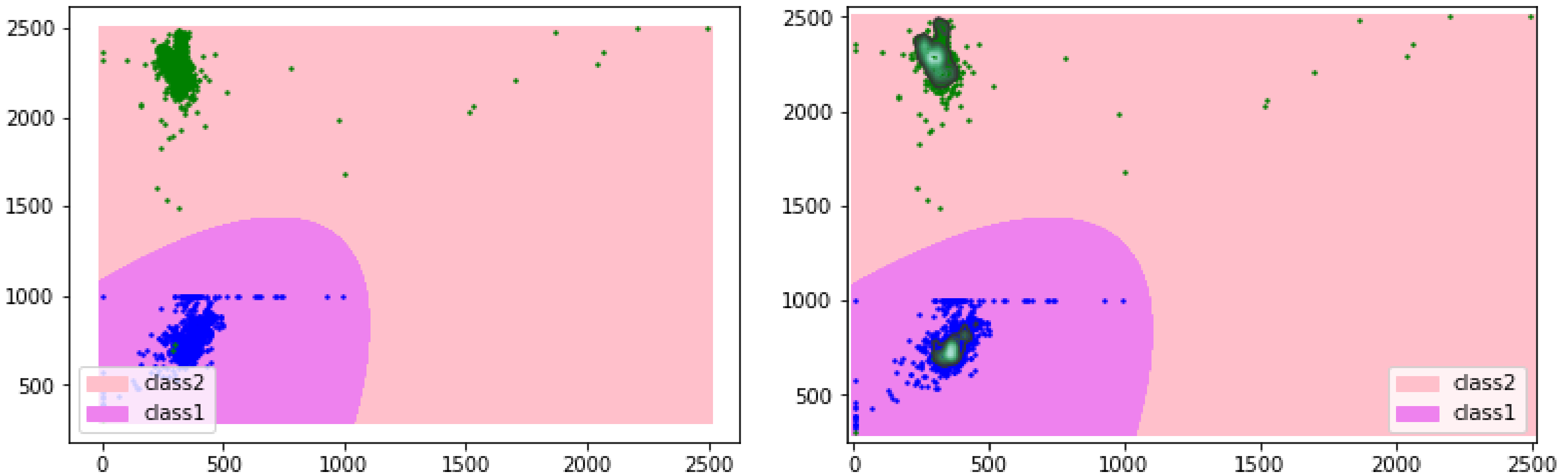
Actual/ Predicted	**Class 1**	**Class 2**
Class 1	181	145
Class 2	145	181

Analysis

	Class 1	**Class 2**
Precision	0.555	0.555
Recall	0.555	0.555
F-measure	0.555	0.555

4.4.3 Real World Data

Class 1 vs Class 2



Accuracy = 100%

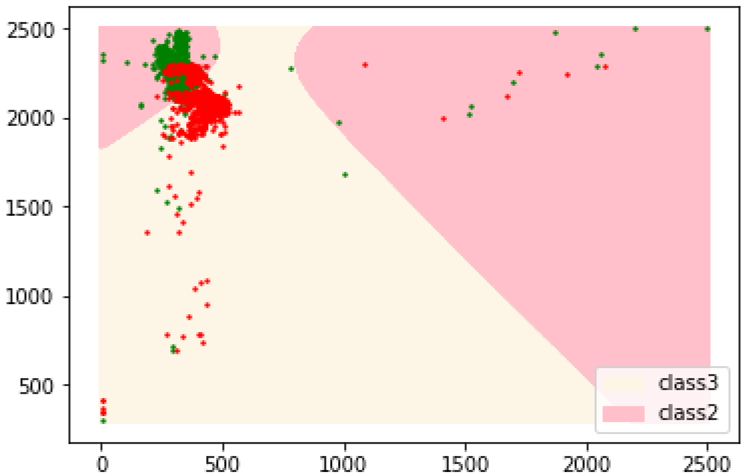
Confusion Matrix

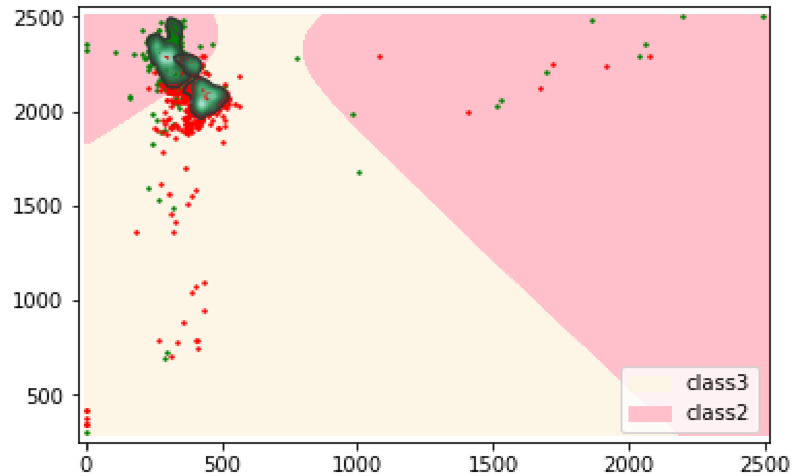
Actual/ Predicted	**Class 1**	**Class 2**
Class 1	622	0
Class 2	0	597

Analysis

	Class 1	**Class 2**
Precision	1.000	1.000
Recall	1.000	1.000
F-measure	1.000	1.000

Class 2 vs Class 3





Accuracy = 89.14%

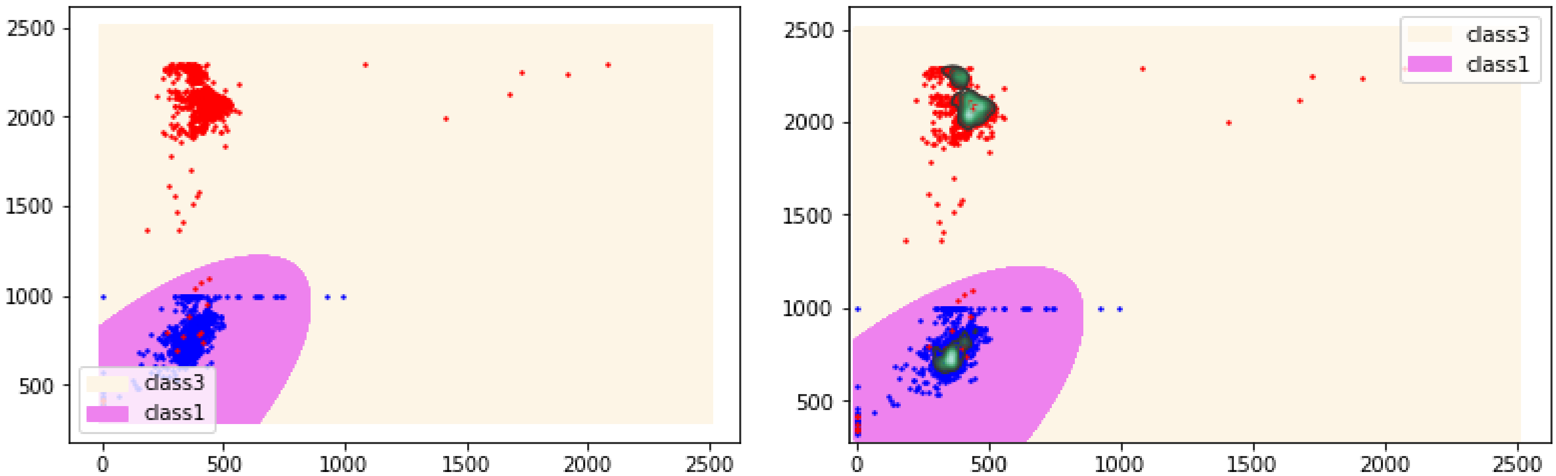
Confusion Matrix

Actual/ Predicted	**Class 2**	**Class 3**
Class 2	558	39
Class 3	88	485

Analysis

	Class 2	**Class 3**
Precision	0.863	0.925
Recall	0.934	.846
F-measure	0.897	0.884

Class 1 vs Class 3



Accuracy = 99.74%

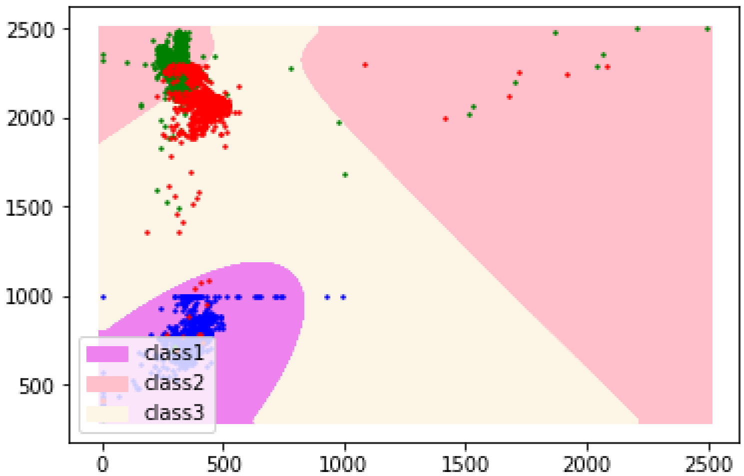
Confusion Matrix

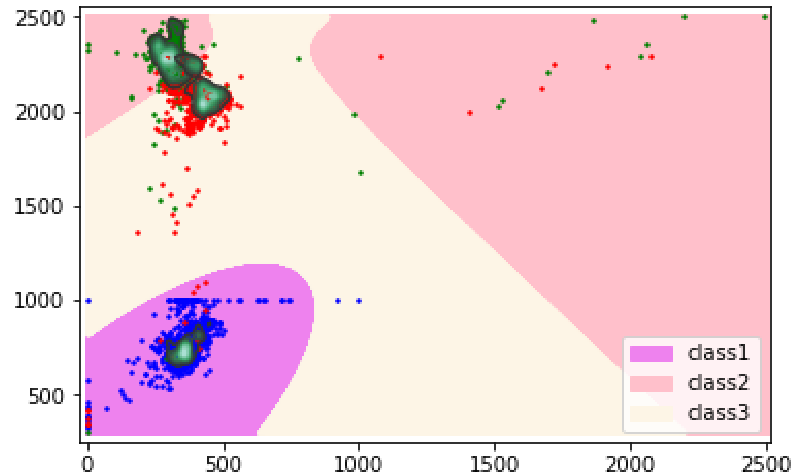
Actual/ Predicted	**Class 1**	**Class 3**
Class 1	621	1
Class 3	2	571

Analysis

	Class 1	**Class 3**
Precision	0.996	0.998
Recall	0.998	0.996
F-measure	0.997	0.997

Class 1 vs Class 2 vs Class 3





Accuracy = 99.83%

Confusion Matrix

Actual/ Predicted	Class 1	Class 2	Class 3
Class 1	621	0	0
Class 2	0	558	39
Class 3	2	88	483

Analysis

	Class 1	Class 2	Class 3
Precision	0.996	0.863	0.923
Recall	0.998	0.934	0.842
F-measure	0.997	0.897	0.881

4.4.4 Inference

1. In this case we consider original covariance matrices of all different classes.
2. This method works well for both linearly and nonlinearly separable data.
3. The decision boundary formed will be a curve of degree two whose shape can be predicted by observing the covariance matrices, that is the decision boundary curve will have concave surface towards the class having maximum variance in that direction and convex towards the other one.
4. This can be inferred from the first and third case where data is almost linearly separable the accuracy is 100% and in the second case where data is not linearly separable the accuracy is 55.52%.

5 Conclusion

1. Bayes classifier works well for the linearly separable data which can be observed from first three classes of data. It also does well in the real world data where the data is not overlapping.
2. But it fails in the non linearly separable case with accuracy as low as 55%. This can be observed from classes that data points are arranged in concentric spirals where the data is non linearly separable.
3. Bayes classifier can give the decision boundary of degree two at the maximum. As the spiral data requires higher degree boundary curve it gives very poor accuracy in bayes classifier.