PATTERN RECOGNITION

BAYES CLASSIFICATION

FOR BIVARIATE GAUSSIAN DISTRIBUTION of LINEARLY AND NONLINEARLY SEPARABLE CLASSES

GROUP - 3

Abhijeet Sharma (B15102)

Akhil Singhal (B15305)

Pranav Gupta (B15227)

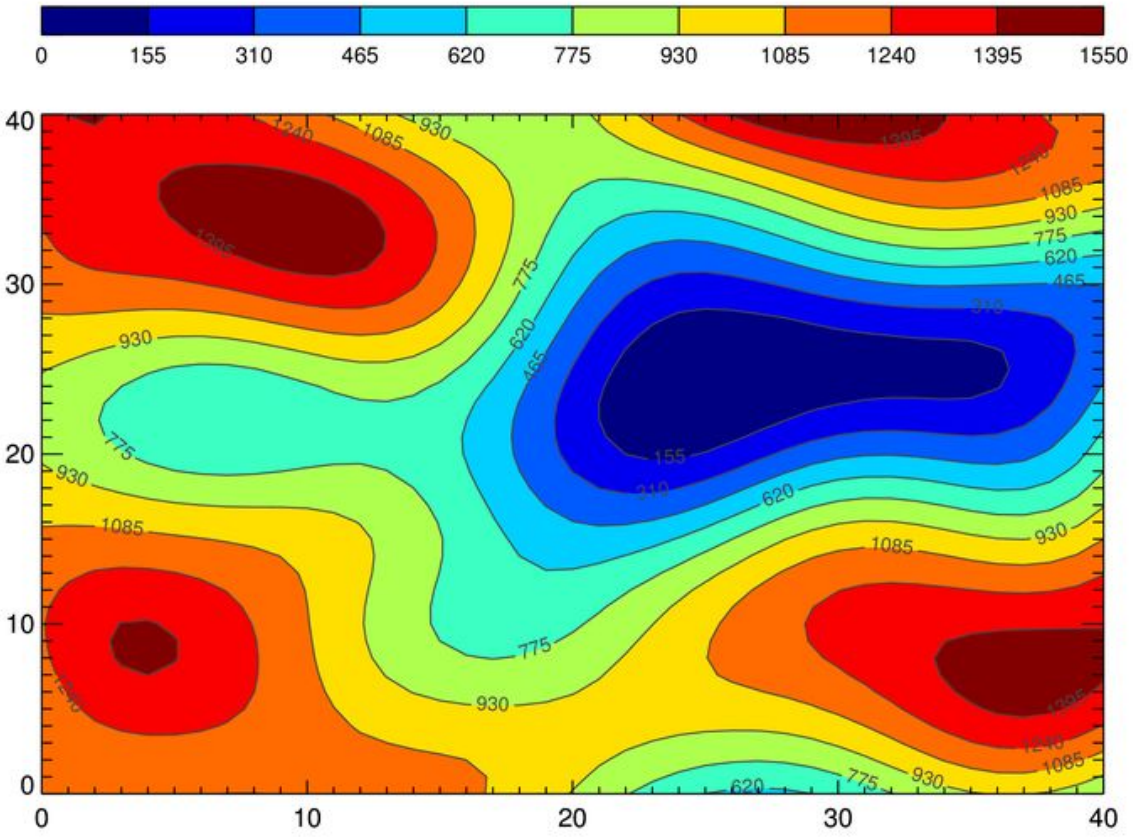


Image cited from images.gootle.com

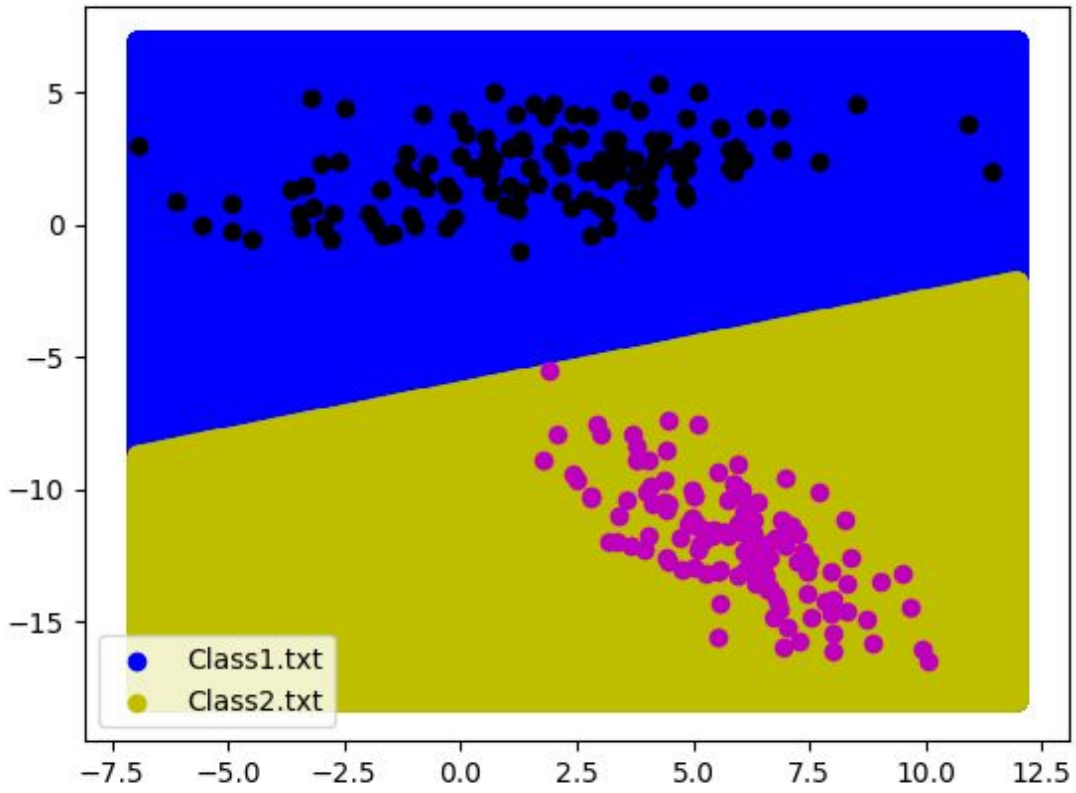
Introduction

The main objective of the assignment is to built Bayes Classifier from various given datasets which includes both linearly and nonlinearly separable data.For each dataset the task to generate Bayes classifier and Bayes decision boundary between each possible combination of classes with certain assumptions which can be listed as :

1) Covariance matrix for each class is same and is equal to $\boldsymbol{\sigma^2 I}$.For this we first take average of covariance matrix of all classes and then averaging the value of all variances.
2) Full Covariance matrix for all the classes is the same and is $\boldsymbol{\sigma}$.Here covariance matrix is obtained by taking average of all covariance matrices.
3) Covariance matrix is diagonal and is different for each class.
4) Full Covariance matrix for each class is different.

Additionally we also have to generate contour plots of different classes from discriminant functions of each class.

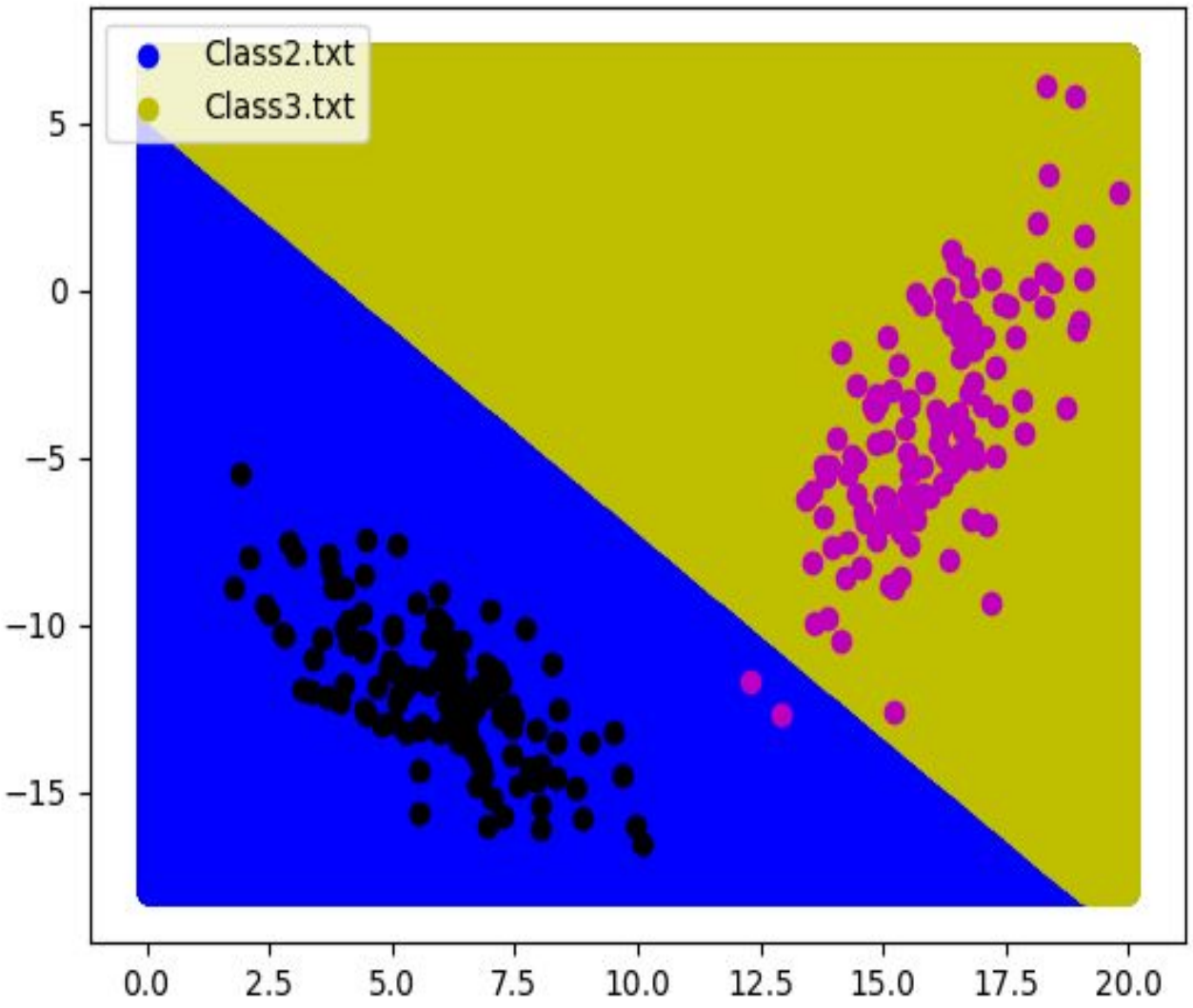
Linearly Separable Data Case 1



1) This case corresponds to define Bayes Classifier and Decision boundary between two classes ,when the classes are linearly separable.
2) In this case covariance matrix is $\boldsymbol{\sigma^2 I}$.
3) The dataset consisted of feature vector having two dimensions and these features are statistically independent.
4) As the dataset is linearly separable as well as our classifier is also linear machine so the results are quite great.

Confusion Matrix For Class1 Class2	
.250	0.1
11.0	124

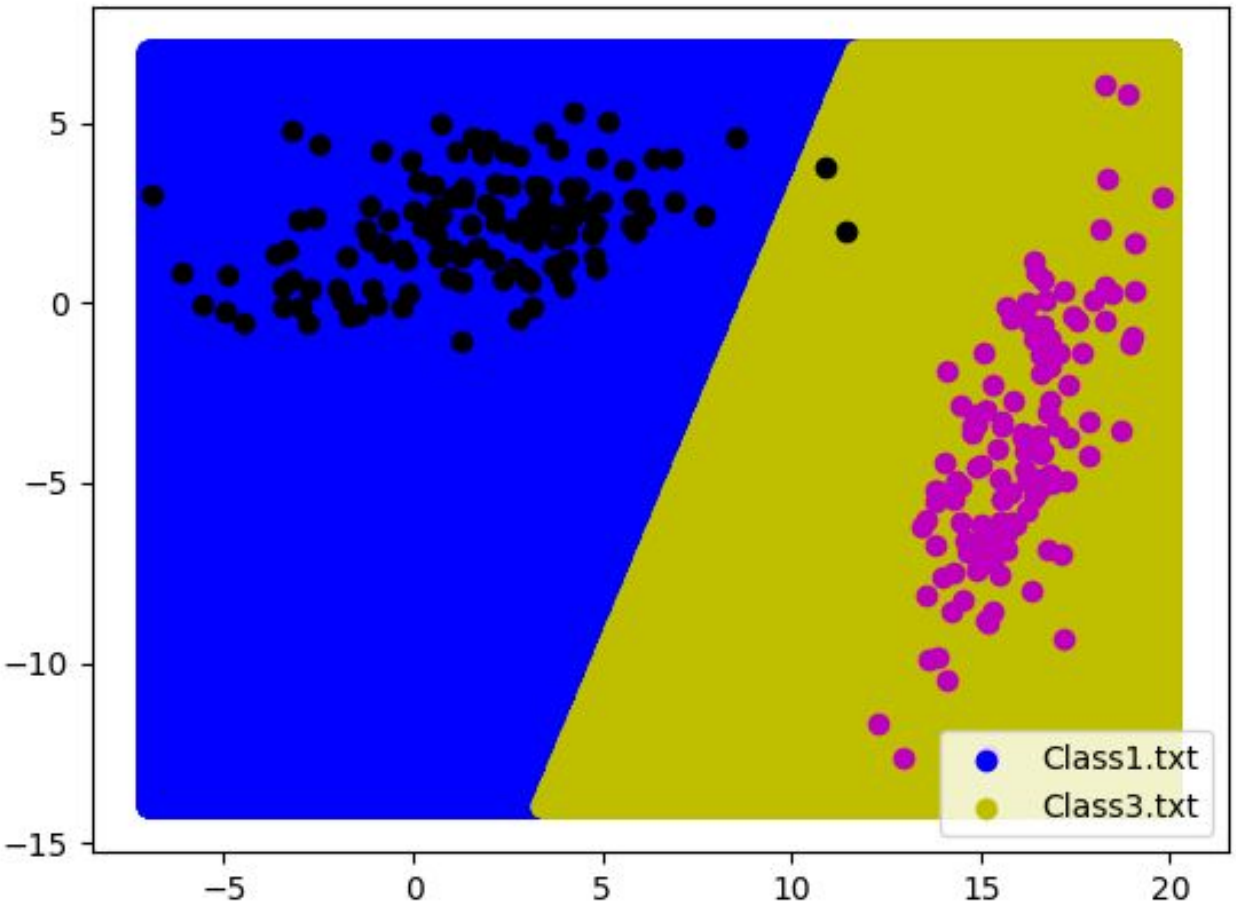
TITLE	VALUE
F-measure for Class1	0.996
F-measure for Class2	0.996
Accuracy of Classification	0.996
Mean recall	0.996
Mean precision	0.996
Mean F-measure	0.996
Recall for class Class1	0.992
Recall for class Class2	1.0
Precision for Class1	1.0
Precision for Class2	0.992



1) This plot also corresponds to Data Set which contains Linearly Separable Data.
2) The Decision boundary is a straight line.
3) As we can see that there is some error associated with the classifier which is shown as two data points which belong to pink class are classified as of black class.
4) Such details about the classifier are exposed in the below tables by various parameters.

Confusion Matrix For Class2 Class3	
125.0	0.0
2.0	123.0

TITLE	VALUE
F-measure for Class2	0.992
F-measure for Class3	0.992
Accuracy of Classification	0.992
Mean recall	0.992
Mean precision	0.992
Mean F-measure	0.992
Recall for Class2.	0.984
Recall for Class3	1.0
Precision for Class2	1.0
Precision for Class3	0.984

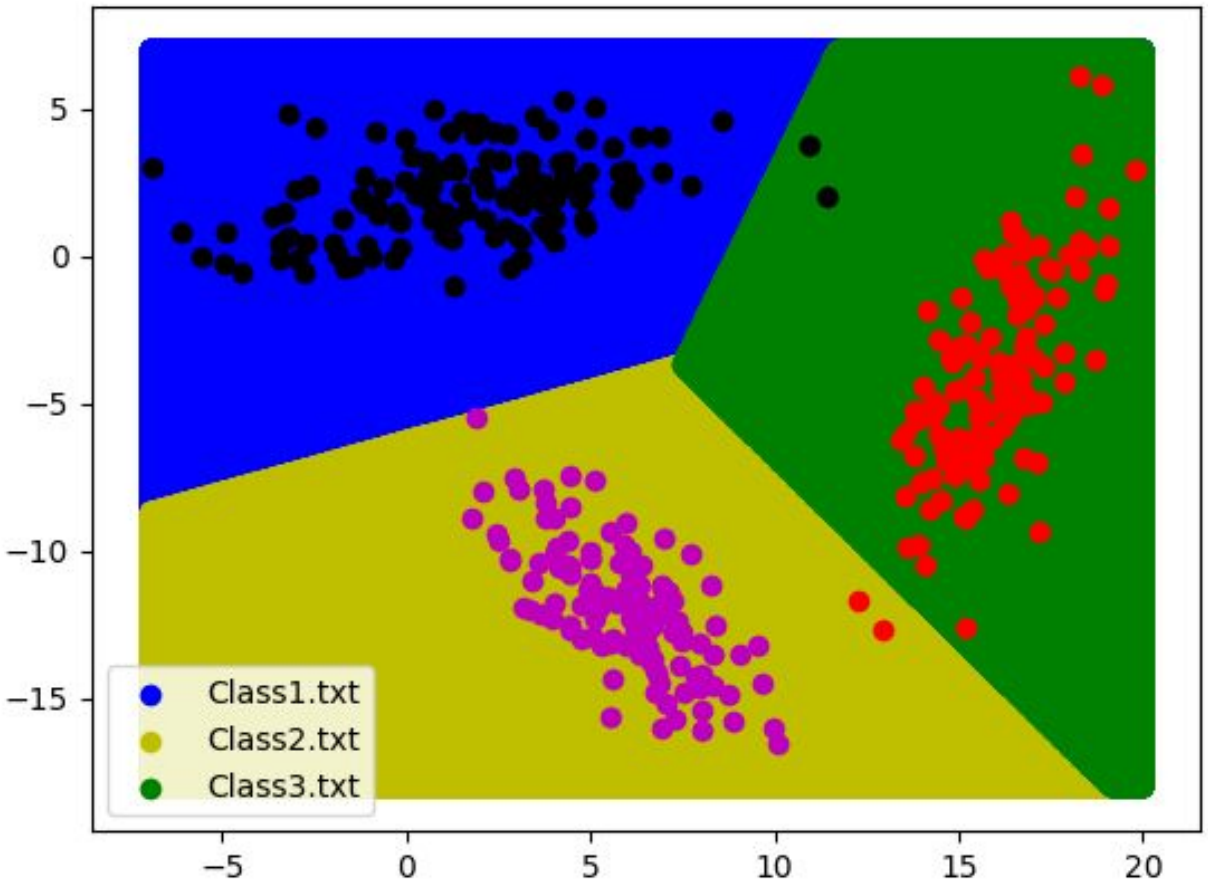


1) The 2D plot corresponds to data with feature vector having two dimensions.
2) Data is linearly separable and hence the decision boundary is linear.
3) There is some error associated with the classification as shown in the figure with wrong classification of two points.

GENERALIZATION FOR D-DIMENSIONS : If covariance matrices for two distributions are equal and proportional to the identity matrix, then the distributions are spherical in d-dimensions and the boundary is a generalized hyperplane in d-1 dimensions, perpendicular to the line joining the means.

Confusion Matrix For Class1 Class3	
123.0	2.0
0.0	125.0

TITLE	VALUE
F-measure for Class1	0.992
F-measure for Class3	0.992
Accuracy of Classification	0.992
Mean recall	0.992
Mean precision	0.992
Mean F-measure	0.992
Recall for Class1	0.984
Recall for Class3	1.0
Precision for Class1	1.0
Precision for Class3	0.984

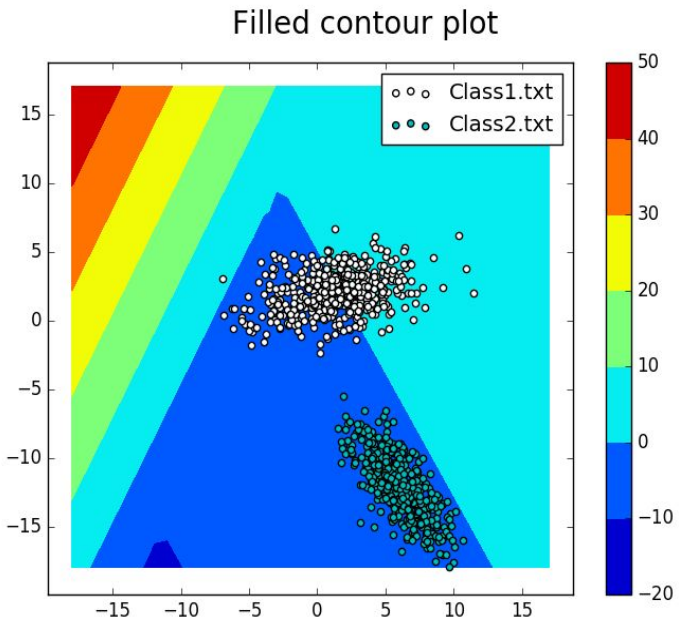


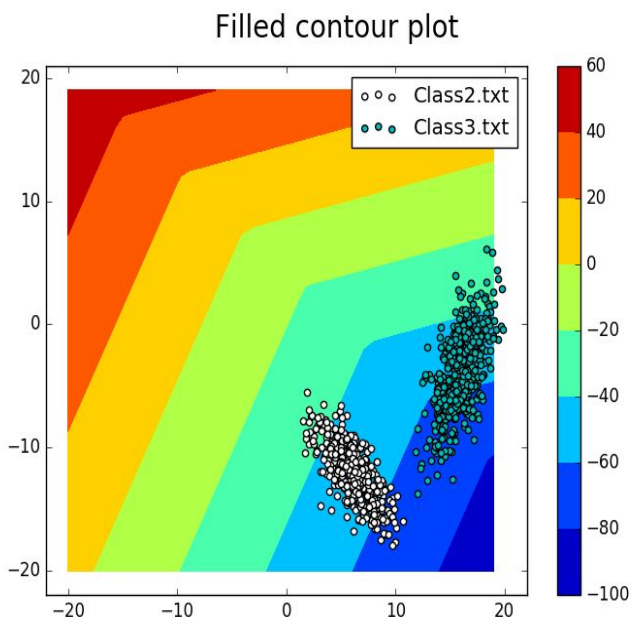
This plot shows formulation of decision boundary between three classes simultaneously and here also the data is linearly separable.

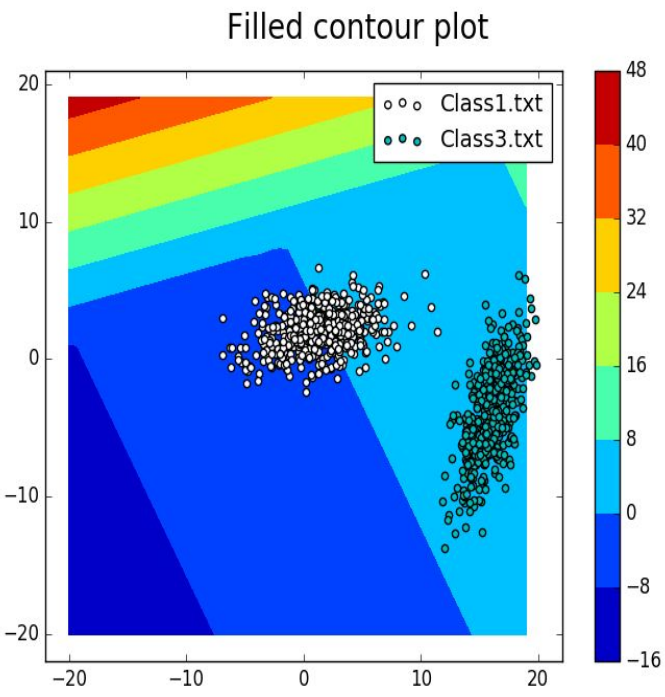
Confusion Matrix For Class1 Class2 Class3		
123.0	0.0	2.0
1.0	124.0	0.0
0.0	2.0	123.0

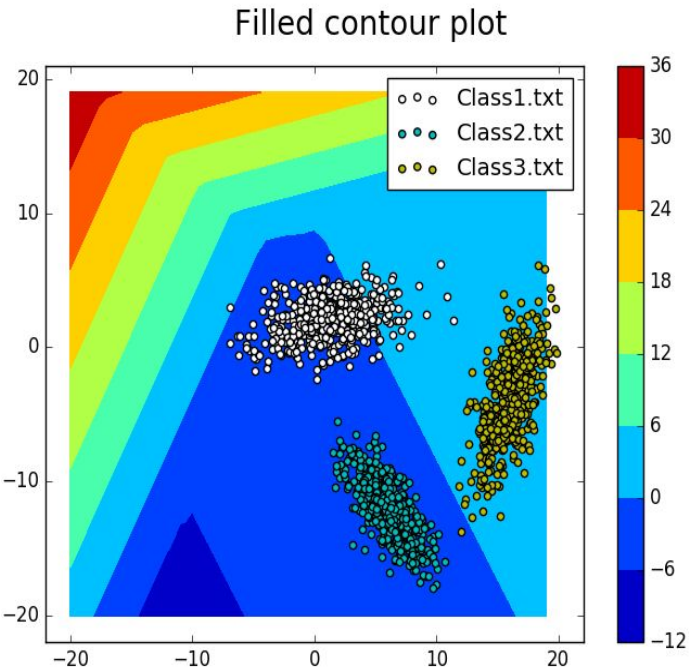
TITLE	VALUE
F-measure for Class1	0.988
F-measure for Class2	0.988
F-measure for Class3	0.984
Accuracy of Classification	0.987
Mean recall	0.987
Mean precision	0.987
Mean F-measure	0.987
Recall for Class1	0.992
Recall for Class2	0.984
Recall for Class3	0.984
Precision for Class1	0.984
Precision for Class2	0.992
Precision for Class3	0.984

Linearly Separable data Case 1 Contour

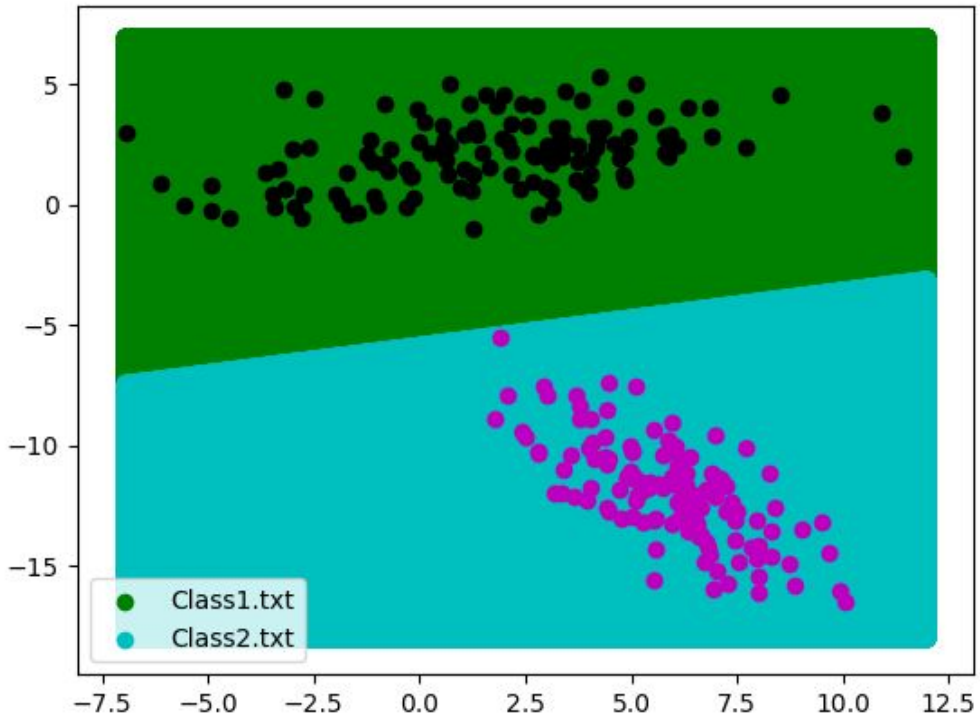








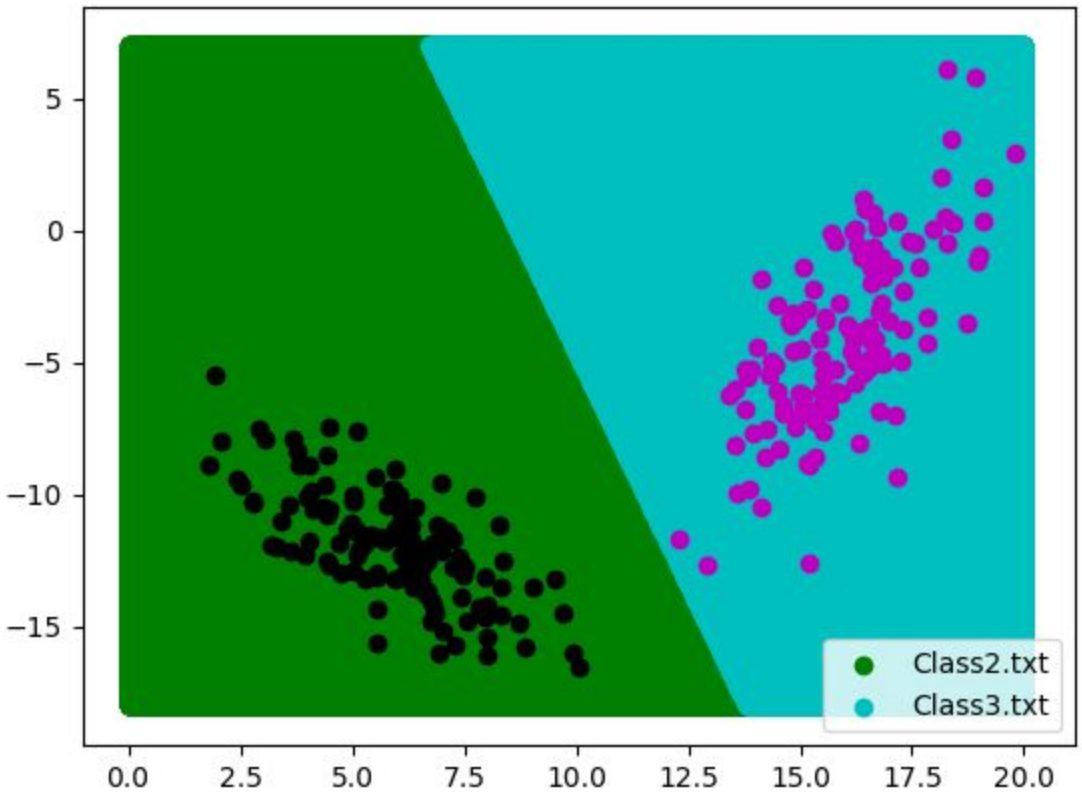
Linearly Separable Data Case 2 : Full covariance matrix for each class is same and is Σ.



1) This case corresponds to defining Bayes Classifier and decision boundary between two classes which are linearly separable but here the covariance matrix for all classes is considered same (by averaging all the covariance matrices of all classes) but it is arbitrary.
2) Decision boundary is this case is also a line which bisects the line joining the centre of means of two classes but unlike the previous case the **decision boundary is not orthogonal** to the line joining means of two classes.
3) It is more refined than the previous case as it is more flexible and gives better result like in the above figure the decision boundary perfectly separates data of two classes.
4) The several features of a feature vector are not statistically independent in this case.

Confusion Matrix For Class1 Class2	
125.0	0.0
0.0	125.0

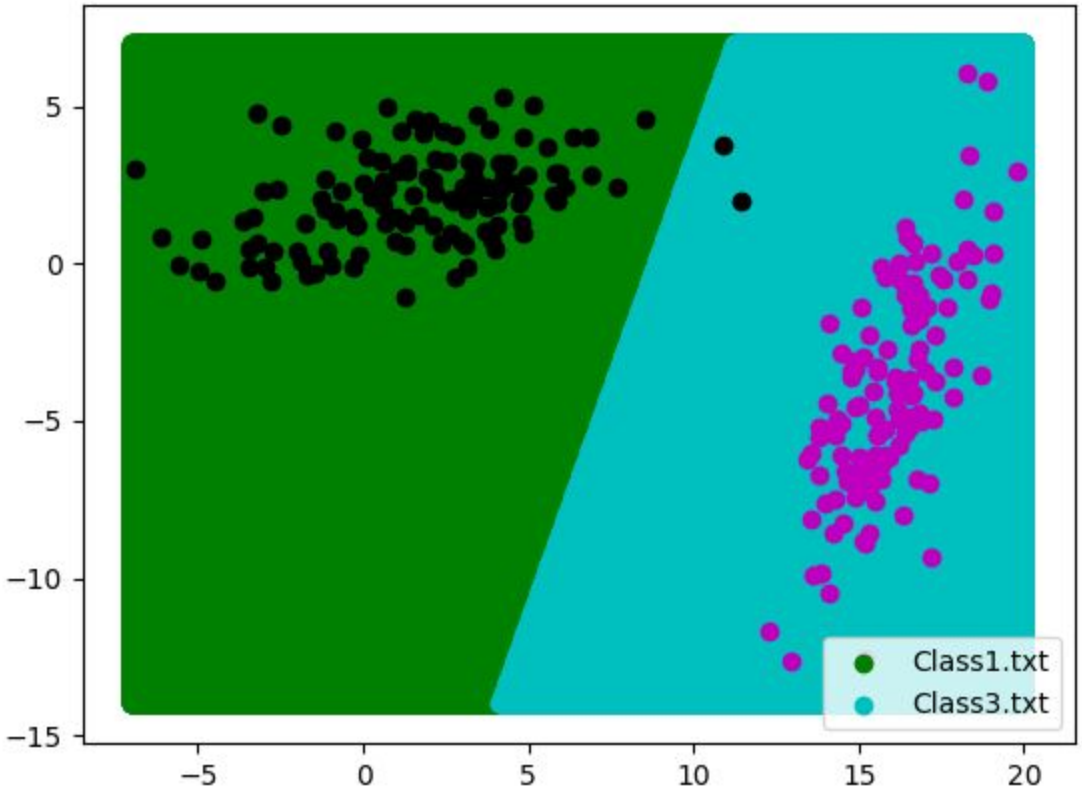
TITLE	VALUE
F-measure for Class1	1.0
F-measure for Class2	1.0
Accuracy of Classification	1.0
Mean recall	1.0
Mean precision	1.0
Mean F-measure	1.0
Recall for class Class1	1.0
Recall for class Class2	1.0
Precision for Class1	1.0
Precision for Class2	1.0



1) This plot corresponds to data set which contains linearly separable data.
2) The decision boundary is straight line.
3) As we can see that there is some error associated with the classifier which is shown as two data points which belong to pink class are classified as of black class.
4) These results can be improved if we apply a nonlinear classifier on this same dataset.

Confusion Matrix For Class2 Class3	
125.0	0.0
2.0	123.0

TITLE	VALUE
F-measure for Class2	1.0
F-measure for Class3	1.0
Accuracy of Classification	1.0
Mean recall	1.0
Mean precision	1.0
Mean F-measure	1.0
Recall for Class2.	1.0
Recall for Class3	1.0
Precision for Class2	1.0
Precision for Class3	1.0



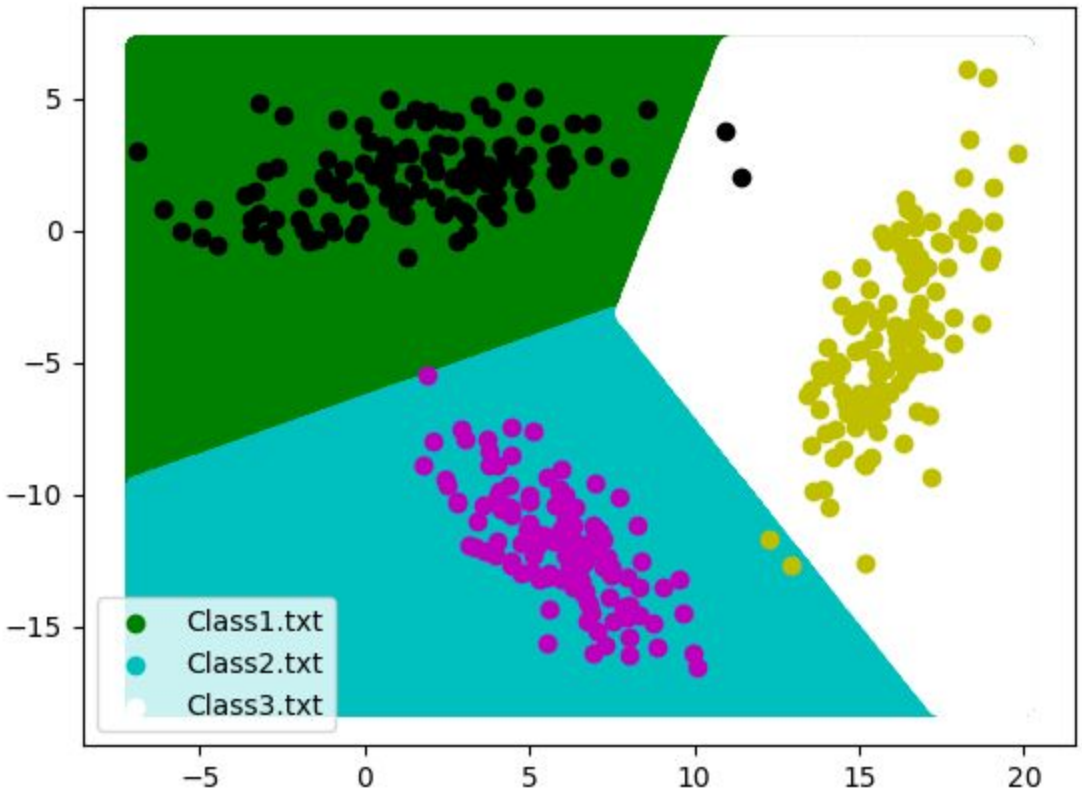
1) Data consists of features with feature vector having two dimensions.
2) Data is linearly separable and hence the decision boundary is linear.
3) There is some error associated with the classification as shown in the figure with wrong classification of two points.

GENERALIZATION FOR D-DIMENSIONS :

If the covariance matrix is arbitrary but is same for all the classes then in d-dimensional space (when feature vector will have d dimensions) the decision hyperplane will not be orthogonal to the line joining their means.

Confusion Matrix For Class1 Class3	
123.0	2.0
0.0	125.0

TITLE	VALUE
F-measure for Class1	0.992
F-measure for Class3	0.992
Accuracy of Classification	0.992
Mean recall	0.992
Mean precision	0.992
Mean F-measure	0.992
Recall for Class1	1.0
Recall for Class3	0.984
Precision for Class1	0.984
Precision for Class3	1.0

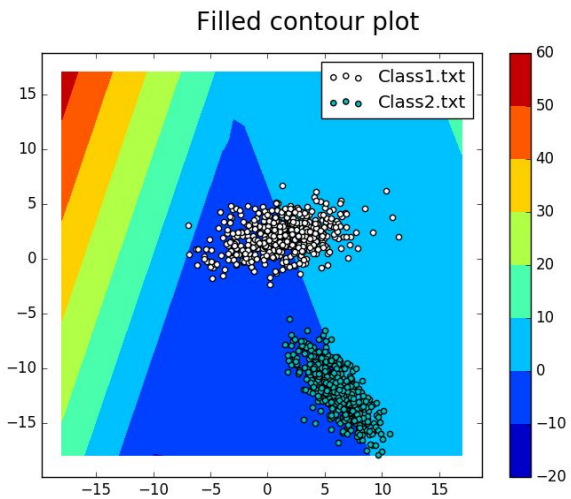


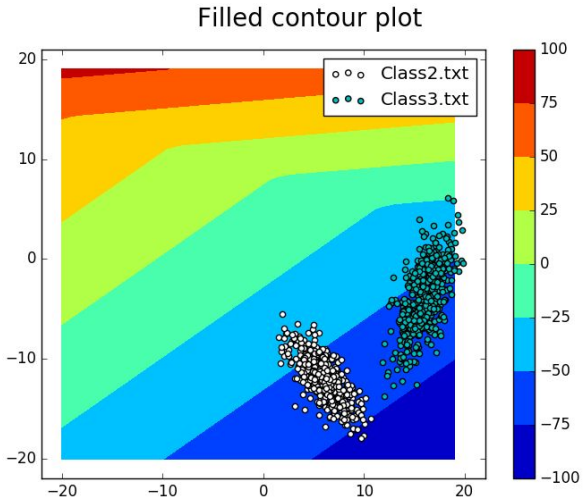
1) This plot shows formulation of decision boundary between all the classes simultaneously.
2) Confusion matrix is shown below which shows the accuracy of prediction of this classifier.

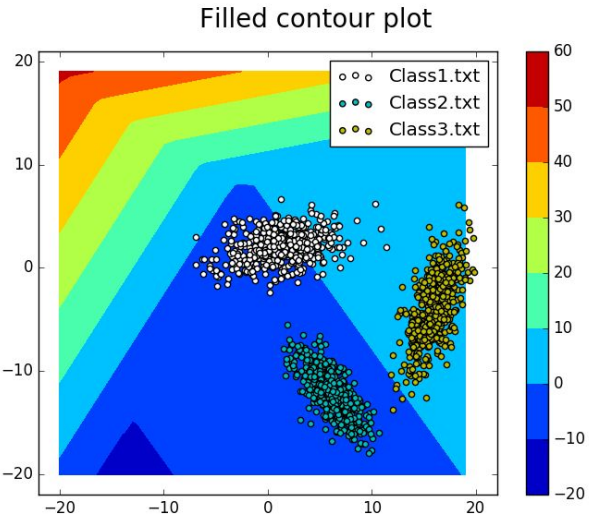
Confusion Matrix For Class1 Class2 Class3		
123.0	**0.0**	**2.0**
1.0	**124.0**	**0.0**
0.0	**2.0**	**123.0**

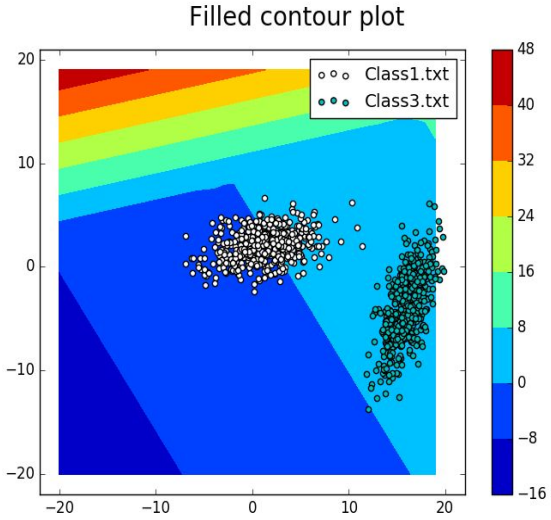
TITLE	VALUE
F-measure for Class1	0.988
F-measure for Class2	0.988
F-measure for Class3	0.984
Accuracy of Classification	0.987
Mean recall	0.987
Mean precision	0.987
Mean F-measure	0.987
Recall for Class1	0.992
Recall for Class2	0.984
Recall for Class3	0.984
Precision for Class1	0.984
Precision for Class2	0.992
Precision for Class3	0.984

Linearly Separable Data Case 2 Contour

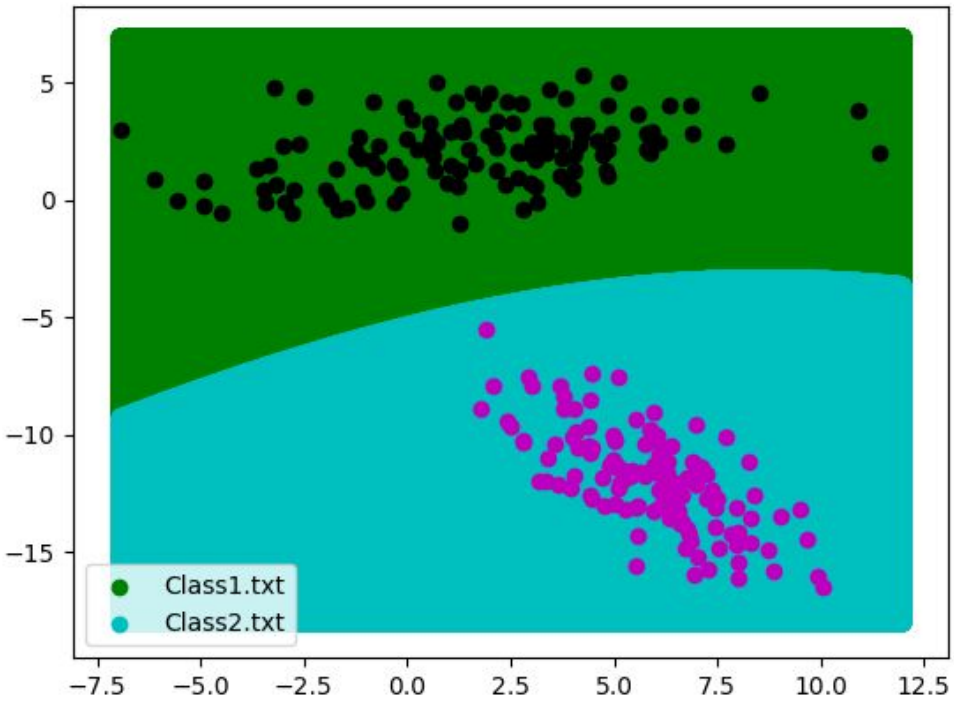








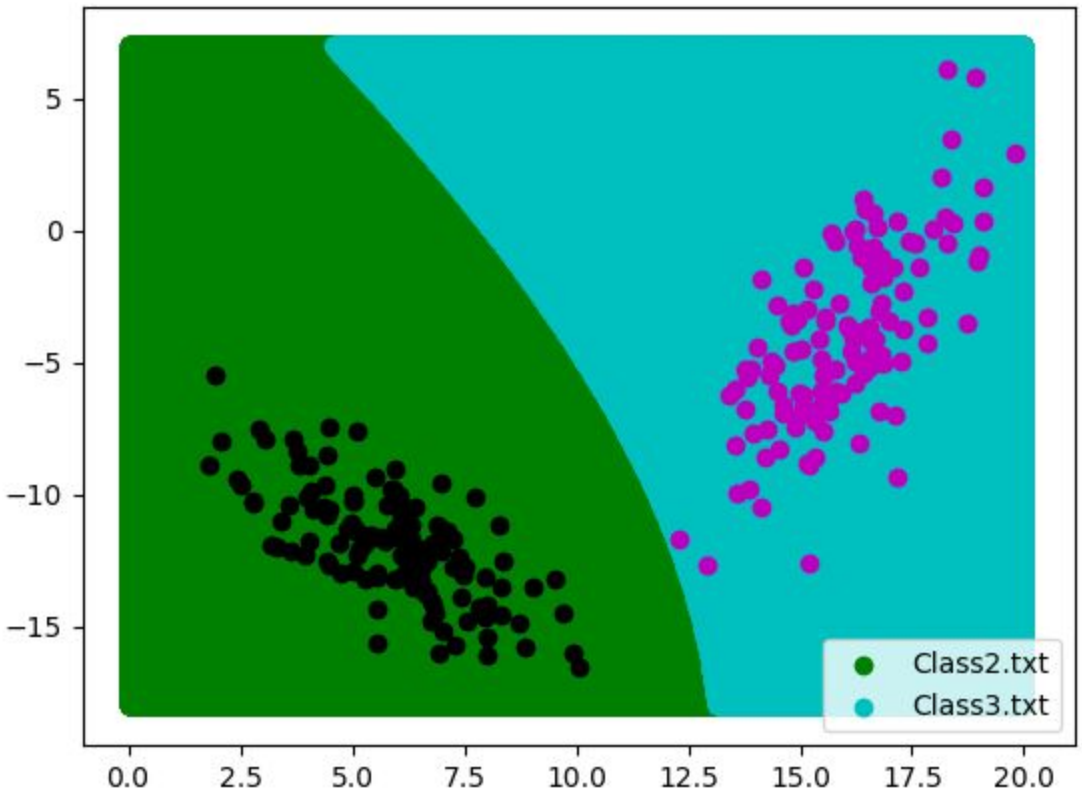
Linearly Separable Data Case 3 : Covariance Matrix is diagonal and different for each class.



1) Here we consider the case when covariance matrix is diagonal but is different for different classes.
2) As the covariance matrix is diagonal we can say the the features involved in feature vector are statistically independent.
3) Decision boundary between the classes is nonlinear and specifically by looking at the discriminant function we can infer that it is quadratic which is verified from the above plot also.
4) As Dataset 1 consisted of data which was linearly separable but defining a nonlinear boundary gave us better results in terms of accuracy but involves a bit more computation.

Confusion Matrix For Class1 Class2	
125.0	0.0
0.0	125.0

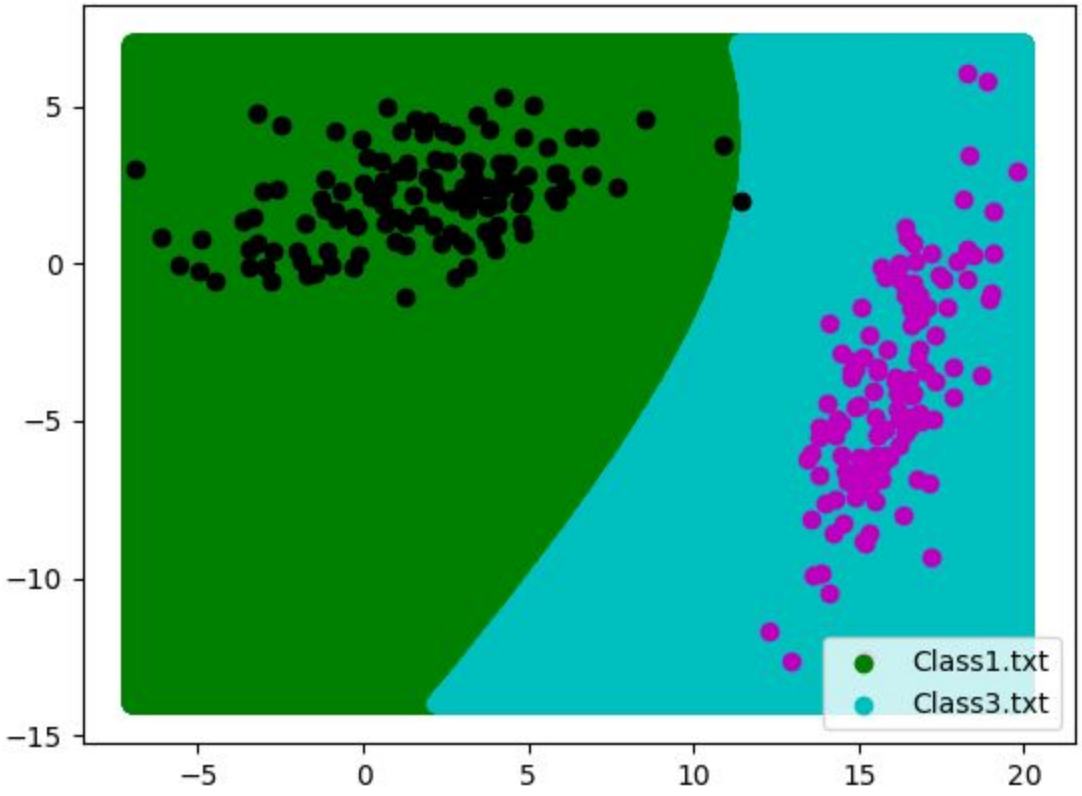
TITLE	VALUE
F-measure for Class1	1.0
F-measure for Class2	1.0
Accuracy of Classification	1.0
Mean recall	1.0
Mean precision	1.0
Mean F-measure	1.0
Recall for class Class1	1.0
Recall for class Class2	1.0
Precision for Class1	1.0
Precision for Class2	1.0



5) This plot corresponds to Dataset 1 in which it was given that dataset is linearly separable and we applied a nonlinear decision boundary over it.
6) As we can see that it acts better than linear machine in performance and for the above case it gives full accuracy but as the data distribution becomes more complicated like a parabolic distribution then this type of decision boundary will not suffice.
7) Various details about the classifier are exposed in the below tables by various parameters.

Confusion Matrix For Class2 Class3	
125.0	0.0
2.0	123.0

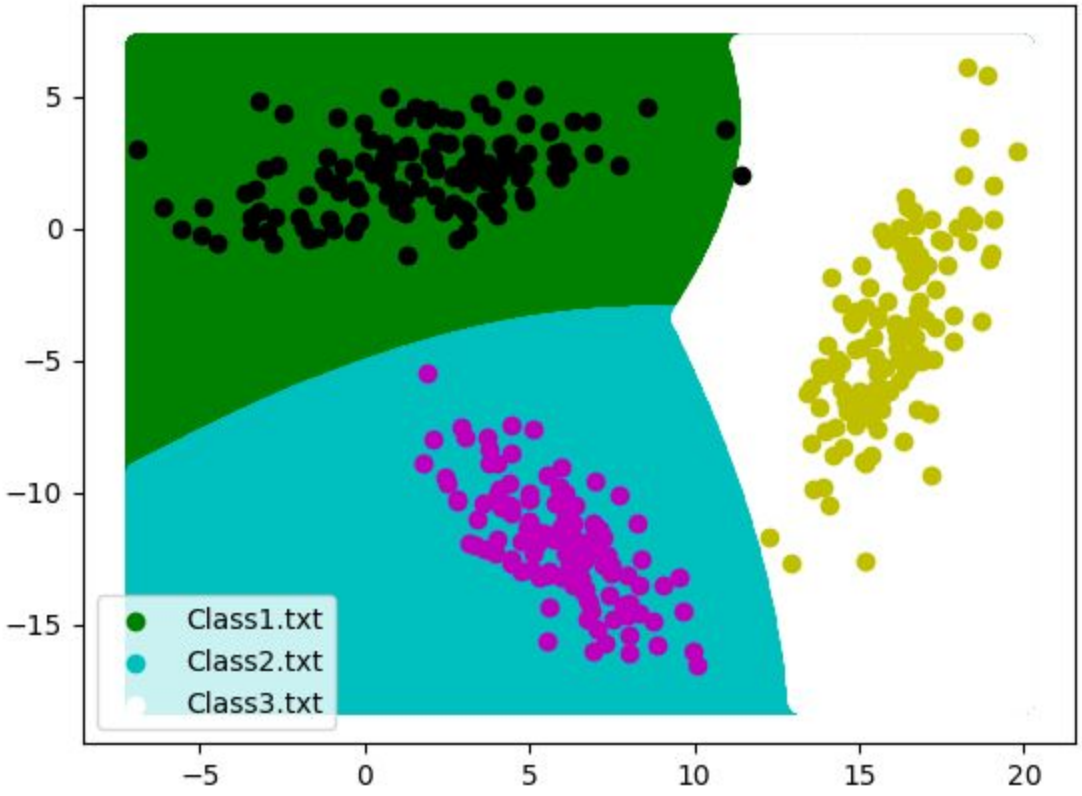
TITLE	VALUE
F-measure for Class2	1.0
F-measure for Class3	1.0
Accuracy of Classification	1.0
Mean recall	1.0
Mean precision	1.0
Mean F-measure	1.0
Recall for Class2.	1.0
Recall for Class3	1.0
Precision for Class2	1.0
Precision for Class3	1.0



1) This is also a case of data with feature vector having two dimensions.
2) Data is linearly separable and hence the decision boundary is linear.
3) There is some error associated with the classification as shown in the figure with wrong classification of two points.

Confusion Matrix For Class1 Class3	
125.0	0.0
0.0	125.0

TITLE	VALUE
F-measure for Class1	1.0
F-measure for Class3	1.0
Accuracy of Classification	1.0
Mean recall	1.0
Mean precision	1.0
Mean F-measure	1.0
Recall for Class1	1.0
Recall for Class3	1.0
Precision for Class1	1.0
Precision for Class3	1.0

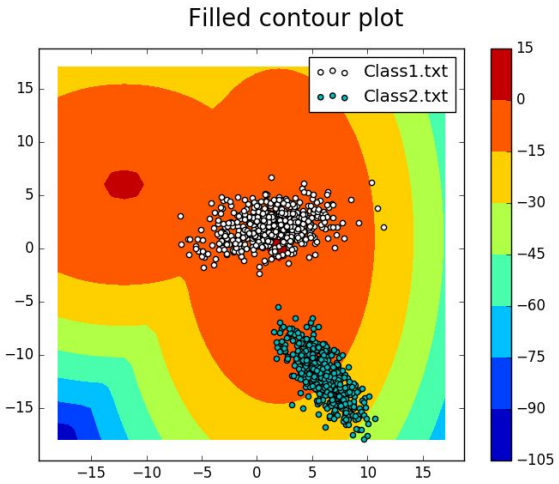


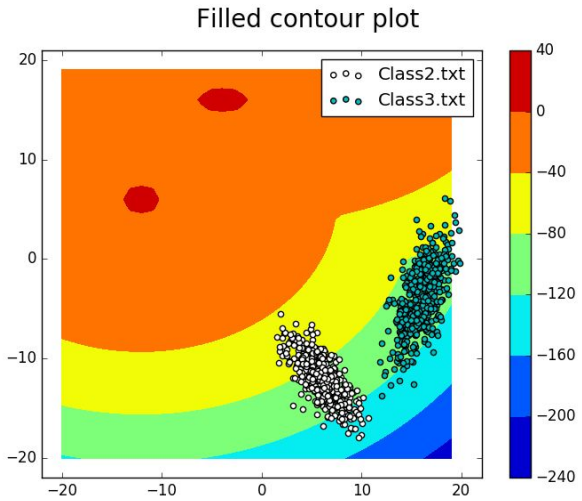
This plot shows formulation of decision boundary between three classes simultaneously.

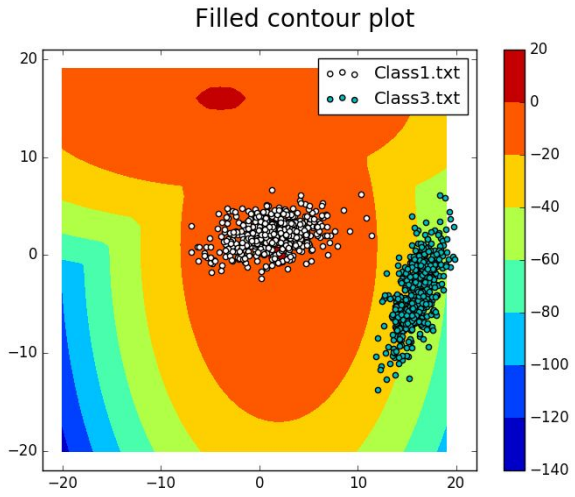
Confusion Matrix For Class1 Class2 Class3		
123.0	0.0	2.0
1.0	124.0	0.0
0.0	2.0	123.0

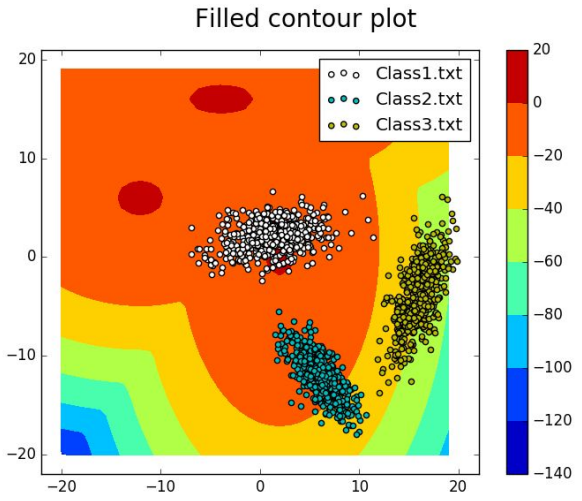
TITLE	VALUE
F-measure for Class1	1.0
F-measure for Class2	1.0
F-measure for Class3	1.0
Accuracy of Classification	1.0
Mean recall	1.0
Mean precision	1.0
Mean F-measure	1.0
Recall for Class1	1.0
Recall for Class2	1.0
Recall for Class3	1.0
Precision for Class1	1.0
Precision for Class2	1.0
Precision for Class3	1.0

Linearly Separable data Case 3 Contour

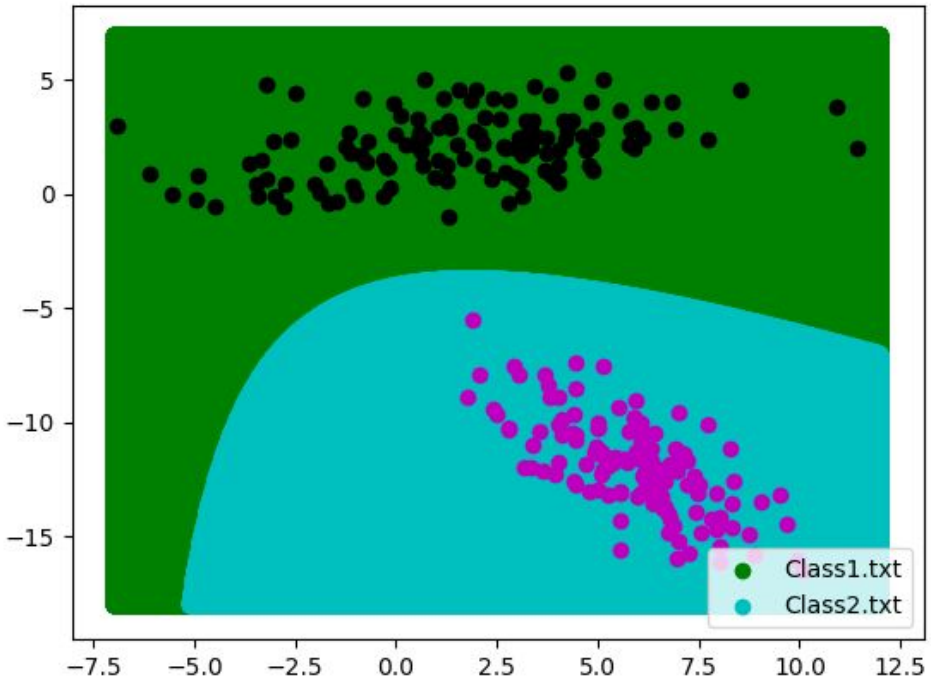








Linearly Separable Data Case 4 : Full Covariance Matrix is Different for Each Class

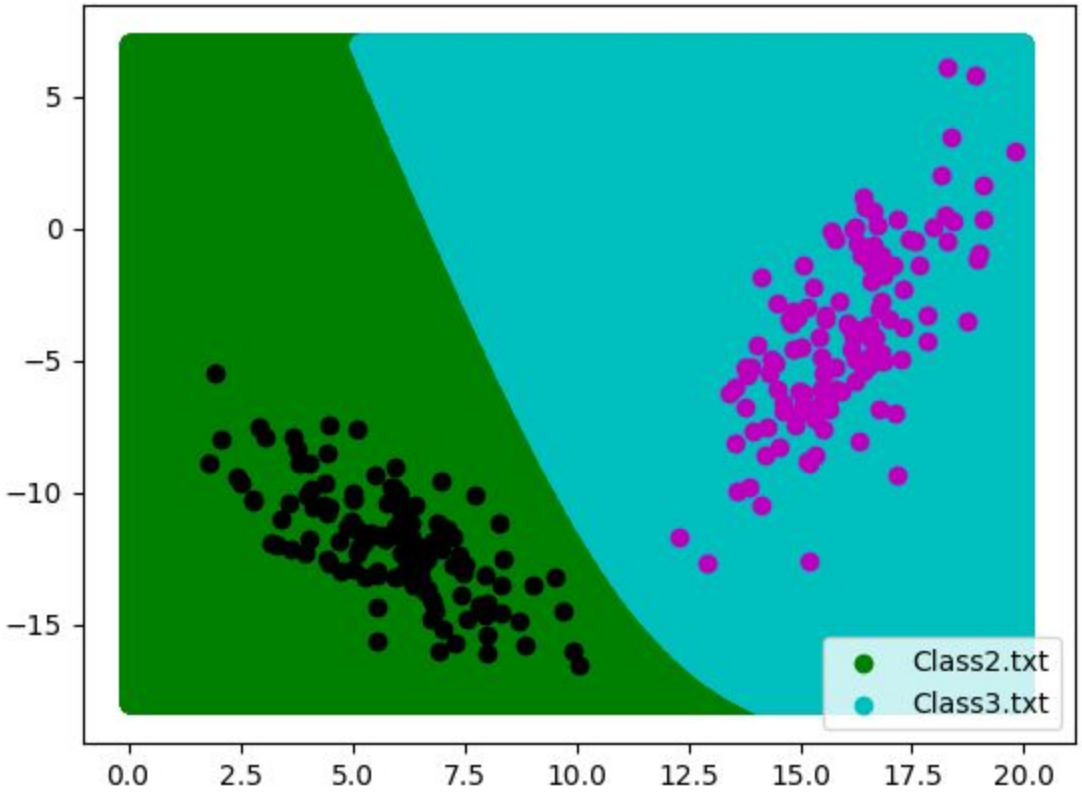


SUMMARISING POINTS :

1) This case corresponds to defining Bayes Classifier and decision boundary between two classes which are linearly separable.
2) Feature vector contains features which are statistically dependent.
3) Nonlinear decision boundary for linearly separable data gave better results in classification.

Confusion Matrix For Class1 Class2	
125.0	0.0
0.0	125.0

TITLE	VALUE
F-measure for Class1	1.0
F-measure for Class2	1.0
Accuracy of Classification	1.0
Mean recall	1.0
Mean precision	1.0
Mean F-measure	1.0
Recall for class Class1	1.0
Recall for class Class2	1.0
Precision for Class1	1.0
Precision for Class2	1.0

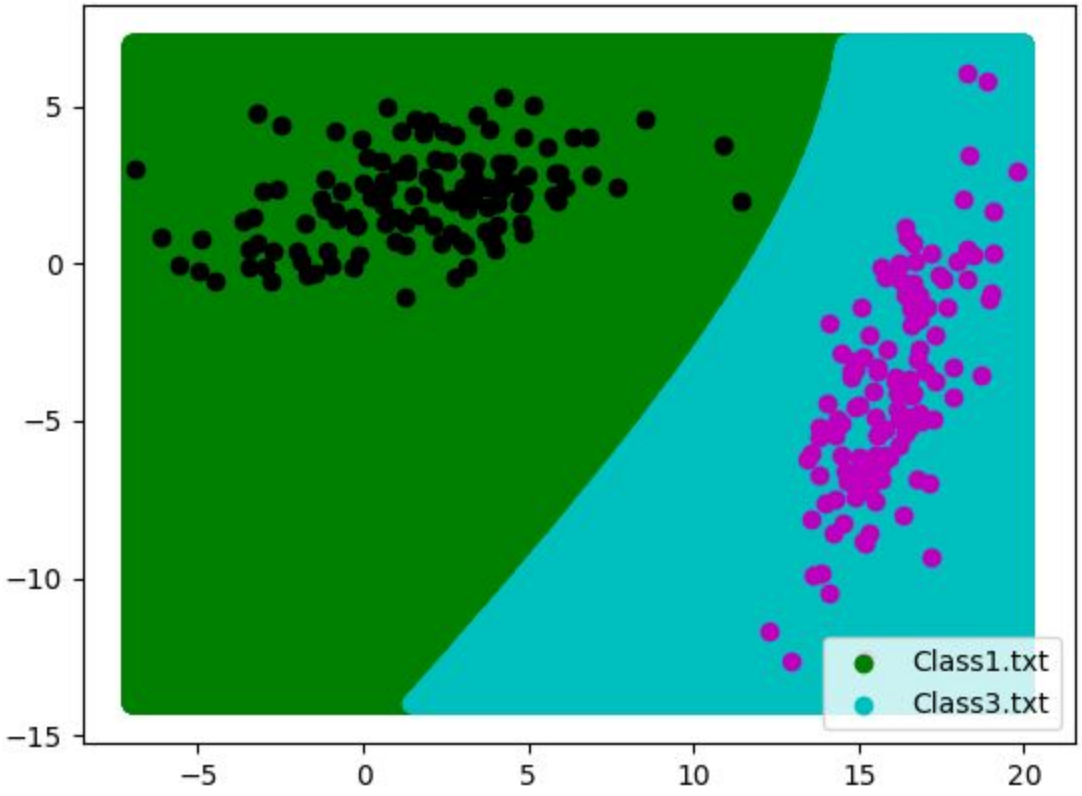


SUMMARISING POINTS :

1) This plot corresponds to data set which contains linearly separable data.
2) The decision boundary is parabolic.
3) As we can see that there is some error associated with the classifier which is shown as two data points which belong to pink class are classified as of black class.
4) Such details about the classifier are exposed in the below tables by various measures.

Confusion Matrix For Class2 Class3	
125.0	0.0
0.0	123.0

TITLE	VALUE
F-measure for Class2	1.0
F-measure for Class3	1.0
Accuracy of Classification	1.0
Mean recall	1.0
Mean precision	1.0
Mean F-measure	1.0
Recall for Class2.	1.0
Recall for Class3	1.0
Precision for Class2	1.0
Precision for Class3	1.0

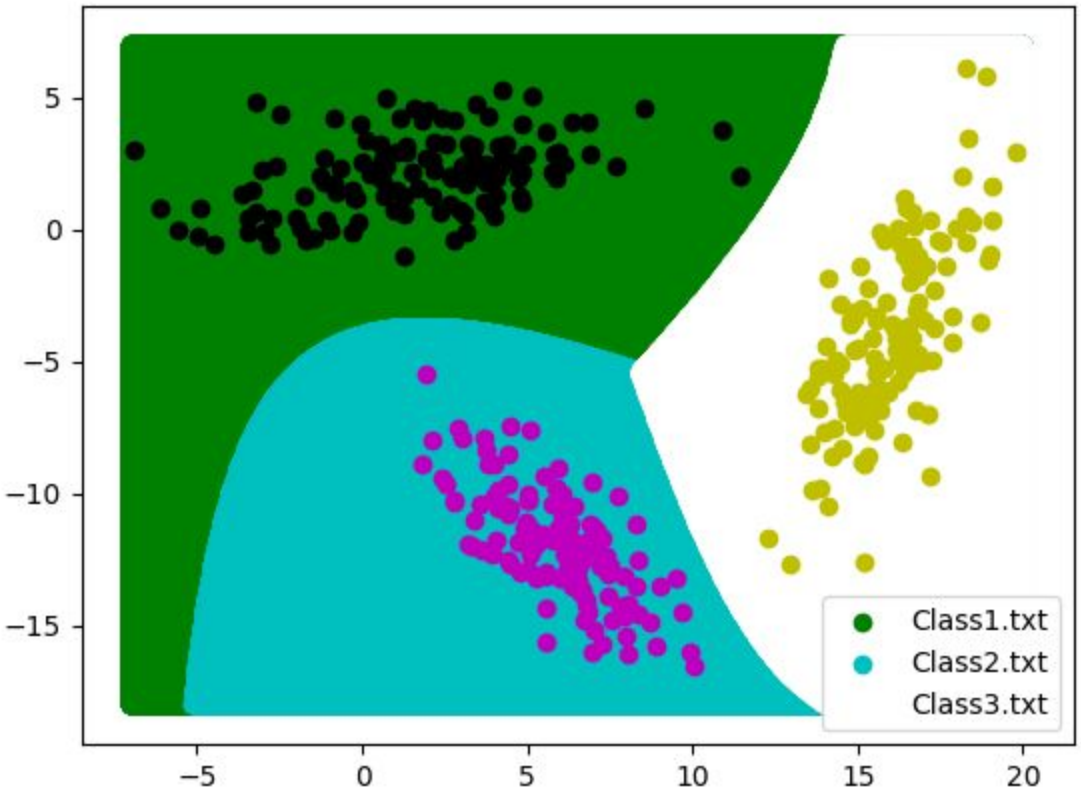


1) Feature vectors have two dimensions.
2) Data is linearly separable but the Bayes classifier applied is nonlinear (as of Case 4).
3) The features of feature vector are not statistically independent.
4) The results have high accuracy in this case.

Confusion Matrix For Class1 Class3

125.0	**0.0**
0.0	**125.0**

TITLE	VALUE
F-measure for Class1	1.0
F-measure for Class3	1.0
Accuracy of Classification	1.0
Mean recall	1.0
Mean precision	1.0
Mean F-measure	1.0
Recall for Class1	1.0
Recall for Class3	1.0
Precision for Class1	1.0
Precision for Class3	1.0

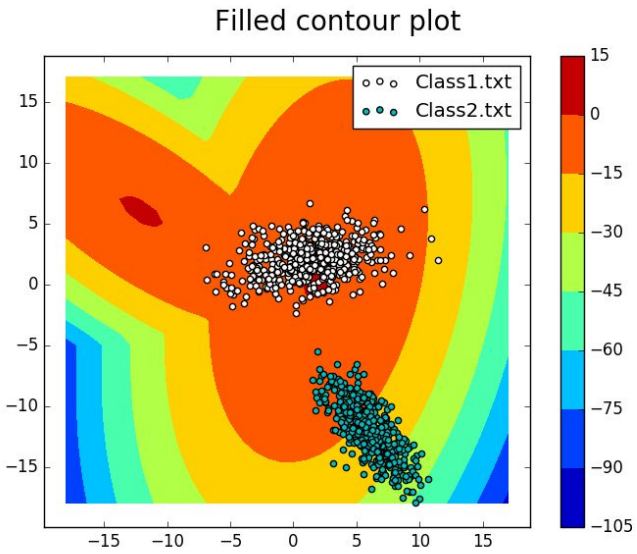


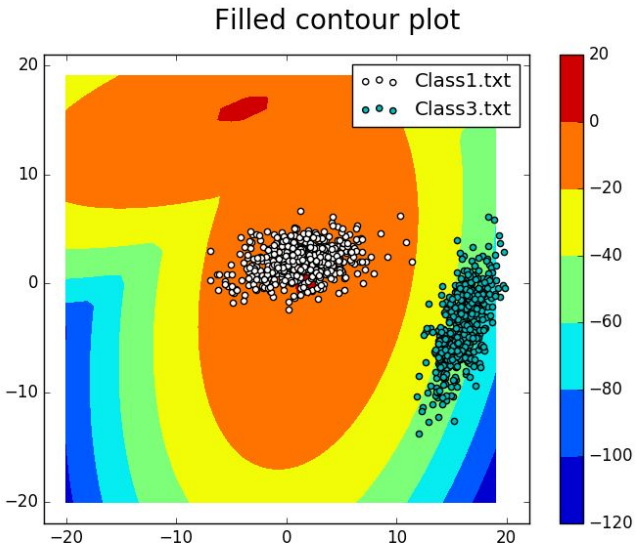
This plot shows formulation of parabolic decision boundary between three classes simultaneously

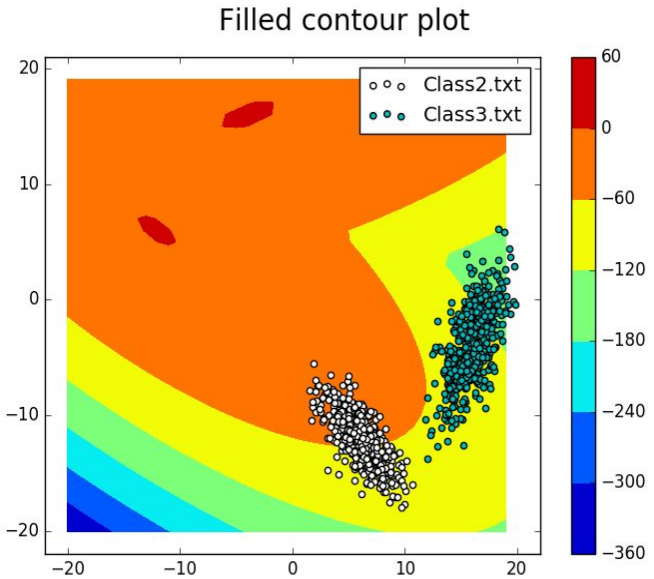
Confusion Matrix For Class1 Class2 Class3		
125.0	0.0	0.0
0.0	125.0	0.0
0.0	0.0	125.0

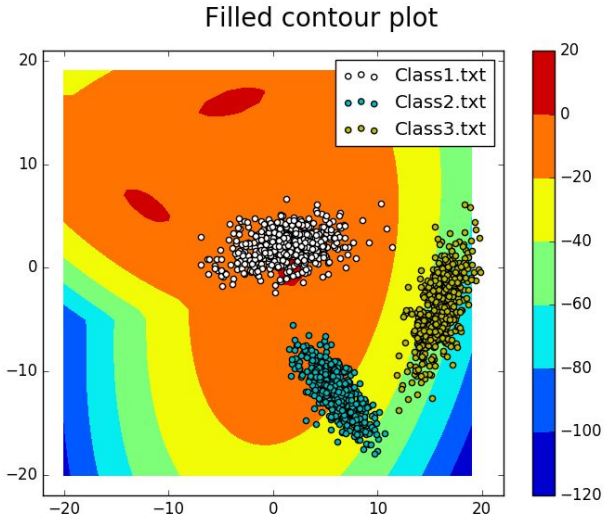
TITLE	VALUE
F-measure for Class1	1.0
F-measure for Class2	1.0
F-measure for Class3	1.0
Accuracy of Classification	1.0
Mean recall	1.0
Mean precision	1.0
Mean F-measure	1.0
Recall for Class1	1.0
Recall for Class2	1.0
Recall for Class3	1.0
Precision for Class1	1.0
Precision for Class2	1.0
Precision for Class3	1.0

Linearly Separable data Case 4 Contour



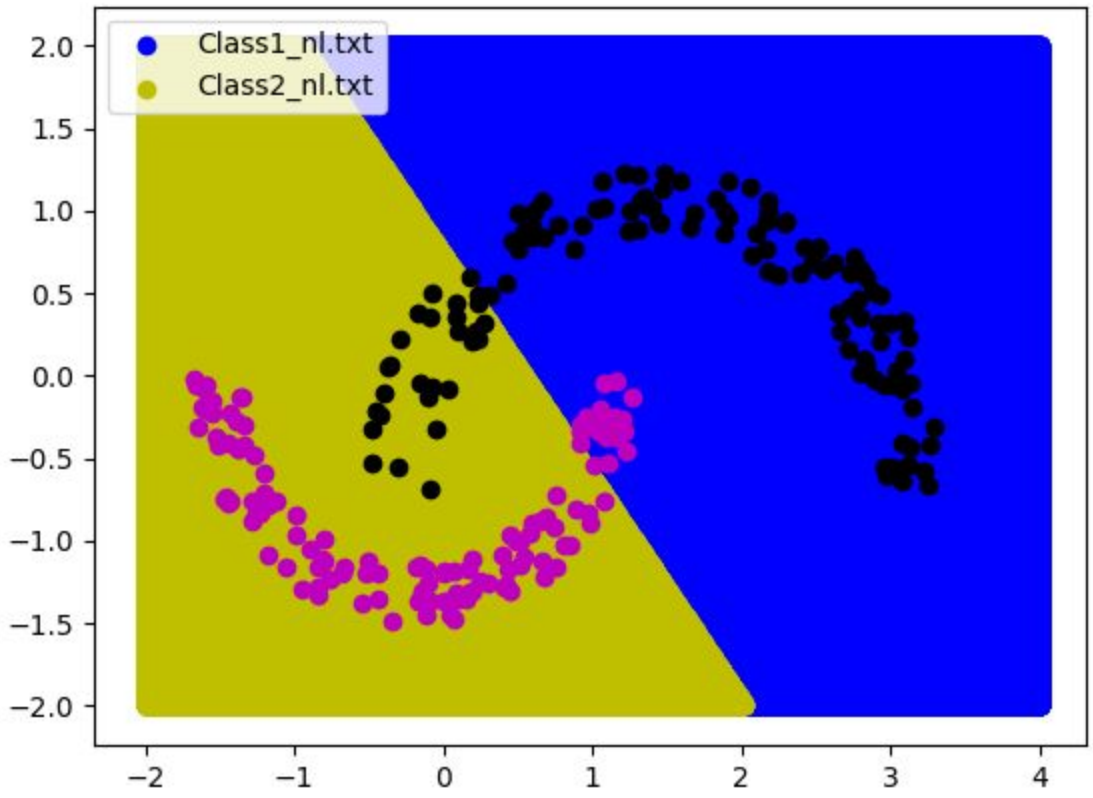






NONLINEAR DATA SET

Non - Linearly Separable Data Case 1

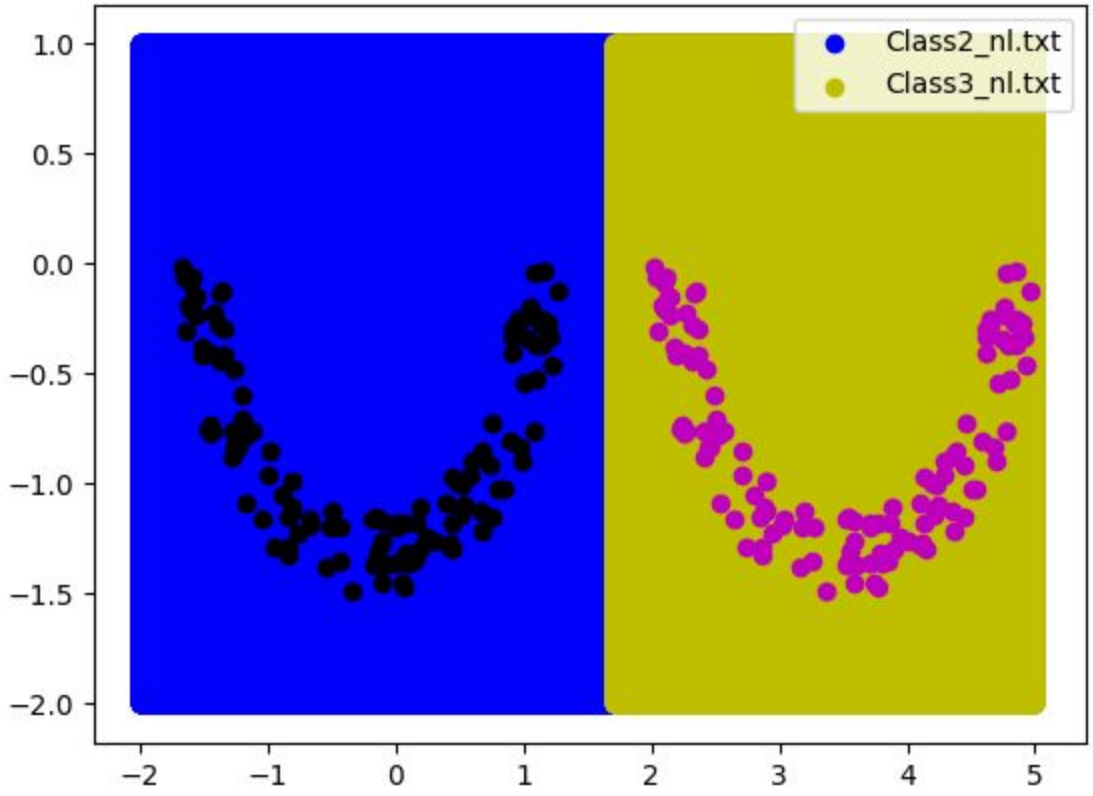


SUMMARISING POINTS :

1) Feature vector is 2 dimensional.
2) The features of feature vector are statistically independent as covariance matrix is diagonal.
3) Dataset given here is not linearly separable and hence defining a linear decision boundary forcefully gives poor results.
4) The results for the above case can be analysed in the various parameters listed below in table.

Confusion Matrix For Class1 Class2	
123.0	2.0
0.0	125.0

TITLE	VALUE
F-measure for Class1	0.992
F-measure for Class2	0.992
Accuracy of Classification	0.992
Mean recall	0.992
Mean precision	0.992
Mean F-measure	0.992
Recall for Class1	0.984
Recall for Class2	1.0
Precision for Class1	1.0
Precision for Class2	0.984

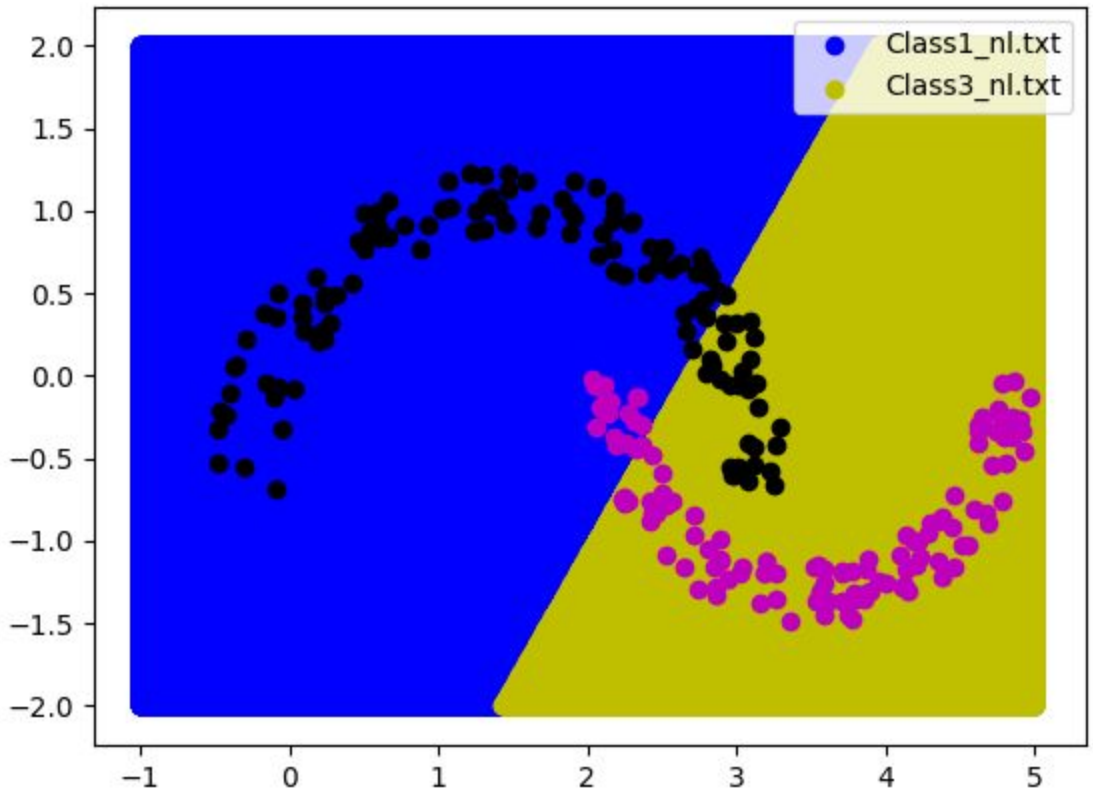


SUMMARISING POINTS :

1) The provided dataset 2 is nonlinearly separable but in this case "by chance" a linear decision boundary is able to classify the data into two separate classes.
2) This shows accuracy of classifier highly depends on distribution of data.
3) These type of cases are rare.

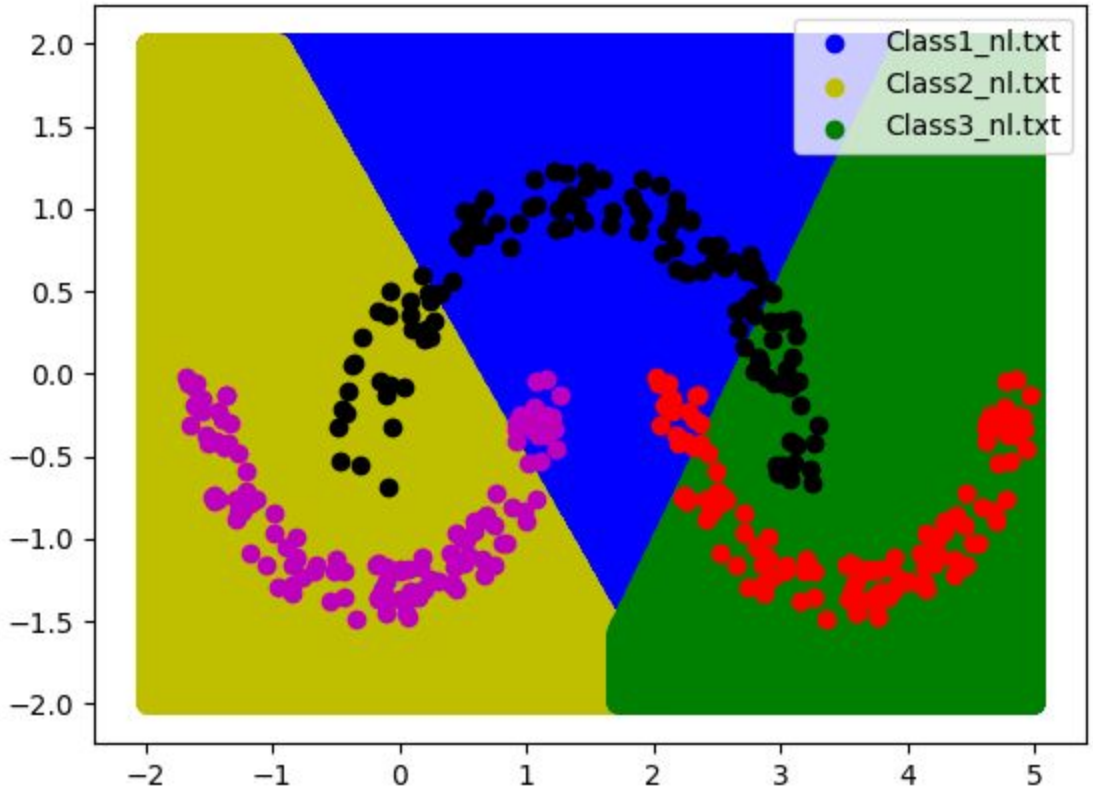
Confusion Matrix For Class2 Class3	
123.0	**2.0**
0.0	**125.0**

TITLE	VALUE
F-measure for Class2	0.992
F-measure for Class3	0.992
Accuracy of Classification	0.992
Mean recall	0.992
Mean precision	0.992
Mean F-measure	0.992
Recall for Class2	0.984
Recall for Class3	1.0
Precision for Class2	1.0
Precision for Class3	0.984



Confusion Matrix For Class1 Class3	
123.0	2.0
0.0	125.0

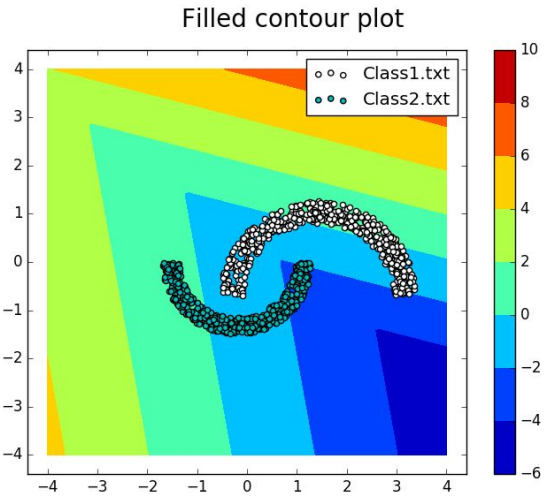
TITLE	VALUE
F-measure for Class1	0.992
F-measure for Class3	0.992
Accuracy of Classification	0.992
Mean recall	0.992
Mean precision	0.992
Mean F-measure	0.992
Recall for Class1	0.984
Recall for Class3	1.0
Precision for Class1	1.0
Precision for Class3	0.984

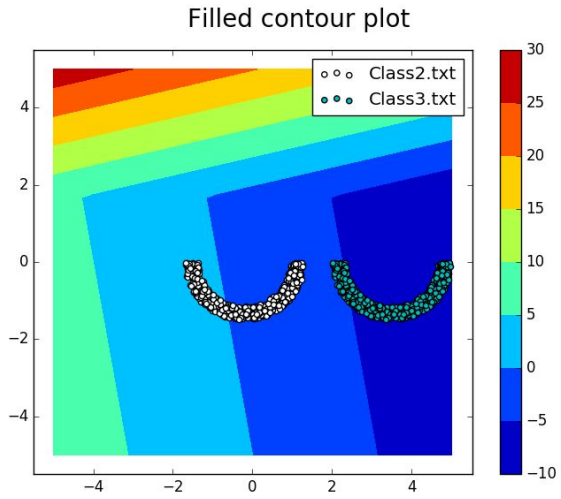


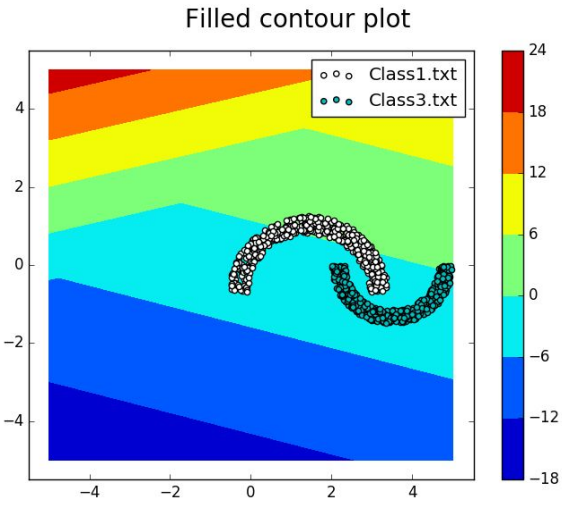
Confusion Matrix For Class1 Class2 Class3		
123.0	0.0	2.0
1.0	124.0	0.0
0.0	2.0	123.0

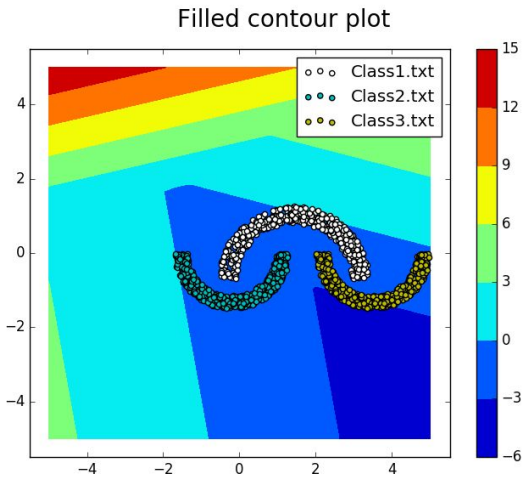
TITLE	VALUE
F-measure for Class1	1.0
F-measure for Class2	1.0
F-measure for Class3	1.0
Accuracy of Classification	1.0
Mean recall	1.0
Mean precision	1.0
Mean F-measure	1.0
Recall for Class1	1.0
Recall for Class2	1.0
Recall for Class3	1.0
Precision for Class1	1.0
Precision for Class2	1.0
Precision for Class3	1.0

Non - Linearly Separable Data Case 1 Contour

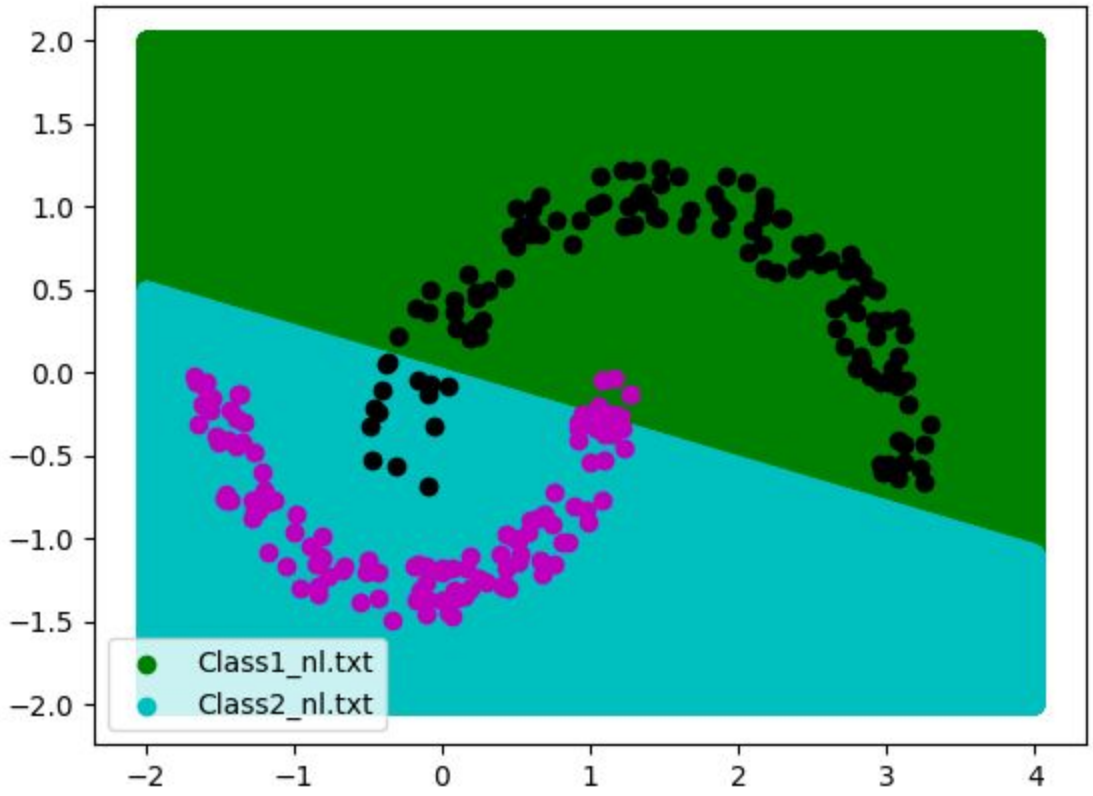








Non - Linearly Separable Data Case 2

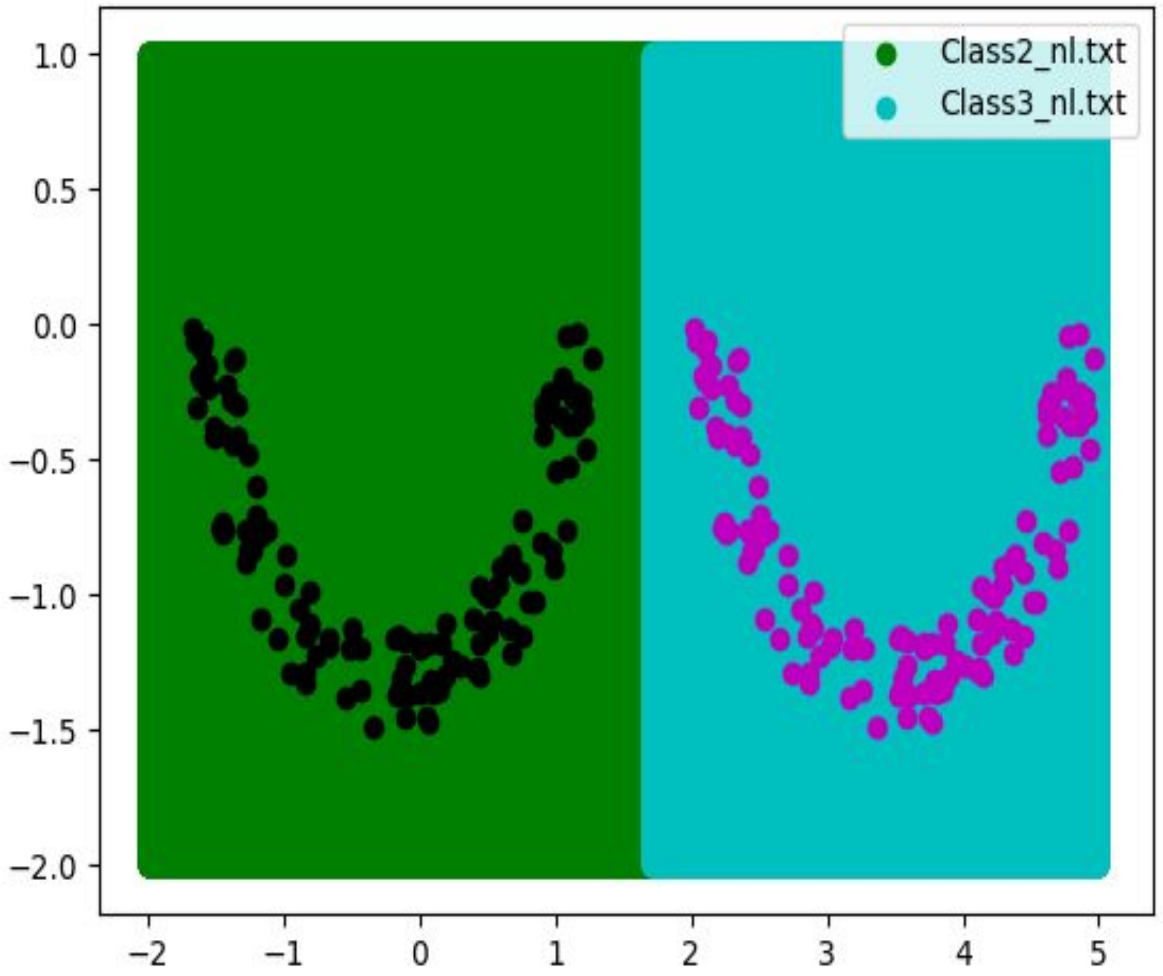


SUMMARISING POINTS :

1) Covariance matrix is full(non diagonal) and the features present in each feature vector are not statistically independent.
2) The resulting decision boundary with constraints of case 1 is linear in nature but the accuracy with such a decision boundary is not that great.

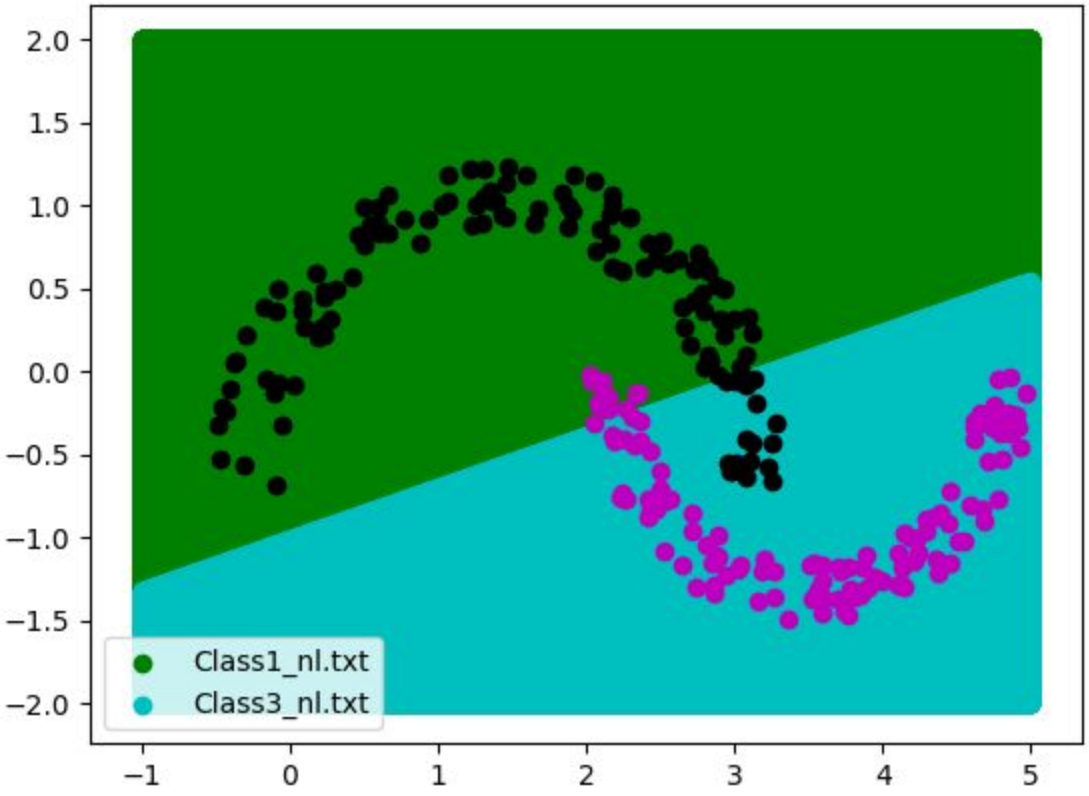
Confusion Matrix For Class1 Class2	
123.0	2.0
0.0	125.0

TITLE	VALUE
F-measure for Class1	0.992
F-measure for Class2	0.992
Accuracy of Classification	0.992
Mean recall	0.992
Mean precision	0.992
Mean F-measure	0.992
Recall for Class1	0.984
Recall for Class2	1.0
Precision for Class1	1.0
Precision for Class2	0.984



Confusion Matrix For Class2 Class3	
123.0	2.0
0.0	125.0

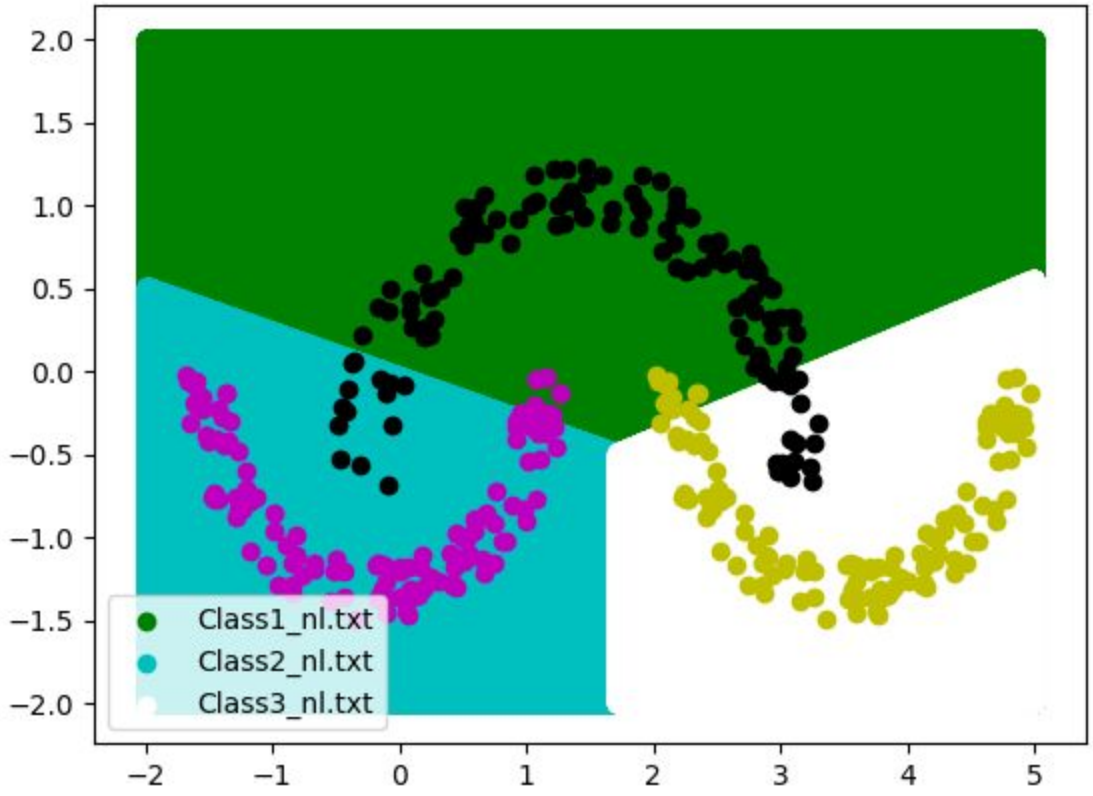
TITLE	VALUE
F-measure for Class2	0.992
F-measure for Class3	0.992
Accuracy of Classification	0.992
Mean recall	0.992
Mean precision	0.992
Mean F-measure	0.992
Recall for Class2	0.984
Recall for Class3	1.0
Precision for Class2	1.0
Precision for Class3	0.984



1) This figure also shows a linear classifier for nonlinearly distributed dataset and as a result the results are not so great and they can get worse with increasing polarization of data points in a nonlinear fashion.

Confusion Matrix For Class1 Class3	
123.0	2.0
0.0	125.0

TITLE	VALUE
F-measure for Class1	0.992
F-measure for Class3	0.992
Accuracy of Classification	0.992
Mean recall	0.992
Mean precision	0.992
Mean F-measure	0.992
Recall for Class1	0.984
Recall for Class3	1.0
Precision for Class1	1.0
Precision for Class3	0.984

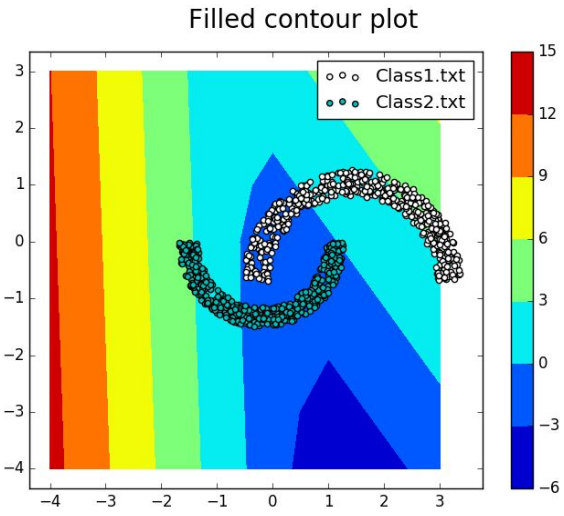


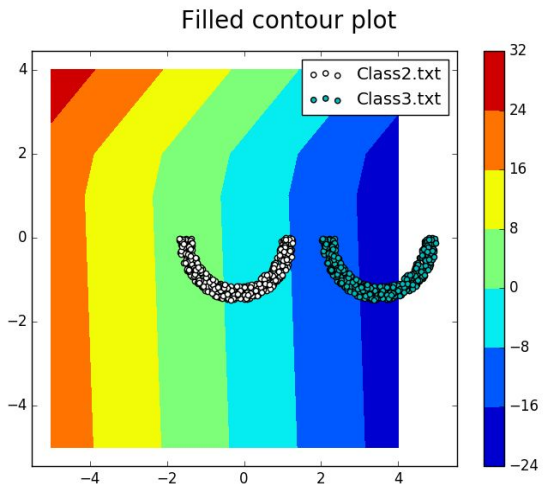
Simultaneous decision boundaries of all 3 classes for Case 2

Confusion Matrix For Class1 Class2 Class3		
123.0	0.0	2.0
1.0	124.0	0.0
0.0	2.0	123.0

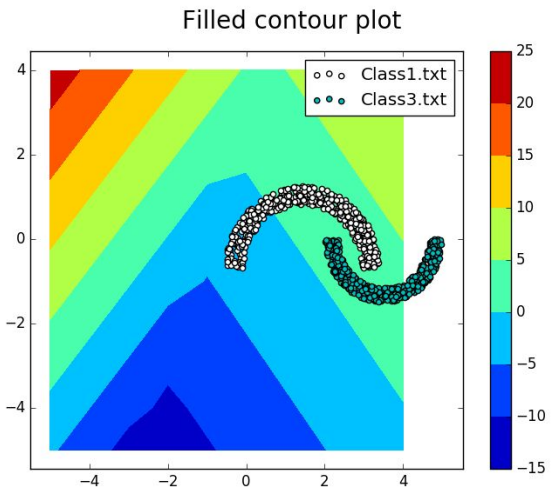
TITLE	VALUE
F-measure for Class1	1.0
F-measure for Class2	1.0
F-measure for Class3	1.0
Accuracy of Classification	1.0
Mean recall	1.0
Mean precision	1.0
Mean F-measure	1.0
Recall for Class1	1.0
Recall for Class2	1.0
Recall for Class3	1.0
Precision for Class1	1.0
Precision for Class2	1.0
Precision for Class3	1.0

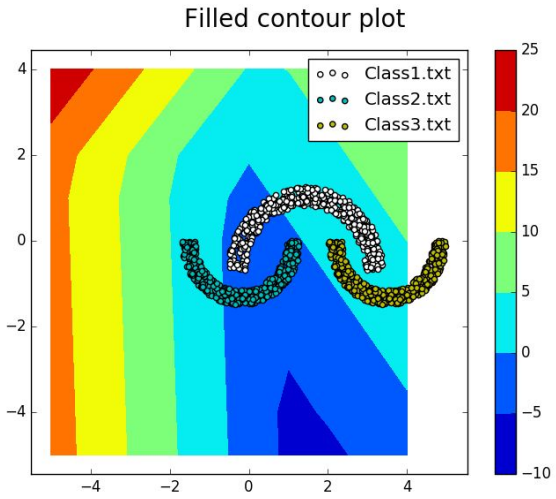
Non - Linearly Separable Data Case 2 Contour



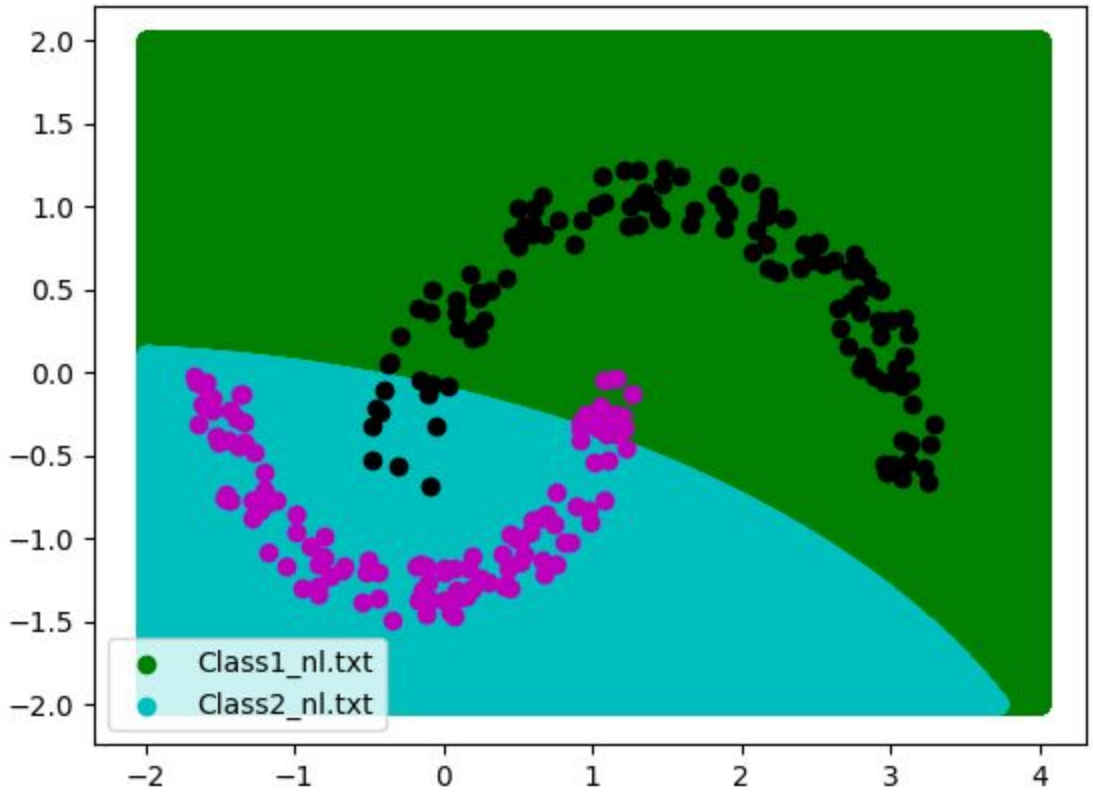


Contour Plots of various combination of classes





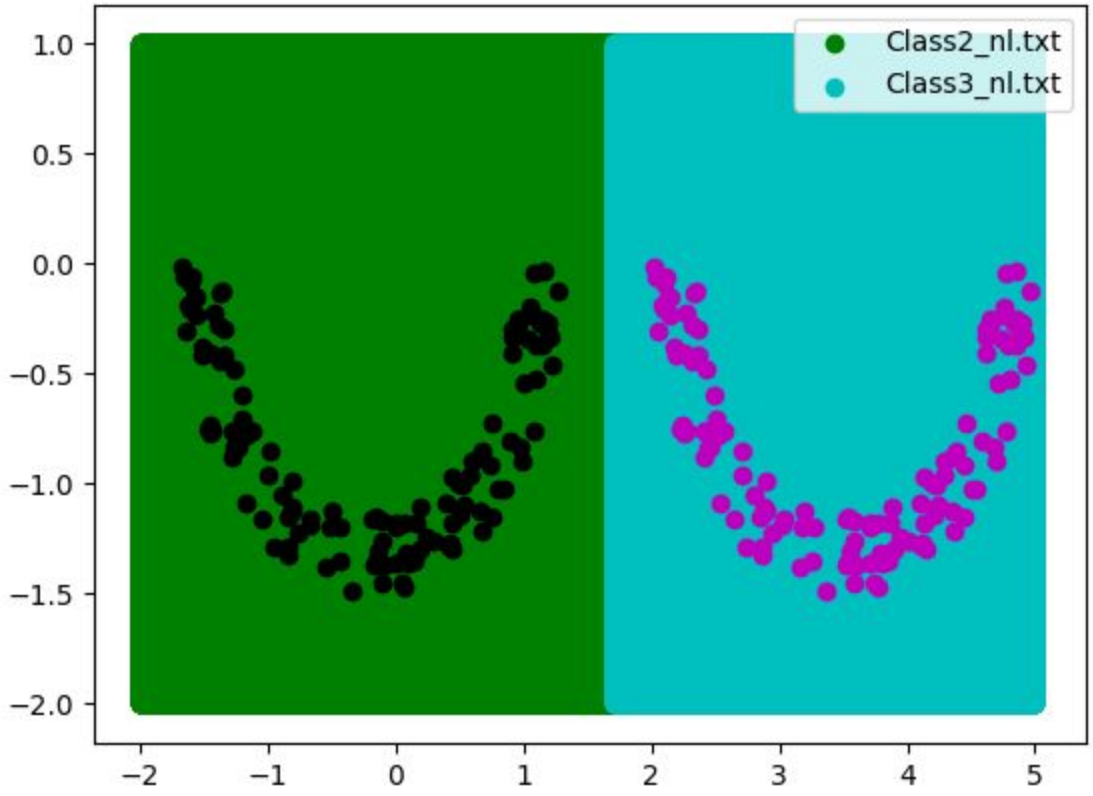
Non - Linearly Separable Data Case 3



1) As the covariance matrix is diagonal hence features considered in feature vector are statistically independent.
2) Decision boundary is parabolic.

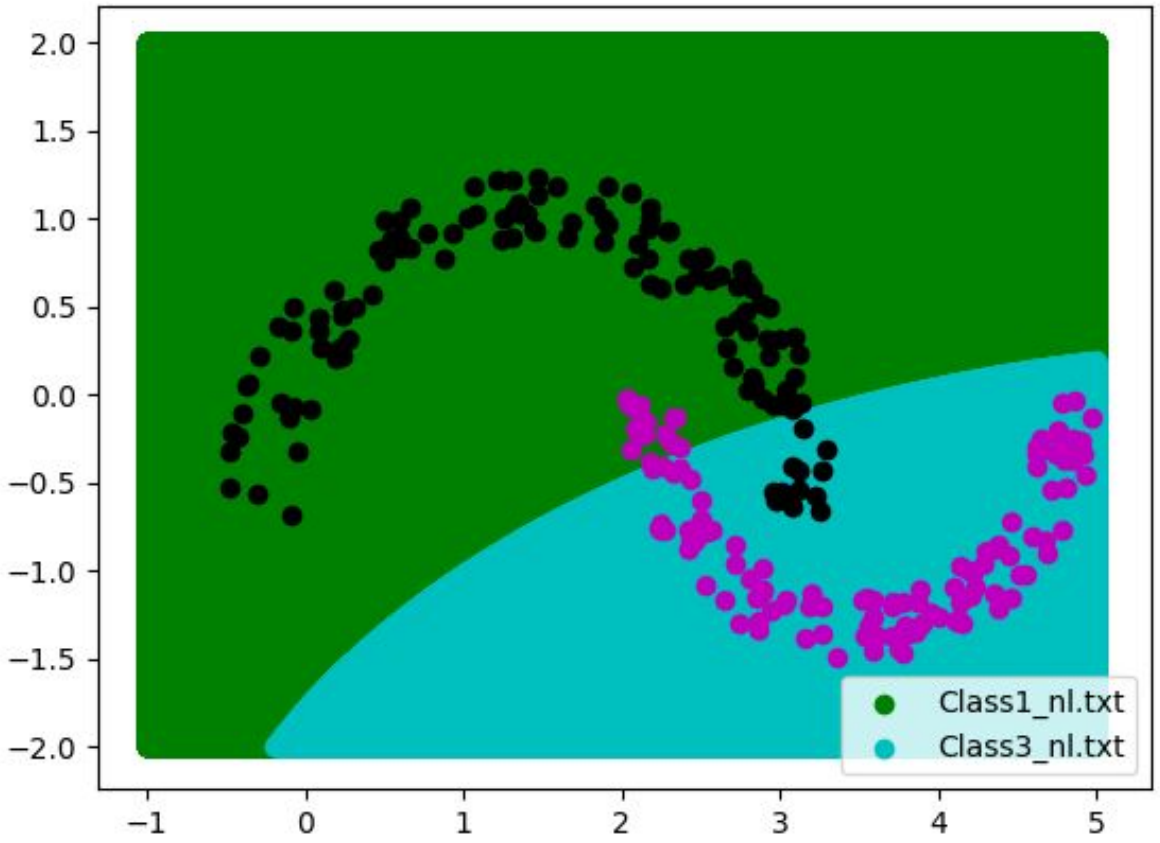
Confusion Matrix For Class1 Class2	
123.0	**2.0**
0.0	**125.0**

TITLE	VALUE
F-measure for Class1	0.992
F-measure for Class2	0.992
Accuracy of Classification	0.992
Mean recall	0.992
Mean precision	0.992
Mean F-measure	0.992
Recall for Class1	0.984
Recall for Class2	1.0
Precision for Class1	1.0
Precision for Class2	0.984



Confusion Matrix For Class2 Class3	
123.0	2.0
0.0	125.0

TITLE	VALUE
F-measure for Class2	0.992
F-measure for Class3	0.992
Accuracy of Classification	0.992
Mean recall	0.992
Mean precision	0.992
Mean F-measure	0.992
Recall for Class2	0.984
Recall for Class3	1.0
Precision for Class2	1.0
Precision for Class3	0.984



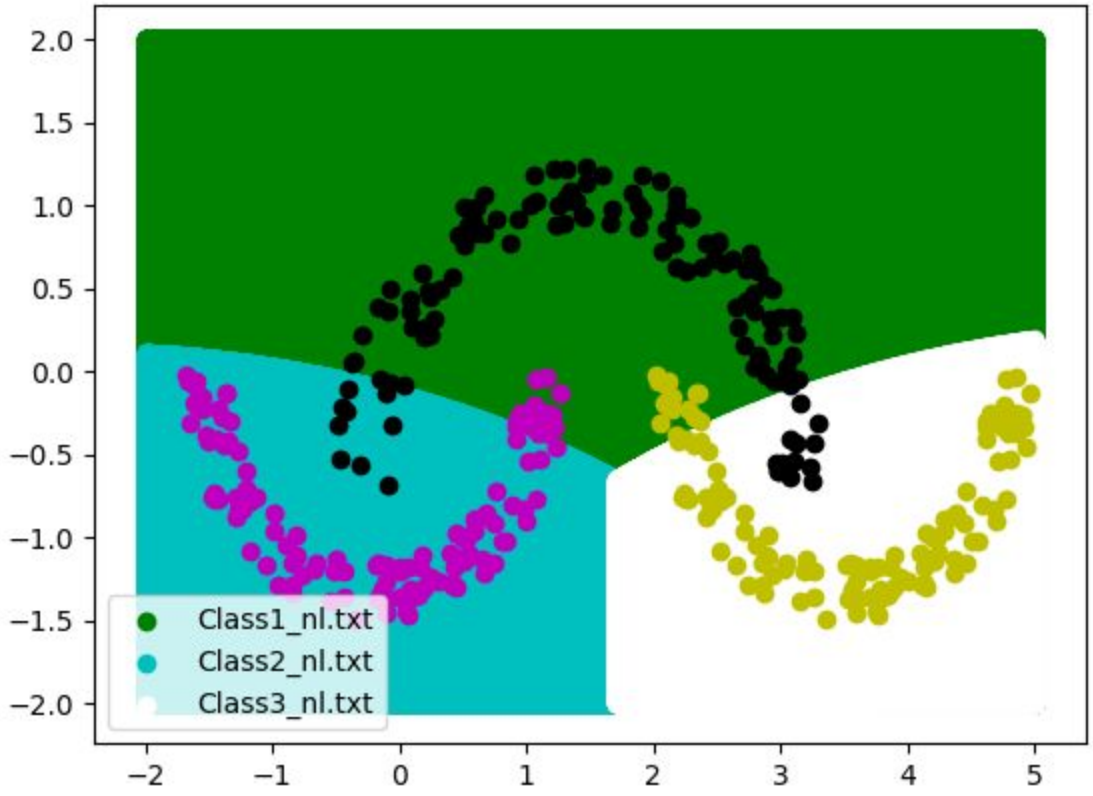
Above plot demonstrates a nonlinear (parabolic to be specific) decision boundary (Case 3) for non linearly separable dataset.

SUMMARISING POINTS :

1) For this case as it can be clearly seen that even a parabolic decision boundary does not completely classify test data correctly into both the classes and the results can turn even worse if the data distribution is more close and polarized.

Confusion Matrix For Class1 Class3	
123.0	2.0
0.0	125.0

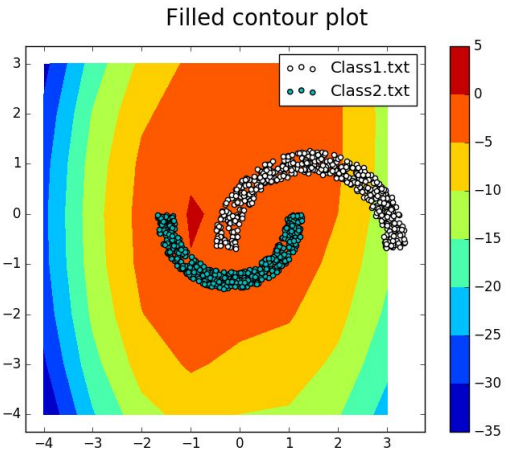
TITLE	VALUE
F-measure for Class1	0.992
F-measure for Class3	0.992
Accuracy of Classification	0.992
Mean recall	0.992
Mean precision	0.992
Mean F-measure	0.992
Recall for Class1	0.984
Recall for Class3	1.0
Precision for Class1	1.0
Precision for Class3	0.984

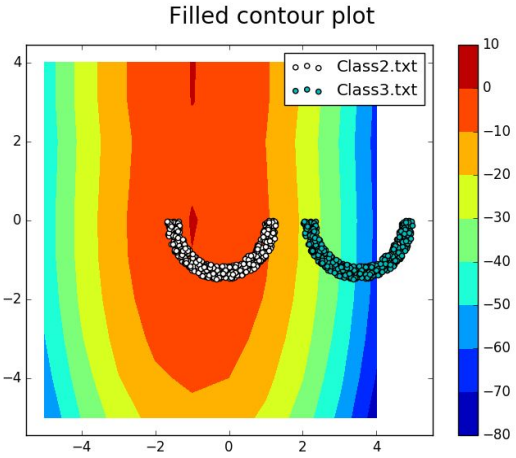


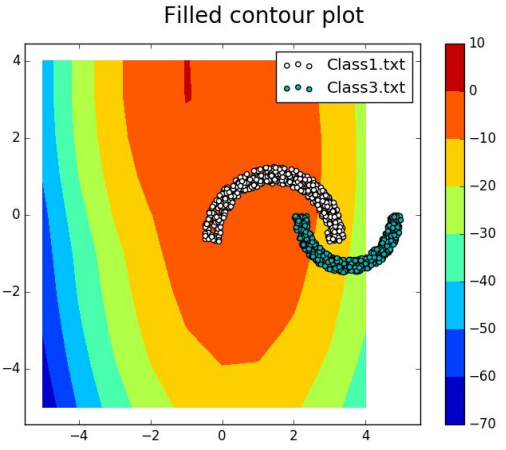
Confusion Matrix For Class1 Class2 Class3		
123.0	0.0	2.0
1.0	124.0	0.0
0.0	2.0	123.0

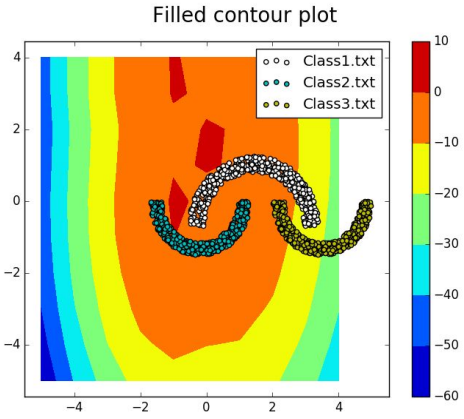
TITLE	VALUE
F-measure for Class1	1.0
F-measure for Class2	1.0
F-measure for Class3	1.0
Accuracy of Classification	1.0
Mean recall	1.0
Mean precision	1.0
Mean F-measure	1.0
Recall for Class1	1.0
Recall for Class2	1.0
Recall for Class3	1.0
Precision for Class1	1.0
Precision for Class2	1.0
Precision for Class3	1.0

Non - Linearly Separable Data Case 3 Contour

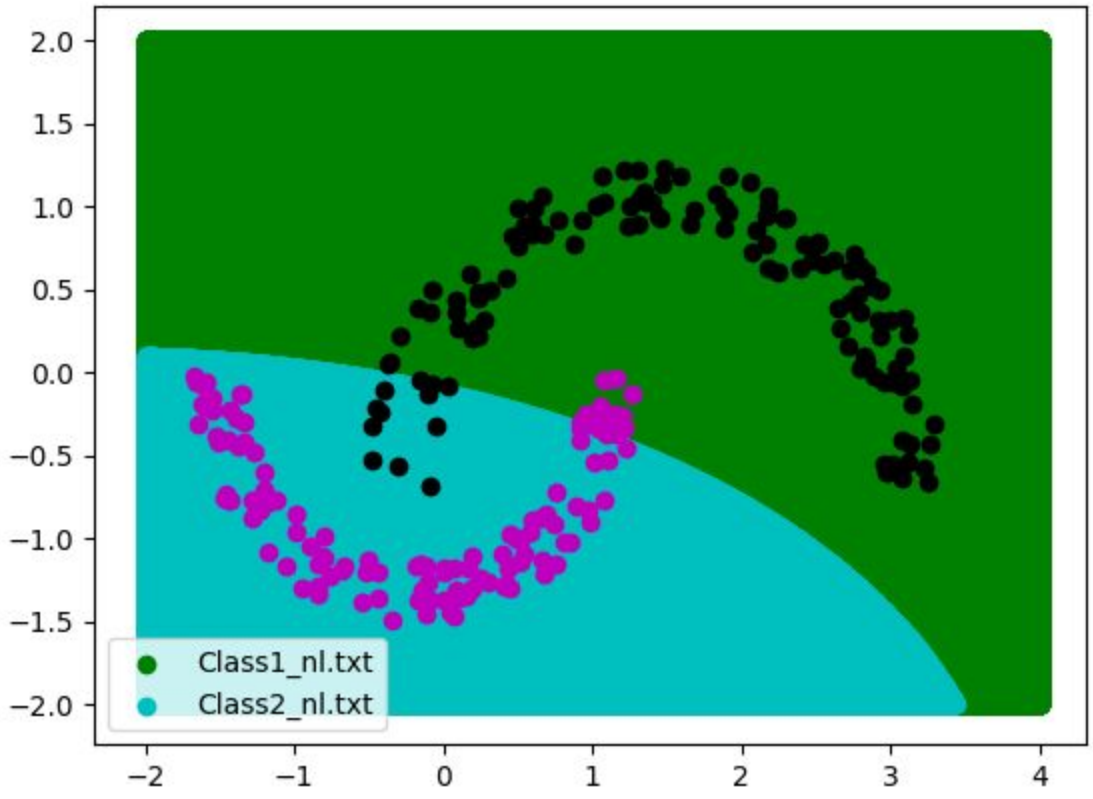








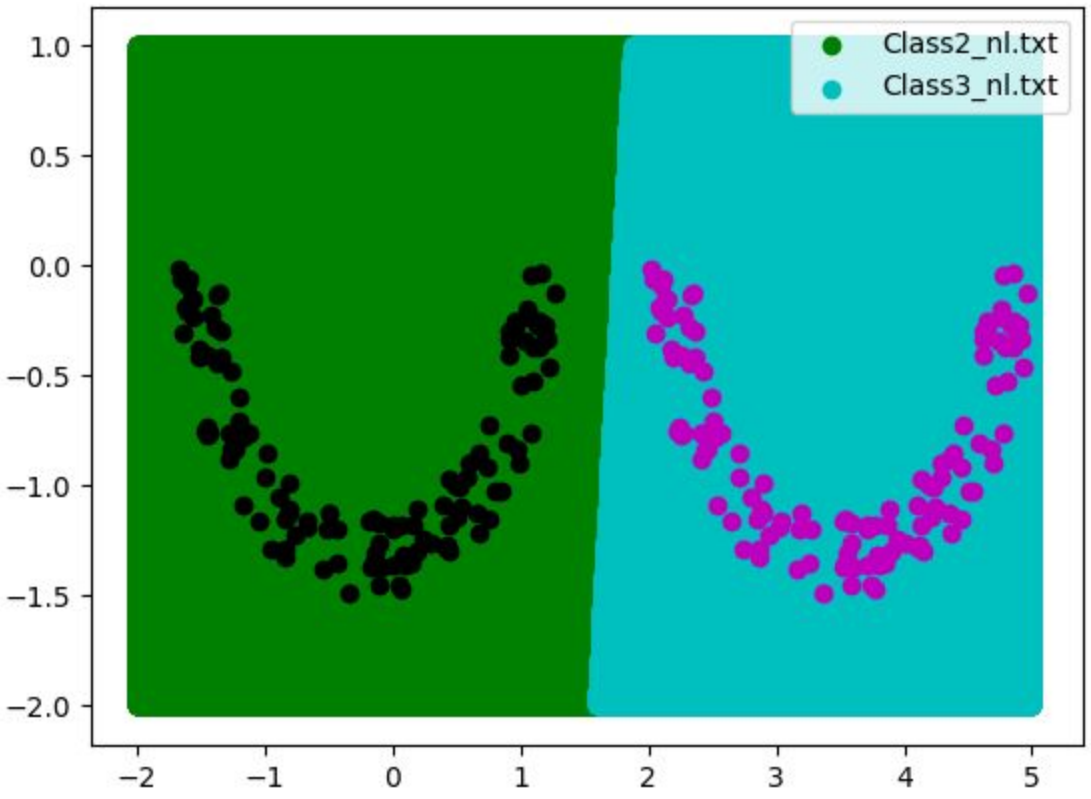
Non Linearly Separable Data Case 4



1) Covariance matrix for each class is different and arbitrary;
2) The feature vector components are not statistically independent.
3) As we know that under the constraints of Case 4 the decision boundary generated in nonlinear and as our dataset is also nonlinearly separable hence this is the most optimized version for this data set.
4) Still there are some wrong assumptions which can be seen in the above plot as well as they are formulated in form of various parameters in below tables.

Confusion Matrix For Class1 Class2	
123.0	2.0
0.0	125.0

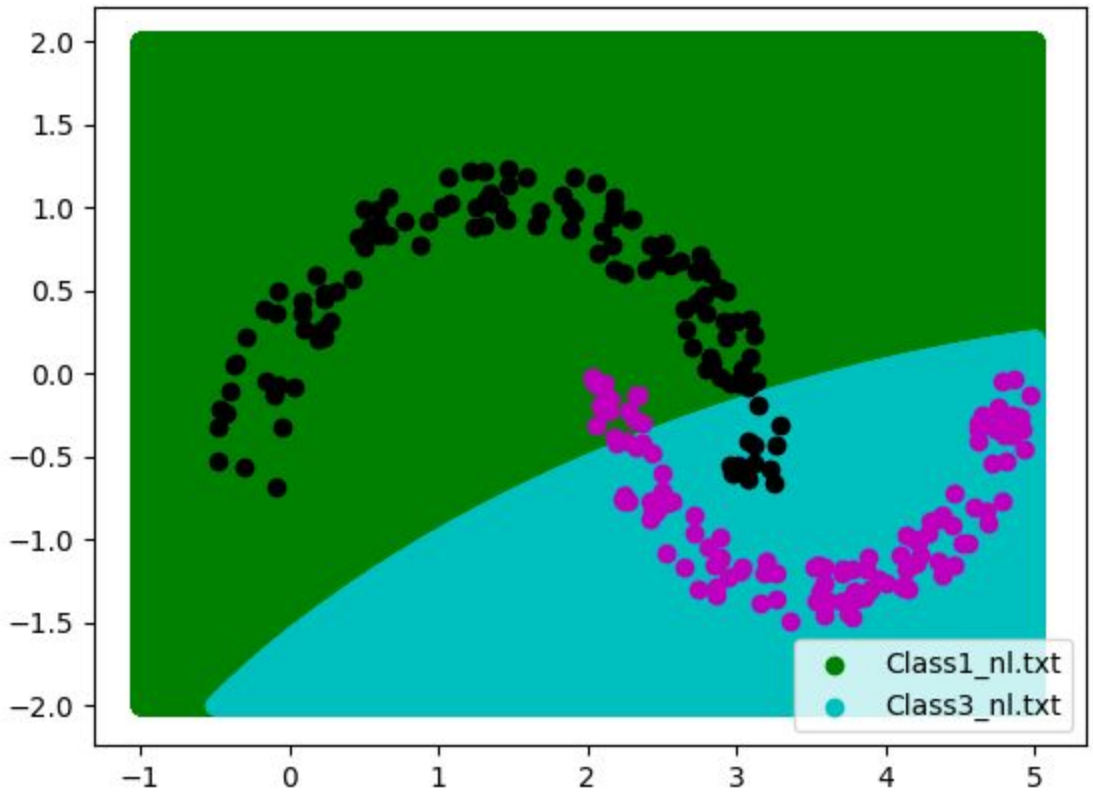
TITLE	VALUE
F-measure for Class1	0.992
F-measure for Class2	0.992
Accuracy of Classification	0.992
Mean recall	0.992
Mean precision	0.992
Mean F-measure	0.992
Recall for Class1	0.984
Recall for Class2	1.0
Precision for Class1	1.0
Precision for Class2	0.984



Plot between Class 2 and Class 3

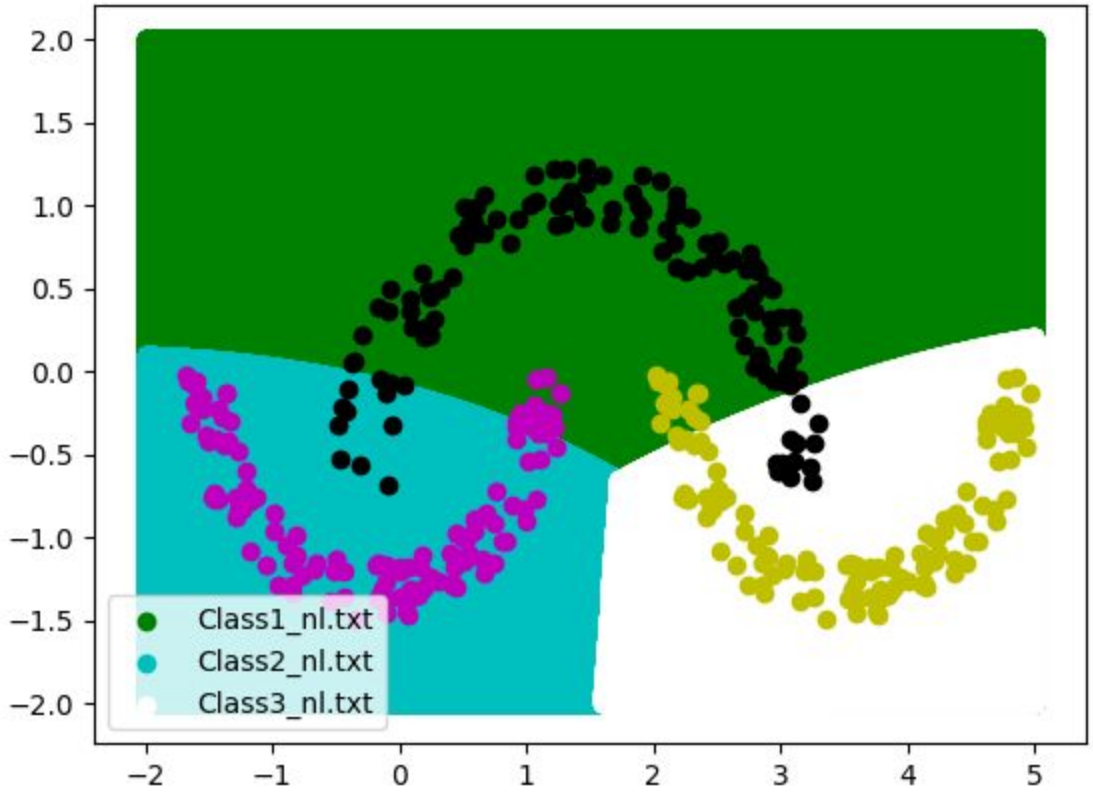
Confusion Matrix For Class2 Class3	
123.0	2.0
0.0	125.0

TITLE	VALUE
F-measure for Class2	0.992
F-measure for Class3	0.992
Accuracy of Classification	0.992
Mean recall	0.992
Mean precision	0.992
Mean F-measure	0.992
Recall for Class2	0.984
Recall for Class3	1.0
Precision for Class2	1.0
Precision for Class3	0.984



Confusion Matrix For Class1 Class3	
123.0	**2.0**
0.0	**125.0**

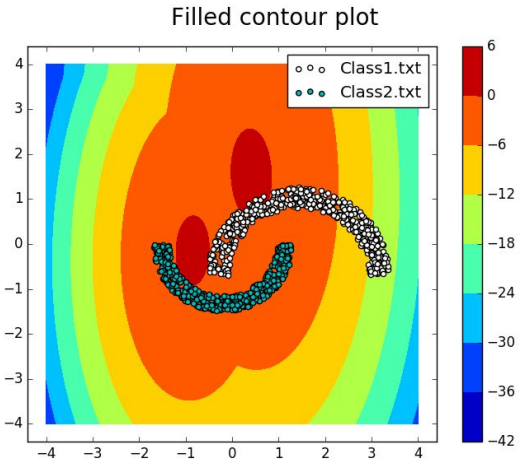
TITLE	VALUE
F-measure for Class1	0.992
F-measure for Class3	0.992
Accuracy of Classification	0.992
Mean recall	0.992
Mean precision	0.992
Mean F-measure	0.992
Recall for Class1	0.984
Recall for Class3	1.0
Precision for Class1	1.0
Precision for Class3	0.984

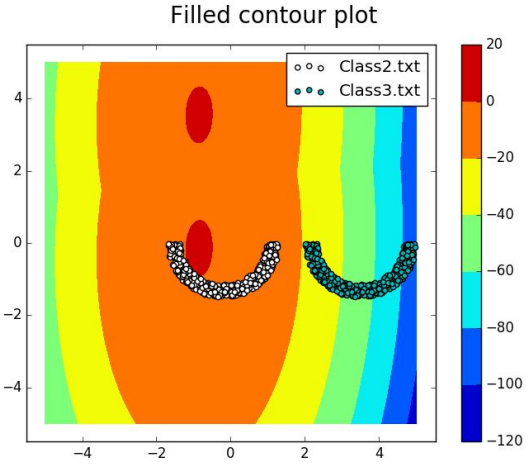


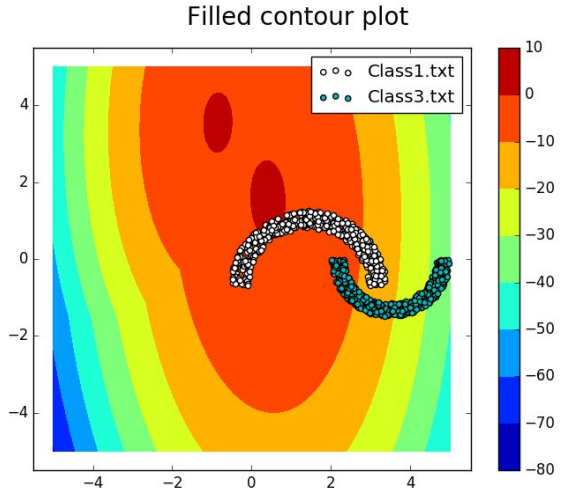
Confusion Matrix For Class1 Class2 Class3		
123.0	0.0	2.0
1.0	124.0	0.0
0.0	2.0	123.0

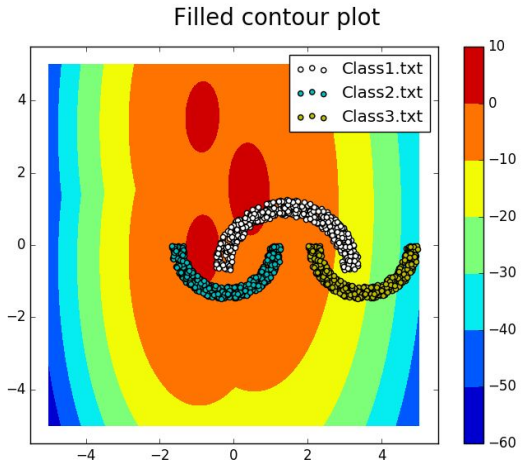
TITLE	VALUE
F-measure for Class1	1.0
F-measure for Class2	1.0
F-measure for Class3	1.0
Accuracy of Classification	1.0
Mean recall	1.0
Mean precision	1.0
Mean F-measure	1.0
Recall for Class1	1.0
Recall for Class2	1.0
Recall for Class3	1.0
Precision for Class1	1.0
Precision for Class2	1.0
Precision for Class3	1.0

Non - Linearly Separable Data Case 4 : Contours









References :

1. Pattern Classification

   Book by David G. Stork, Peter E. Hart, and Richard O. Duda

2. MatPlotLib: https://matplotlib.org/
3. Numpy : https://www.scipy.org/scipylib/download.html
4. Python 3.X https://docs.python.org/3.5/