CS669: Pattern Recognition

Programming Assignment 1

Instructor: Dileep A.D.

By-

Group 12

Ayush Garg (B13111)

Himanshu Singal (B13312)

Hitesh Tarani (B13139)

Introduction

Pattern recognition is the research area that studies the operation and design of systems that recognize patterns in data. It encloses sub disciplines like discriminant analysis, feature extraction, error estimation and cluster analysis. In this assignment we were given many artificial as well as real world datasets and assuming each dataset to follow Gaussian distribution we built Bayes as well as Naïve-Bayes classifier and thus calculated decision boundary for each of the datasets.

We changed the nature of the covariance matrices of the classes and observed the changes in the decision boundaries as a result of that. Then finally we tested the classification accuracy of each classifier on the test data and made a confusion matrix for each of the cases as a result. We also plotted the decision region for every pair of the classes as well as the decision region plot for all the classes together with the training data superposed.

Bayesian Decision Theory

Bayesian decision theory is a fundamental statistical approach to the problem of pattern classification. It is considered the ideal case in which the probability structure underlying the categories is known perfectly. While this sort of situation rarely occurs in practice, it permits us to determine the optimal (Bayes) classifier against which we can compare all other classifiers. Moreover, in some problems it enables us to predict the error we will get when we generalize to novel patterns. This approach is based on quantifying the tradeoffs between various classification decisions using probability and the costs that accompany such decisions. It makes the assumption that the decision problem is posed in probabilistic terms, and that all of the relevant probability values are known.

The Bayes formula is based on prior probability and the class conditional density function the classes (likelihood). Bayes formula:

$$P(w_j|x) = \frac{p(x|w_j)P(w_j)}{p(x)}$$

Here, $P(w_j|x)$ denotes the probability of class w_j given x. Also, $p(x|w_j)$ denotes the class conditional density function also called as the likelihood and $P(w_j)$ is the prior probability of the class.

Based on this formula we calculated the discriminant function of different classes and thus the decision surface was obtained.

Discriminant function

As the evidence i.e. $p(\bar{x})$ is same for all the classes hence it can be neglected in the calculation for the discriminant function. Also, the value of the posterior i.e. the likelihood function multiplied by prior probability of the class can be very small sometimes i.e. why to increase its range we multiply it by an increasing function. In our case we took ln on the function because it also simplifies the equation to a great extent especially the Gaussian distribution.

$$g_i(\bar{x}) = \ln(p(\bar{x}|y_i)P(y_i))$$

$$g_i(\bar{x}) = \ln p(\bar{x}|y_i) + \ln P(y_i)$$

Expression for the decision surface

According to the min error rate classifier we assign $\bar{x}$ to the class y_i when,

$$g_i(\bar{x}) > g_j(\bar{x}) \quad \forall\ j \neq i$$

Therefore, the equation for the decision surface is:

$$g(\bar{x}) = g_i(\bar{x}) - g_j(\bar{x}) = 0 \quad \forall\ j \neq i$$

$$g(\bar{x}) = \ln p(\bar{x}|y_i) + \ln P(y_i) - \ln p\left(\bar{x}\middle|y_j\right) - \ln P\left(y_j\right)$$

$$g(\bar{x}) = \ln \frac{p(\bar{x}|y_i)}{p(\bar{x}|y_j)} + \ln \frac{P(y_i)}{P(y_j)}$$

Now, as we are assuming the distribution to be Gaussian therefore the likelihood is given as follows:

$$p(\bar{x}|y_i) = \frac{1}{(2\pi)^{\frac{d}{2}}\ |\Sigma|^{\frac{1}{2}}} e^{-(\bar{x}-\bar{\mu}_i)^T \Sigma_i^{-1} (\bar{x}-\bar{\mu}_i)}$$

Hence, the discriminant function $g_i(\bar{x})$ is as follows:

$$g_i(\bar{x}) = -\frac{1}{2}(\bar{x} - \bar{\mu}_i)^T \Sigma_i^{-1} (\bar{x} - \bar{\mu}_i) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_i| + \ln \mathrm{P}(y_i)$$

Naïve Bayes Classifier

We use Naïve Bayes Classifier for i.i.d. (identically independent data). Here, we assume the features of the classes to be independent from each other and hence the covariance matrix is a diagonal matrix while in real world this is rarely possible so we force the covariance matrix to be a diagonal one.

For some types of probability models, Naïve Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for Naïve Bayes models uses the method of maximum likelihood; in other words, one can work with the Naïve Bayes model without accepting Bayesian probability or using any Bayesian methods. Despite their Naïve design and apparently oversimplified assumptions, Naïve Bayes classifiers have worked quite well in many complex real-world situations.

Result of Studies

Data-Set I (Artificial Data)

1) Linearly Separable

1.1) Bayes Classifier

1.1.a) Covariance Matrix Same

1.1.a.i) Average of Covariance Matrices of All Classes

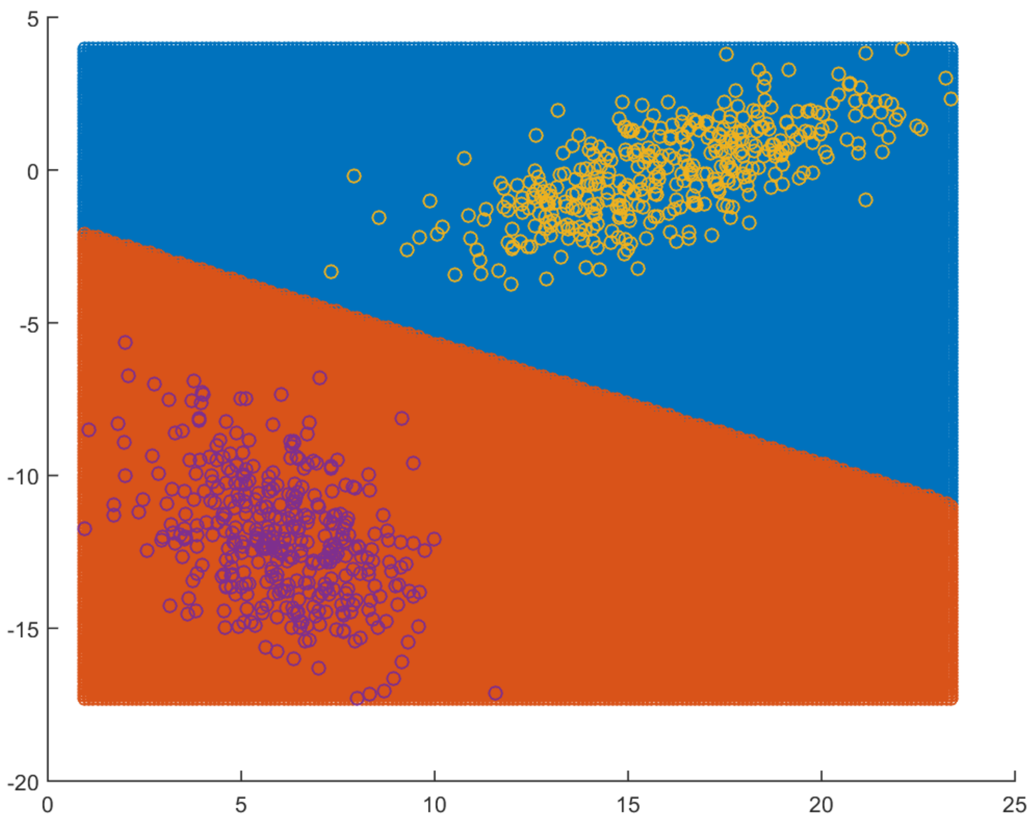
Classification Accuracy: - 100%

Confusion Matrix

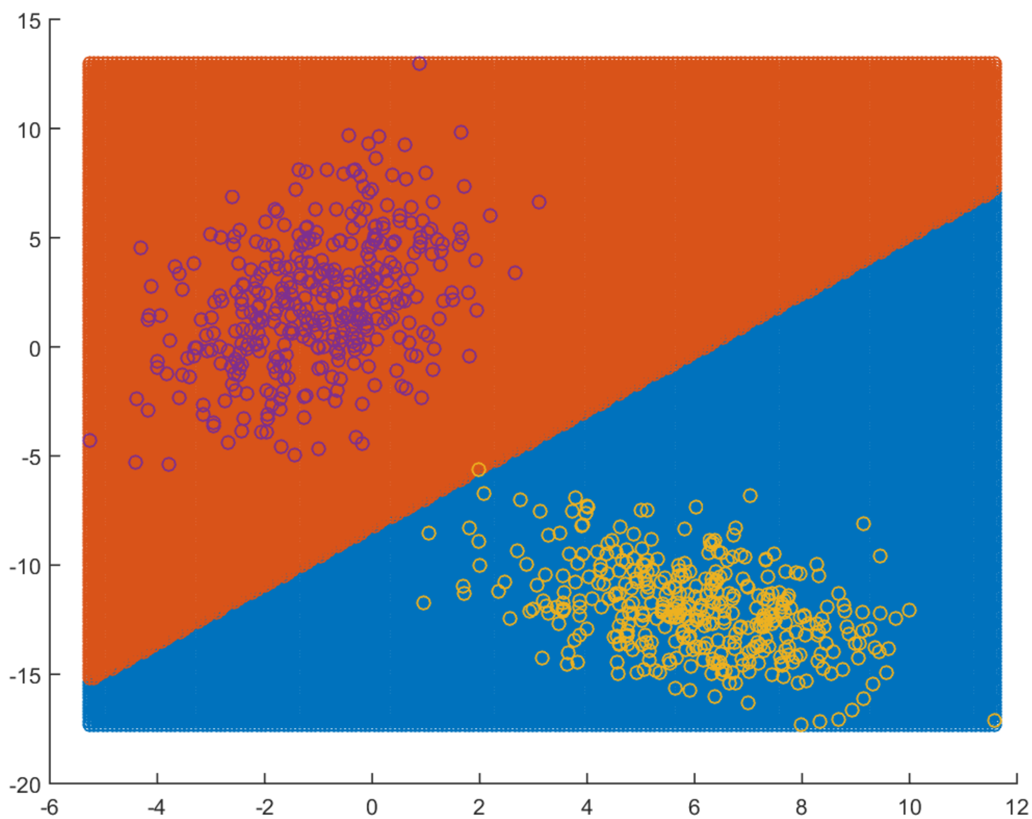
	Class1	Class2	Class3
Class1	125	0	0
Class2	0	125	0
Class3	0	0	125

Decision Region Plot for Pair

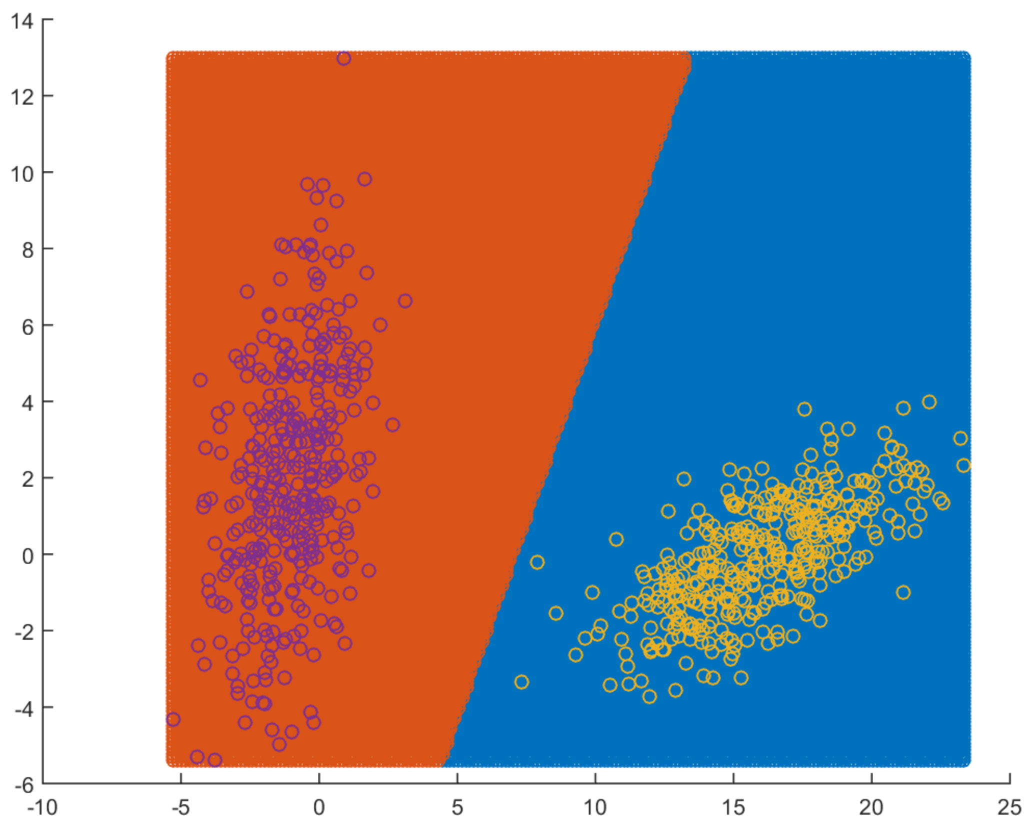
Class 1, Class 2



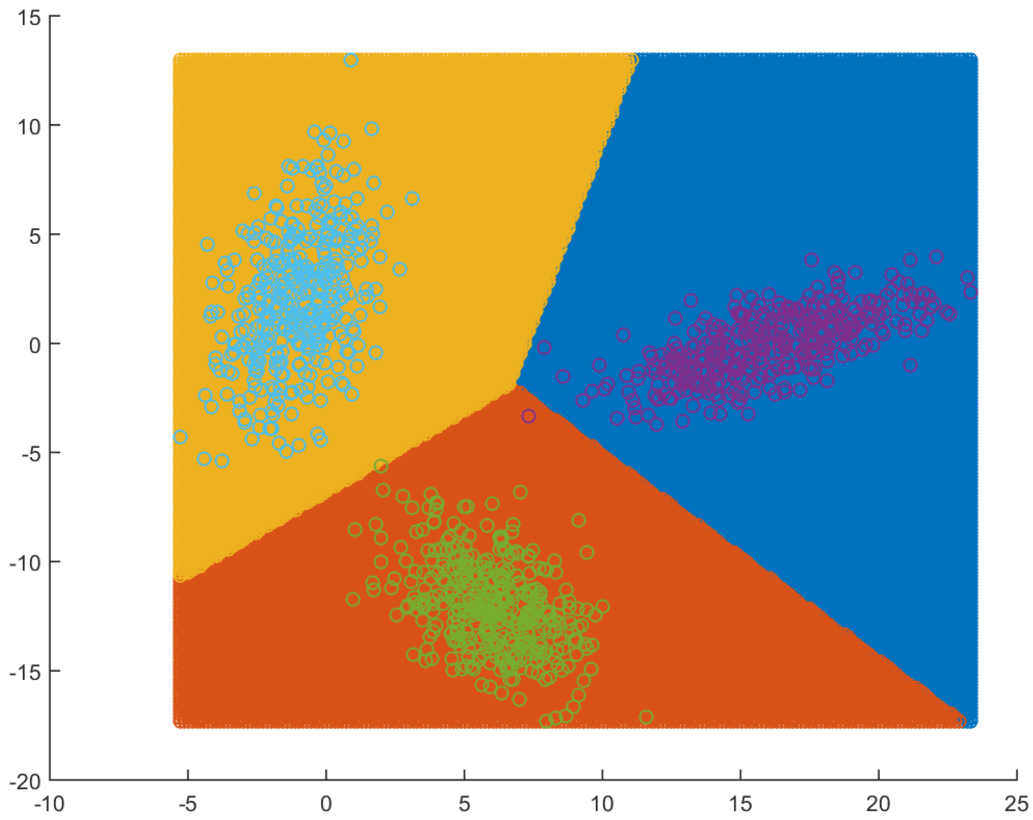
Class 2, Class 3



Class 1, Class 3



Decision Region Plot for All Classes



Observation

After making the covariance matrix same, the decision surface comes out to be linear. The line passes through the midpoint of the line joining the mean of the pair of classes.

The data being linearly separable, can be easily distinguished into different classes by using this classifier and hence the confusion matrix comes out to be a diagonal matrix with 100% correct recognition rate.

1.1.a.ii) Covariance Matrix from all Training Data Combined

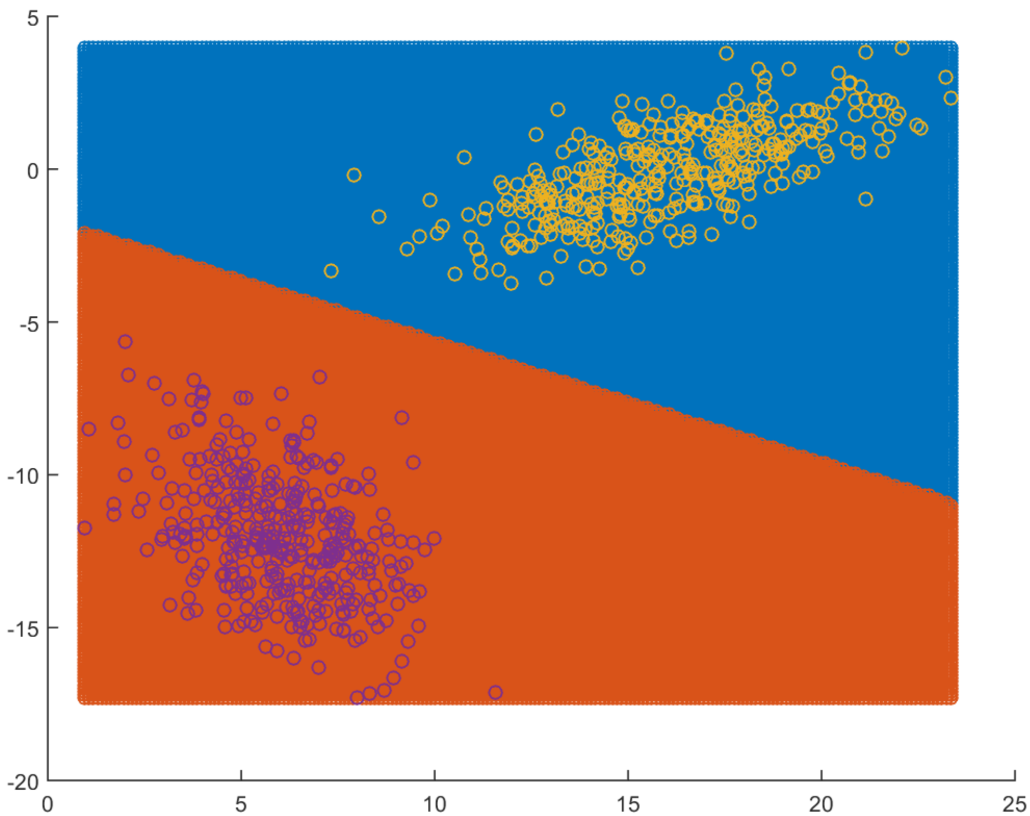
Classification Accuracy: - 100%

Confusion Matrix

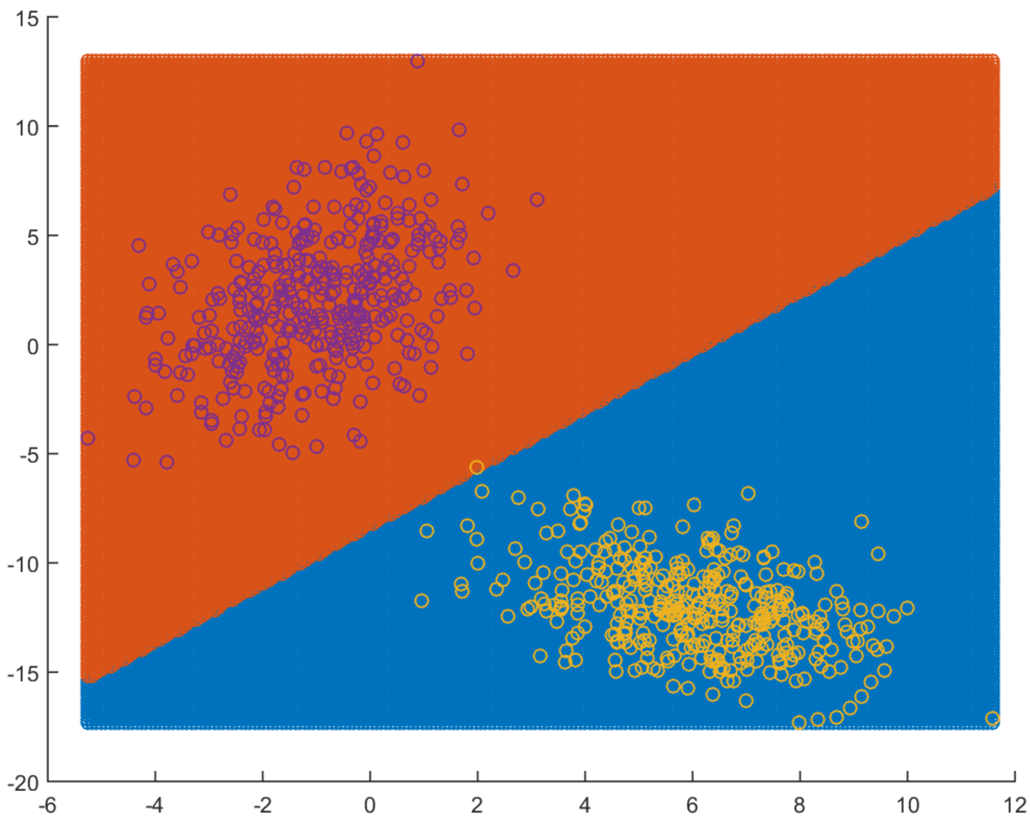
	Class1	Class2	Class3
Class1	125	0	0
Class2	0	125	0
Class3	0	0	125

Decision Region Plot for Pair

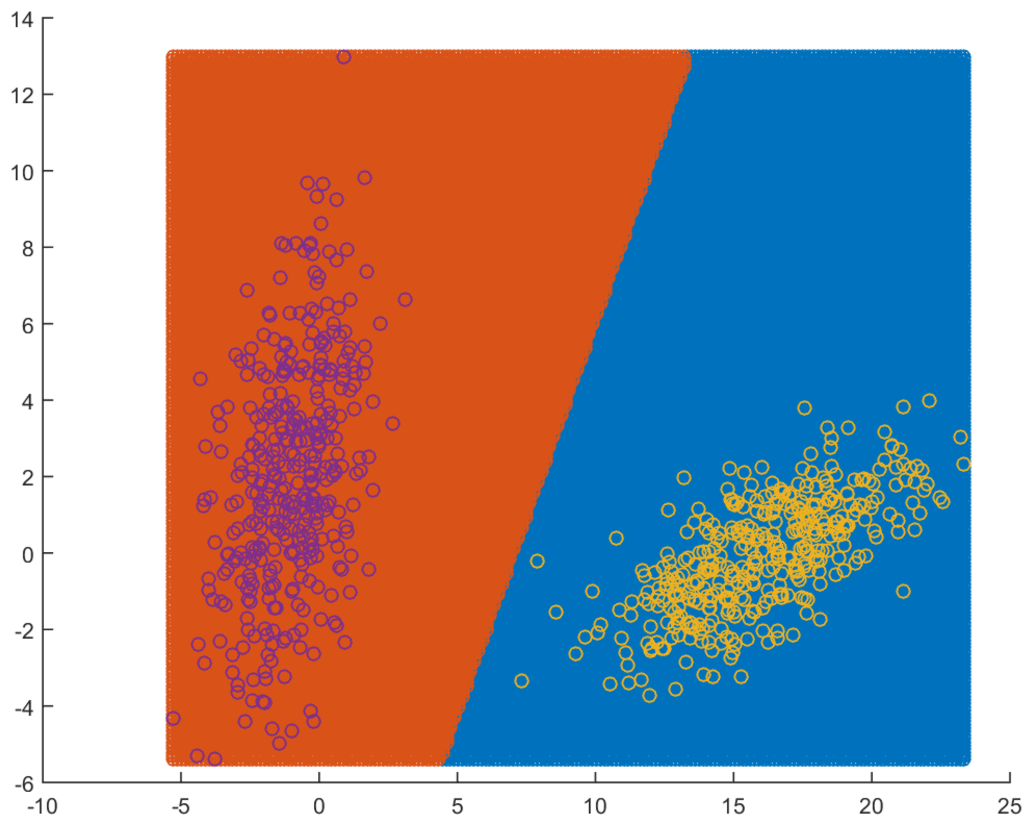
Class 1, Class 2



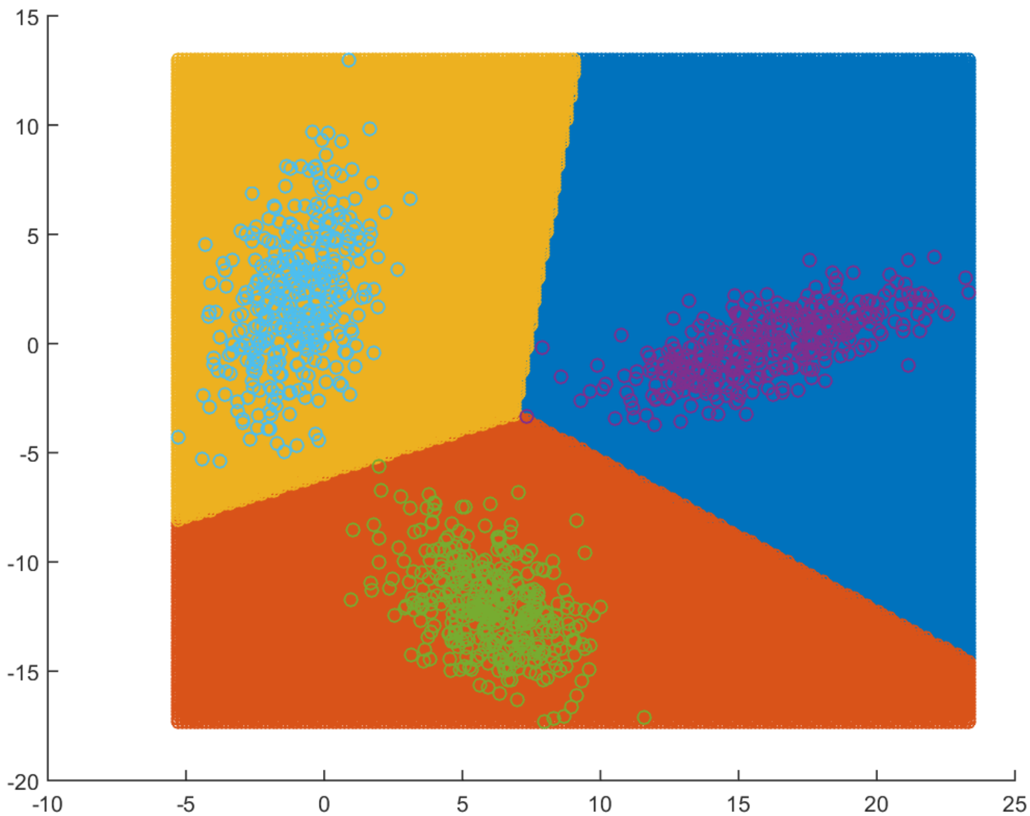
Class 2, Class 3



Class 1, Class 3



Decision Region Plot for All Classes



Observation

After making the covariance matrix same, the decision surface comes out to be linear. The slope for the lines changes as compared to previous case as the method for calculating the covariance is different for both.

The data being linearly separable, can be easily distinguished into different classes by using this classifier and hence the confusion matrix comes out to be a diagonal matrix with 100% correct recognition rate.

1.1.b) Covariance Matrix Different

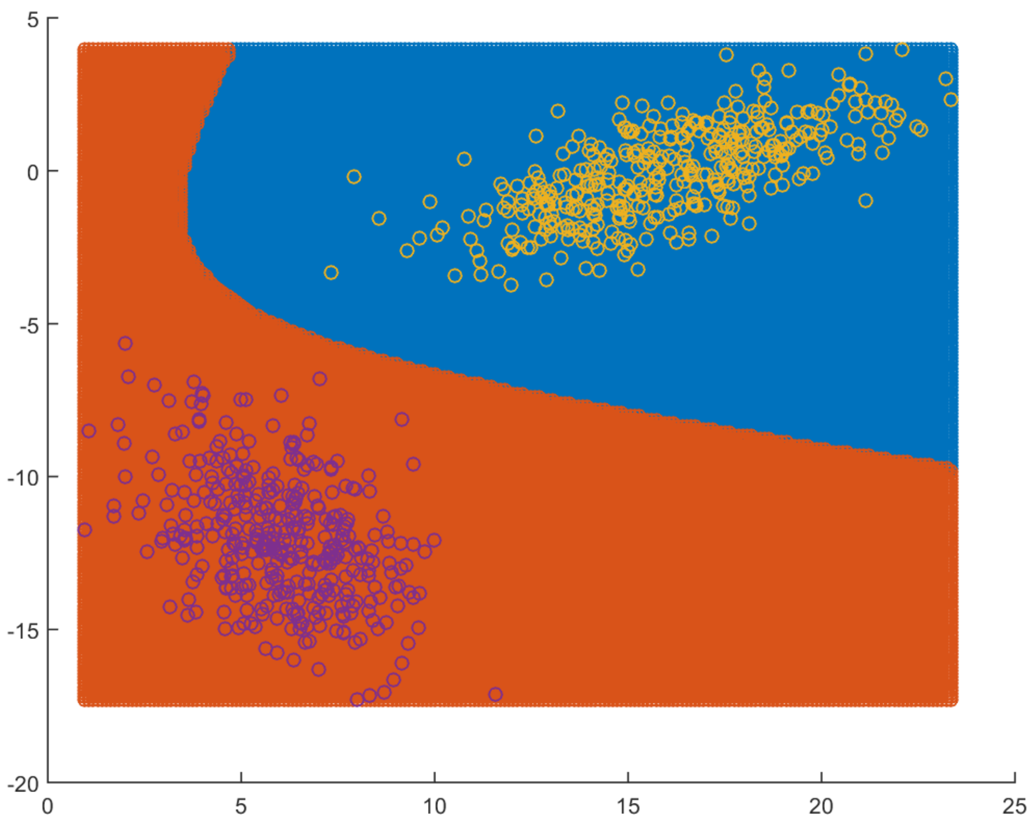
Classification Accuracy: - 100%

Confusion Matrix

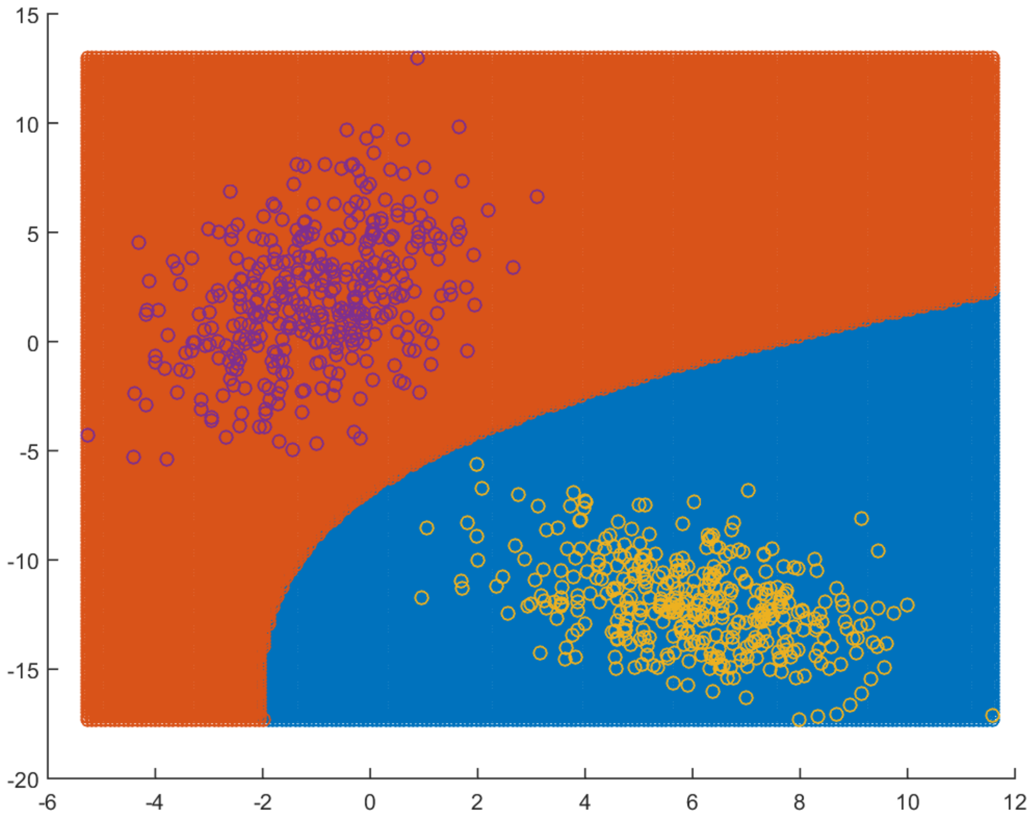
	Class1	Class2	Class3
Class1	125	0	0
Class2	0	125	0
Class3	0	0	125

Decision Region Plot for Pair

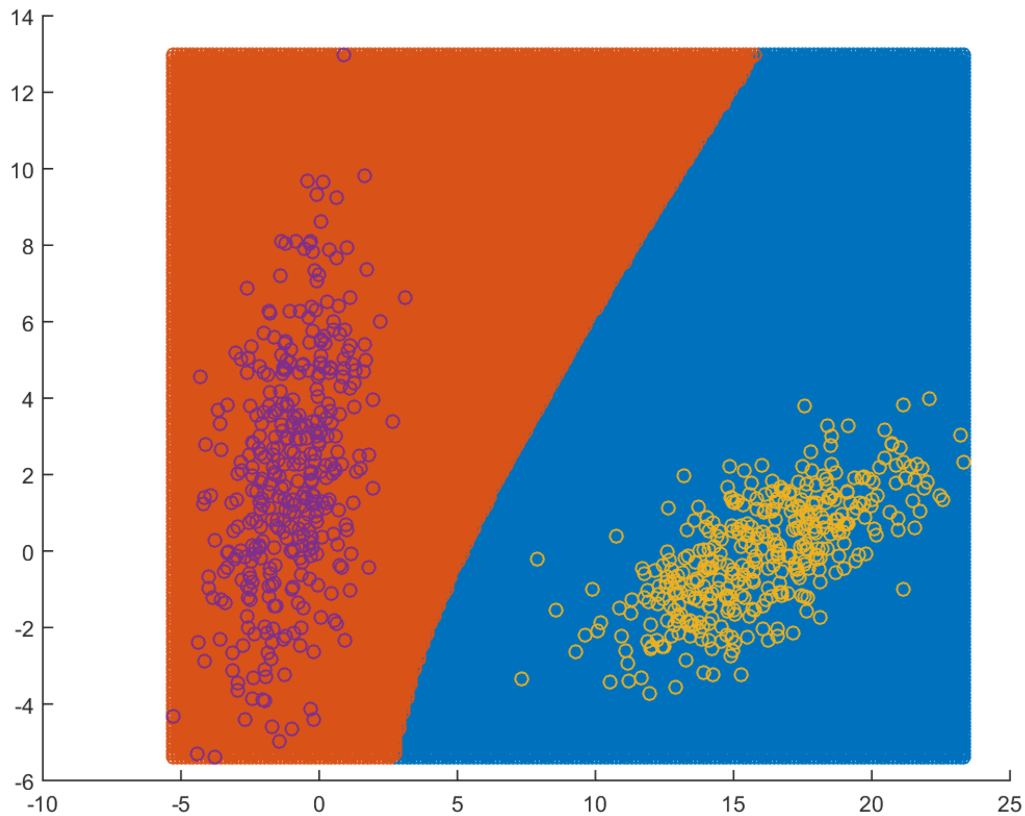
Class 1, Class 2



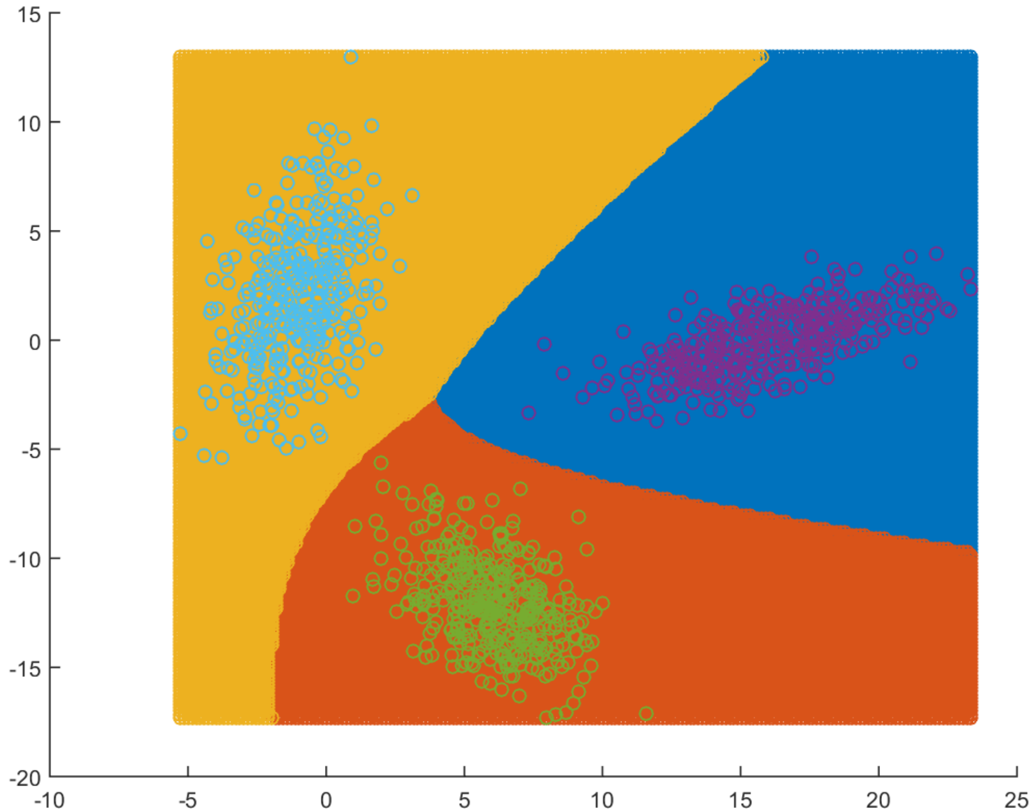
Class 2, Class 3



Class 1, Class 3



Decision Region Plot for All Classes



Observation

After making the covariance matrix different, the decision surface comes out to be hyperbolic.

The reason being that the covariance matrix is different for all the classes taken into consideration. The data could easily be distinguished and hence the confusion matrix comes out to be a diagonal matrix with 100% correct recognition rate.

1.2) Naïve Bayes Classifier

1.2.a) Covariance Matrix Same (σ^2I)

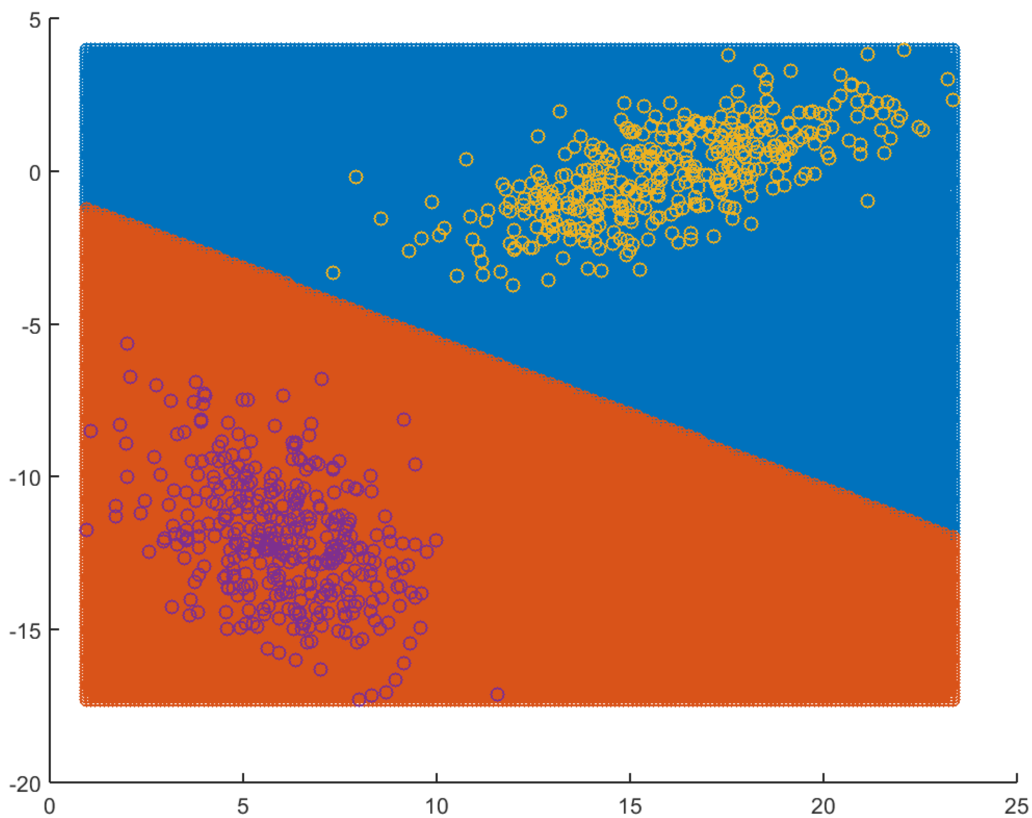
Classification Accuracy: - 100%

Confusion Matrix

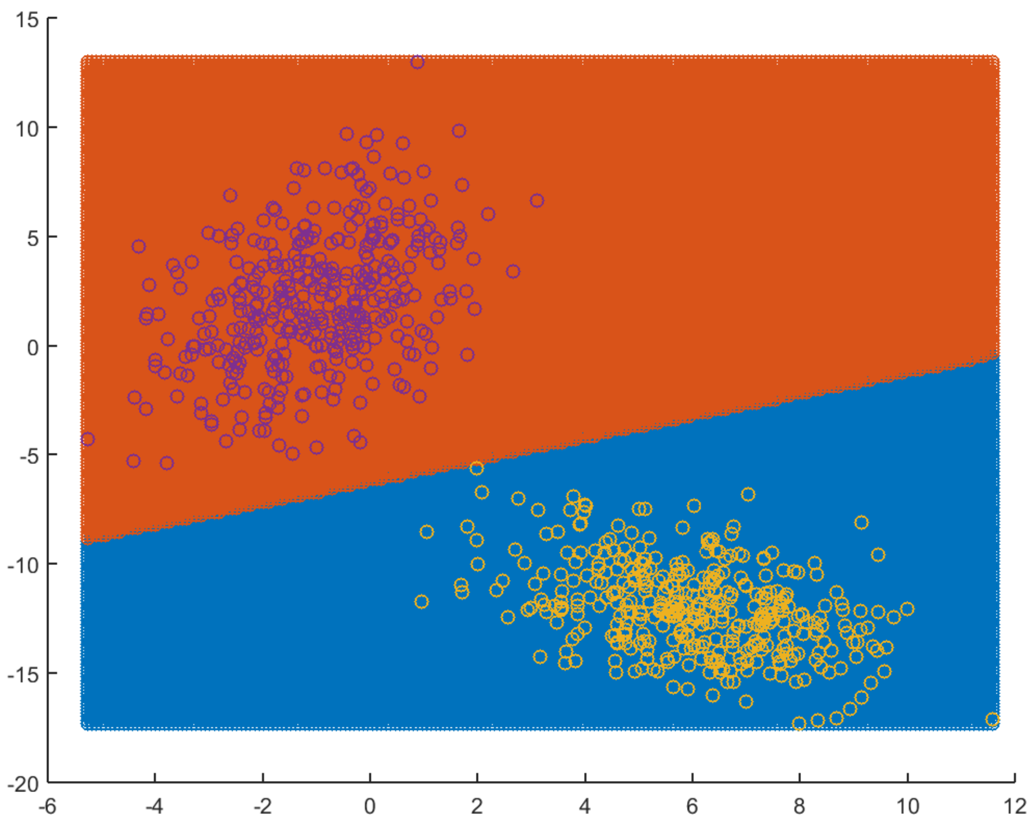
	Class1	Class2	Class3
Class1	125	0	0
Class2	0	125	0
Class3	0	0	125

Decision Region Plot for Pair

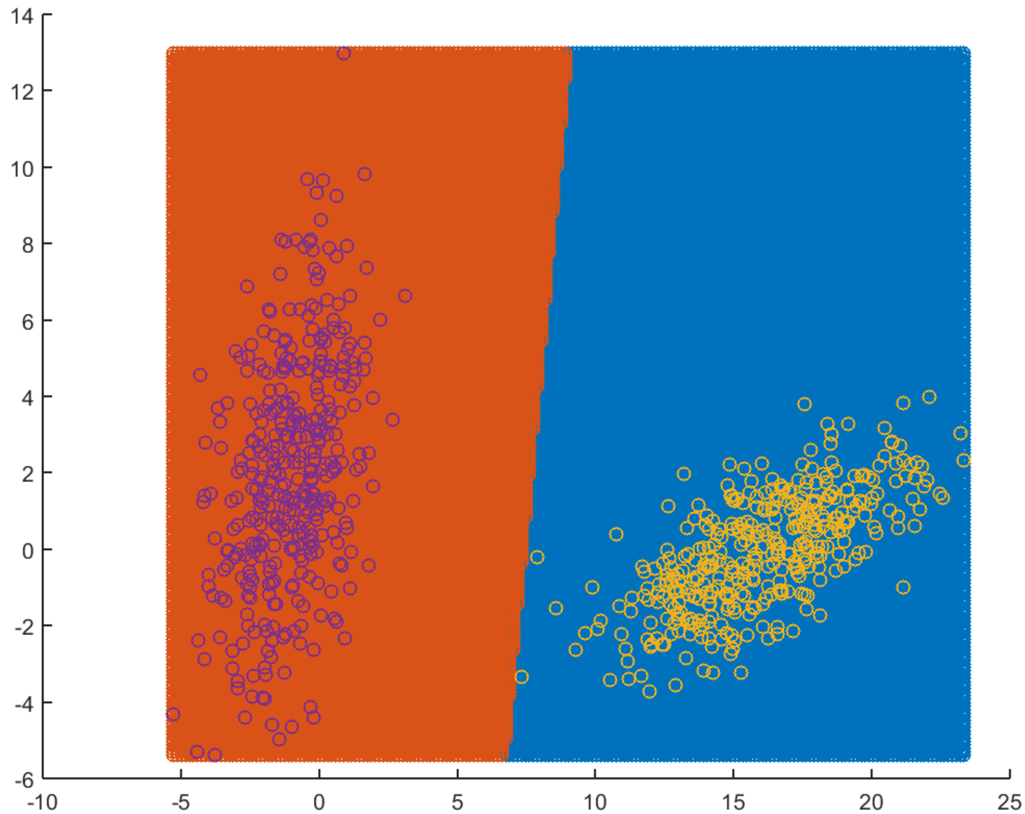
Class 1, Class 2



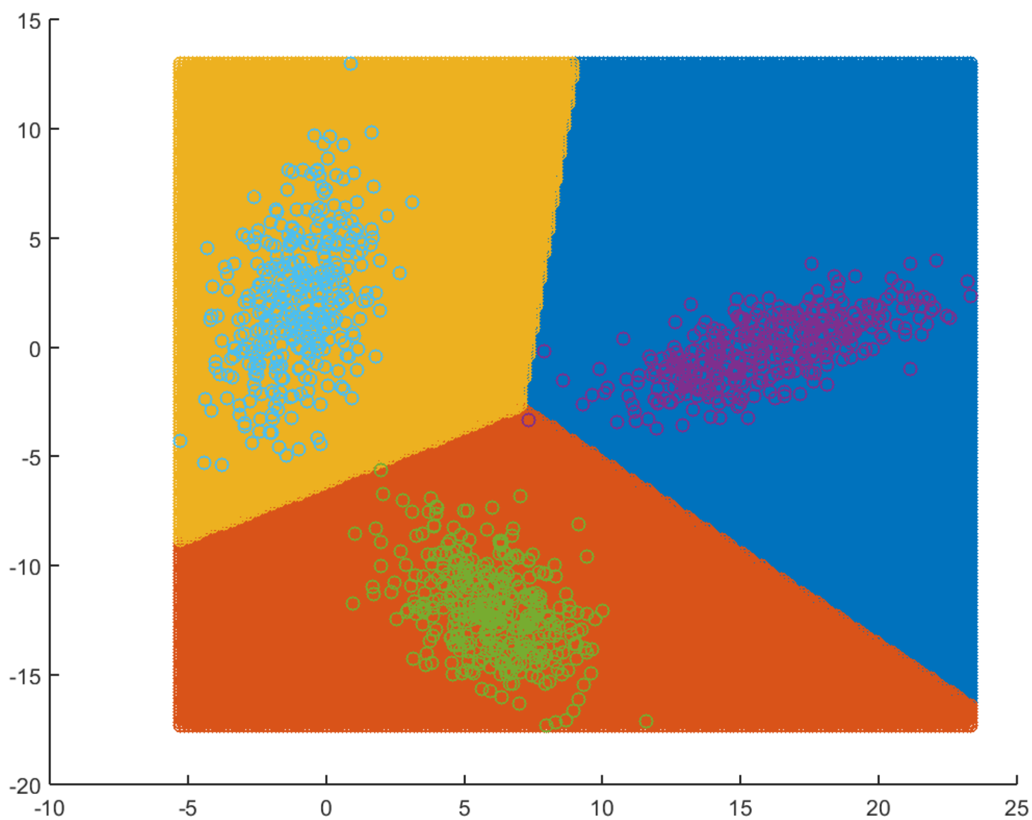
Class 2, Class 3



Class 1, Class 3



Decision Region Plot for All Classes



Observation

After making the covariance matrix same, the decision surface comes out to be linear. The line passes through the midpoint of the line joining the mean of the pair of classes and is perpendicular to it as we have taken covariance matrix of the form σ^2I.

The data being linearly separable, can be easily distinguished into different classes by using this classifier and hence the confusion matrix comes out to be a diagonal matrix with 100% correct recognition rate.

1.2.b) Covariance Matrix Same (C)

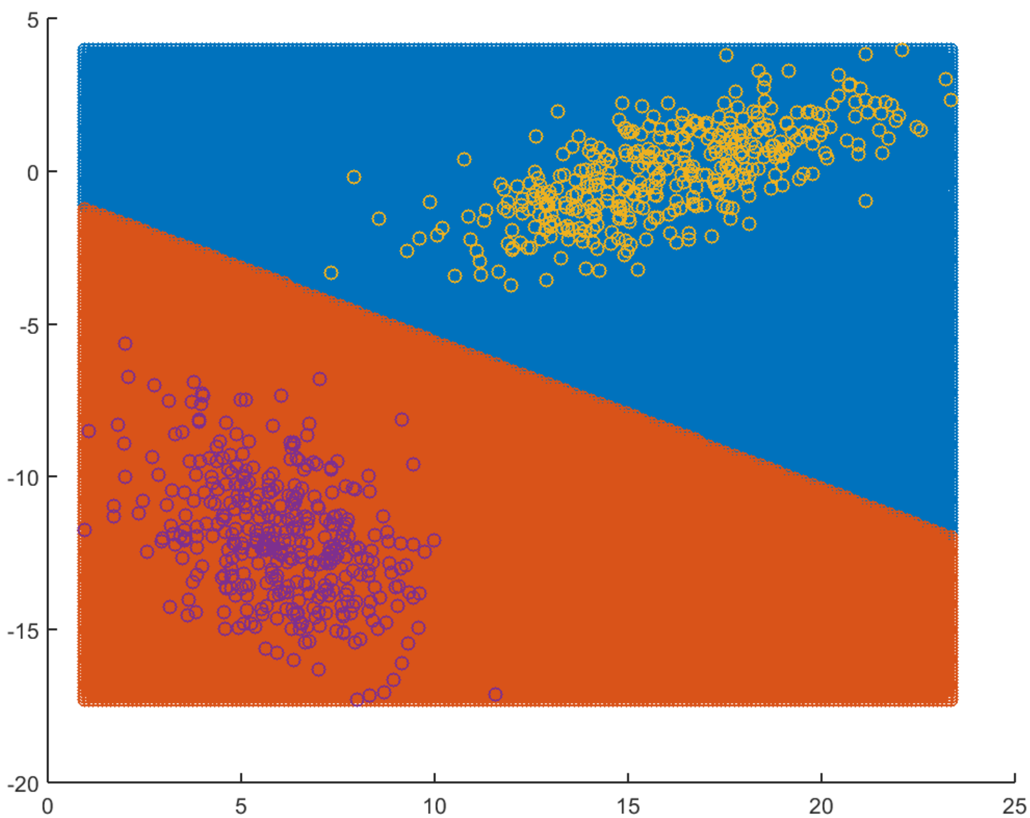
Classification Accuracy: - 100%

Confusion Matrix

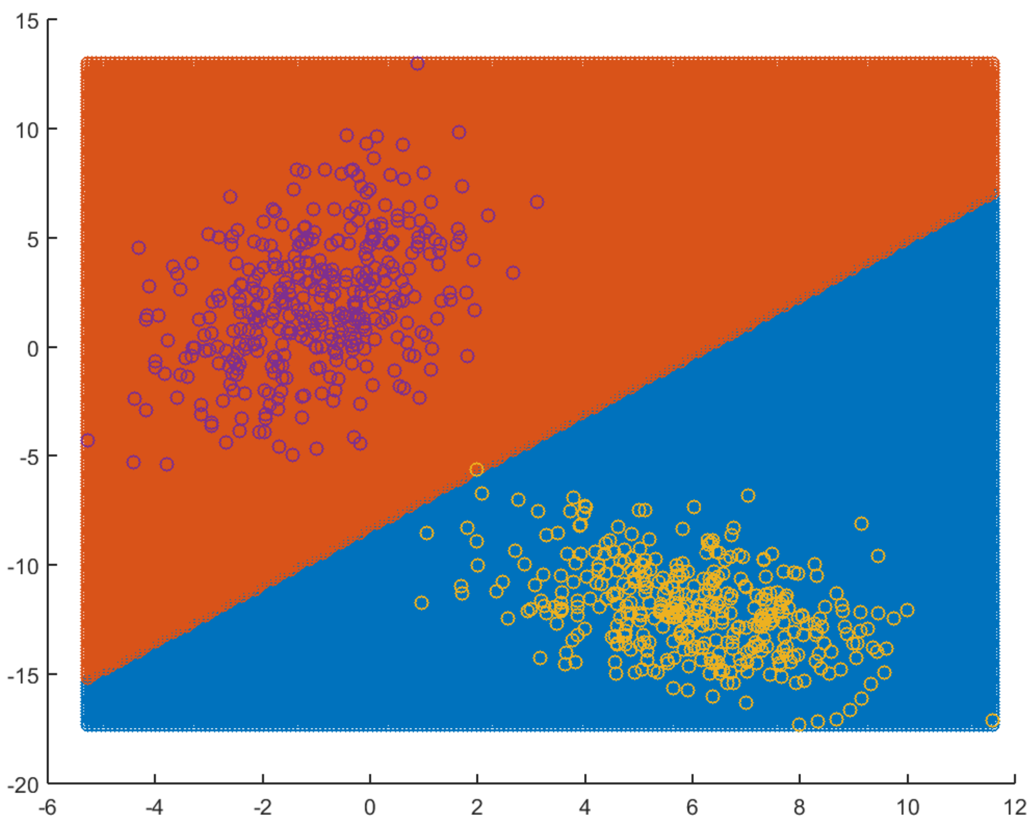
	Class1	Class2	Class3
Class1	125	0	0
Class2	0	125	0
Class3	0	0	125

Decision Region Plot for Pair

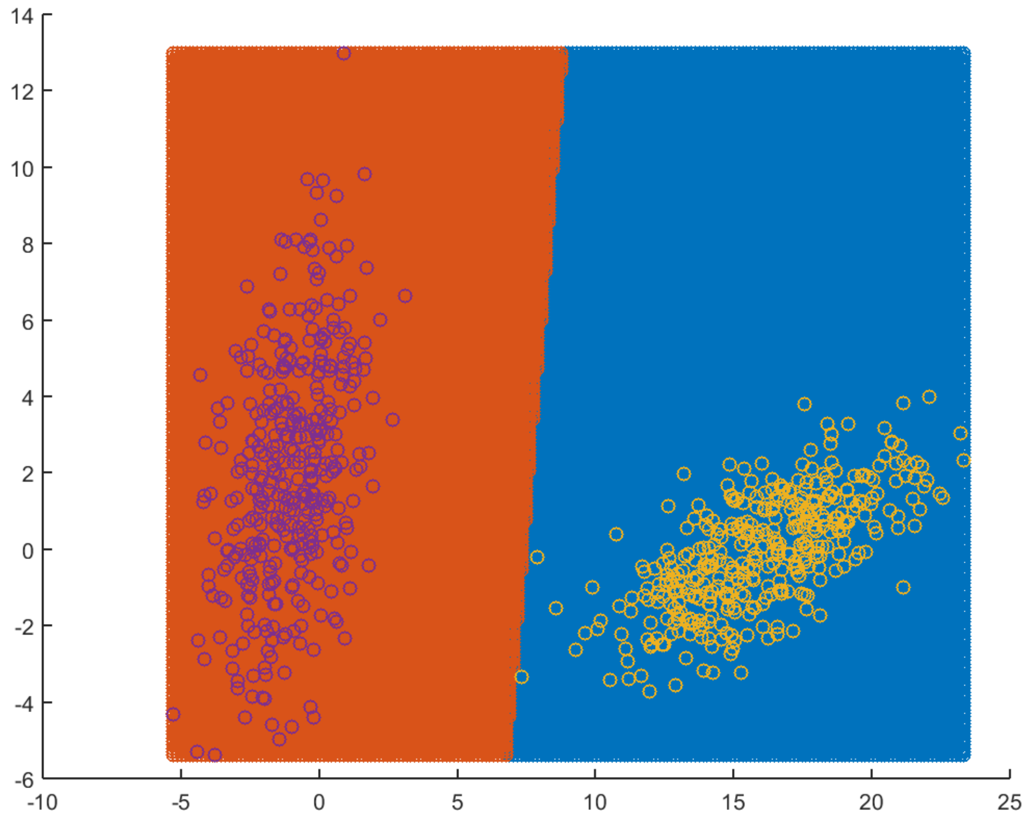
Class 1, Class 2



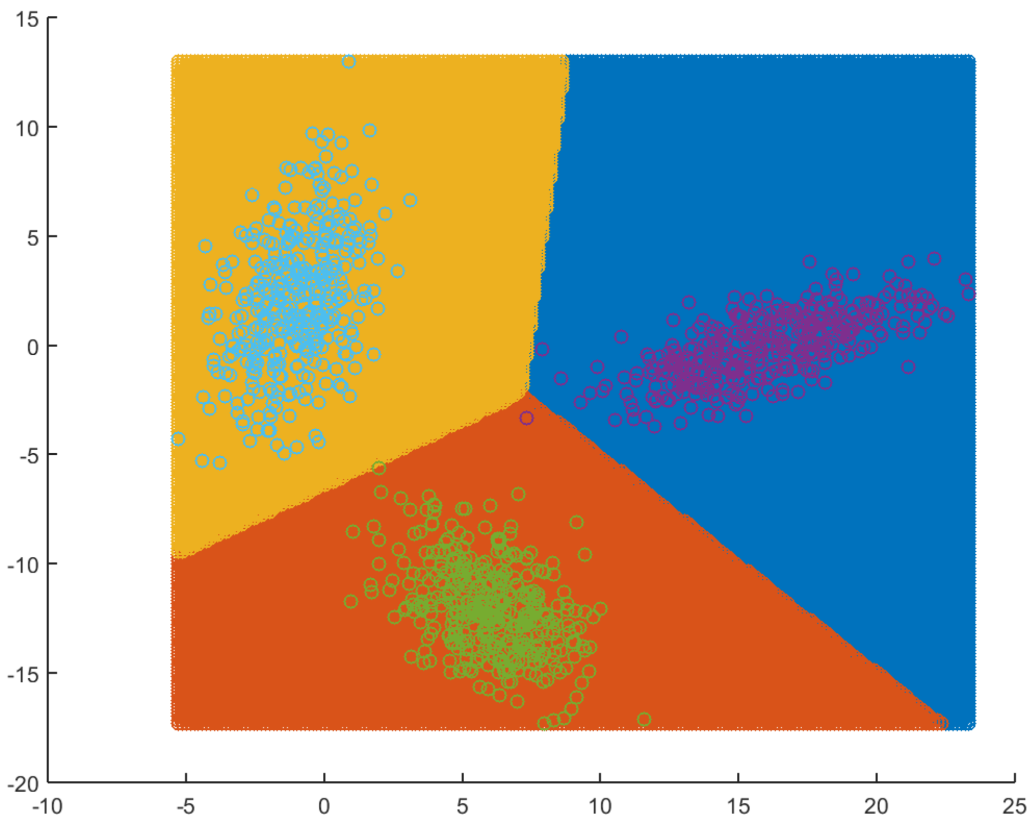
Class 2, Class 3



Class 1, Class 3



Decision Region Plot for All Classes



Observation

After making the covariance matrix same, the decision surface comes out to be linear. The line passes through the midpoint of the line joining the mean of the pair of classes as we have taken covariance matrix same.

The data being linearly separable, can be easily distinguished into different classes by using this classifier and hence the confusion matrix comes out to be a diagonal matrix with 100% correct recognition rate.

1.2.c) Covariance Matrix Different

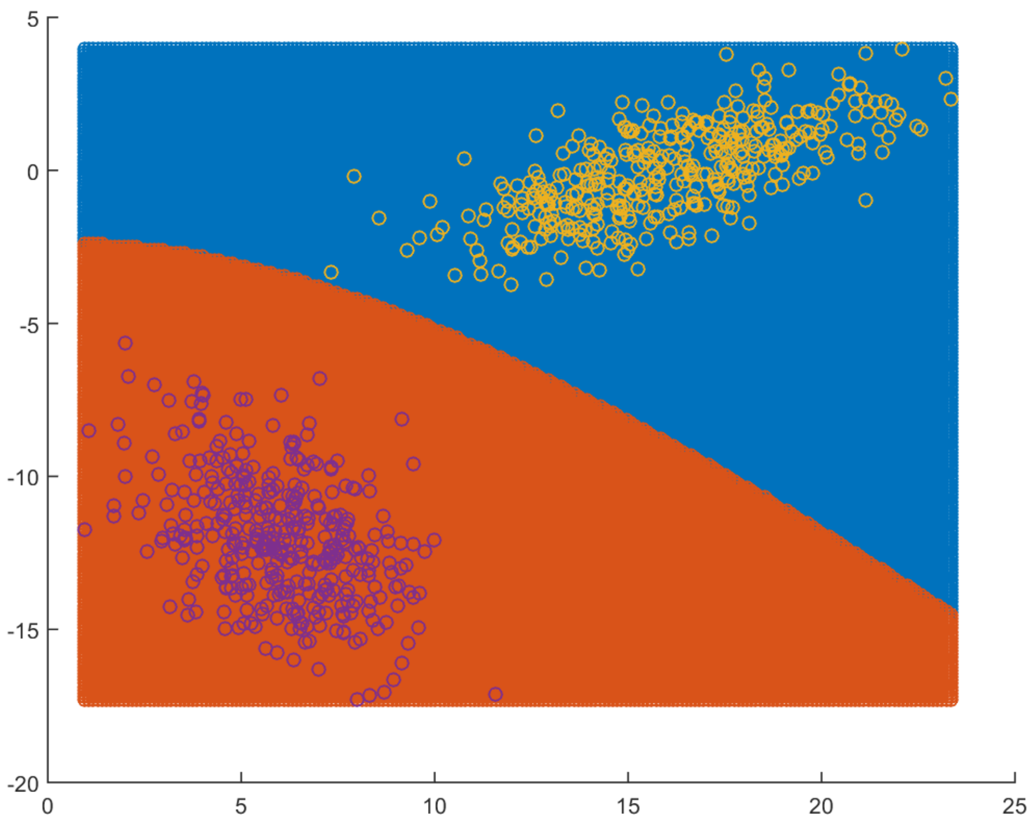
Classification Accuracy: - 99.733%

Confusion Matrix

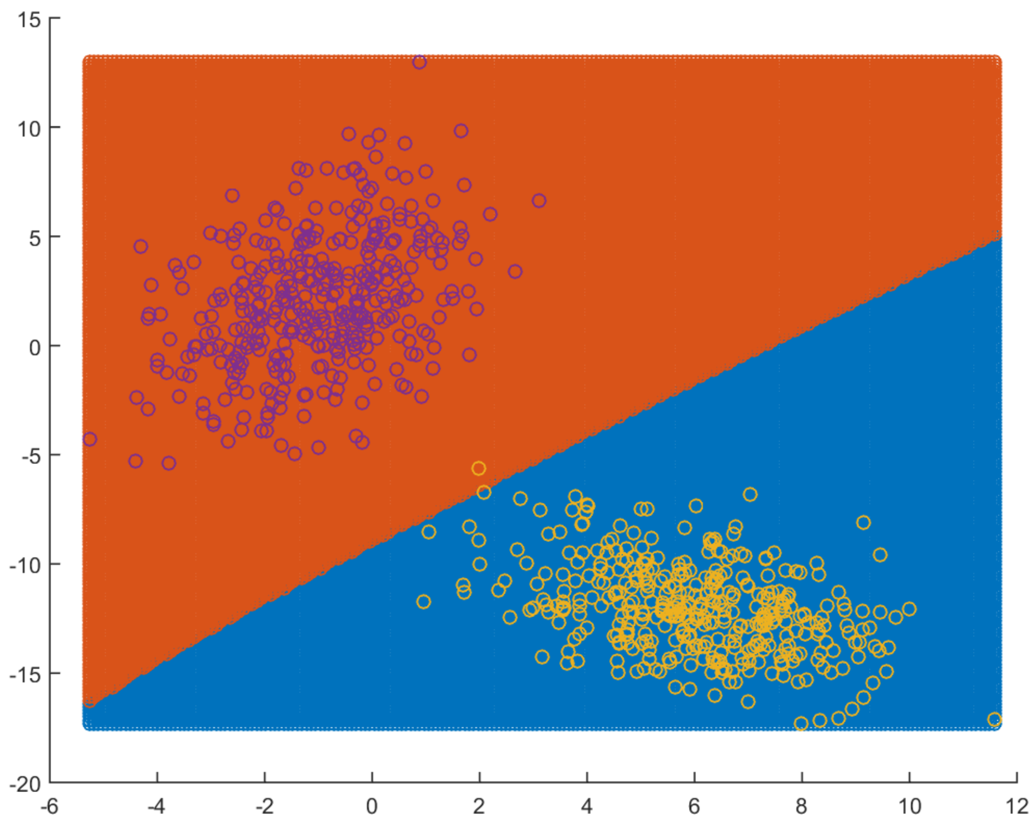
	Class1	Class2	Class3
Class1	125	0	0
Class2	0	124	0
Class3	0	0	125

Decision Region Plot for Pair

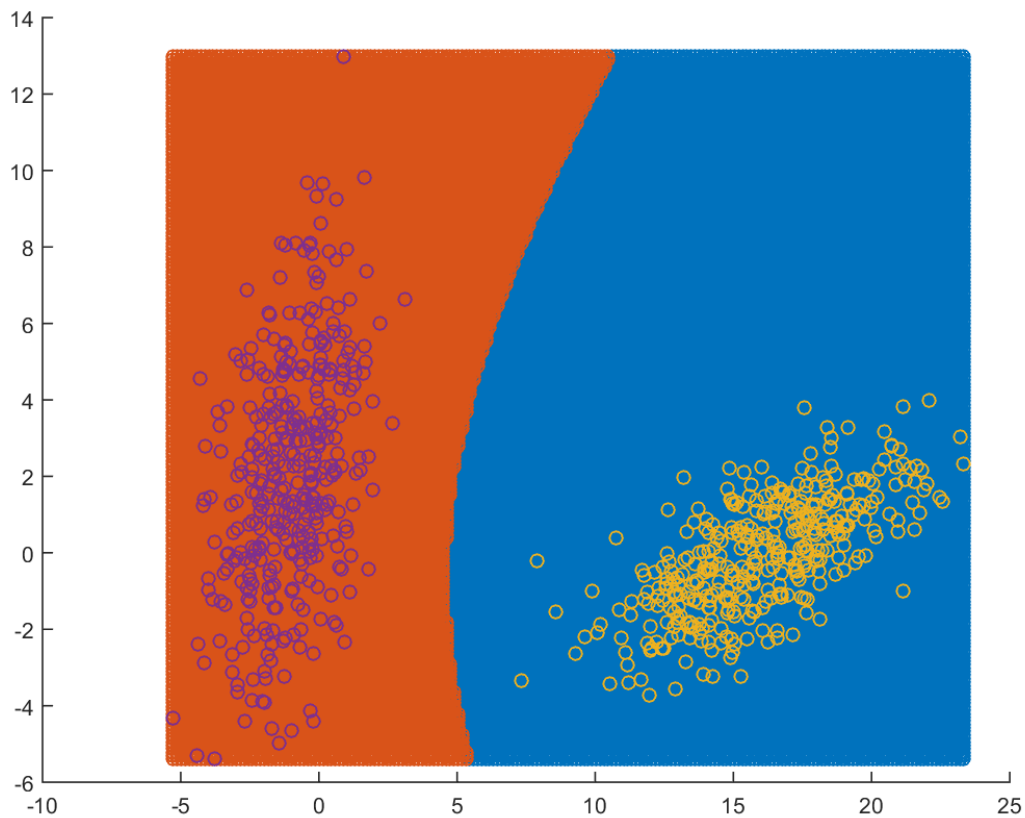
Class 1, Class 2



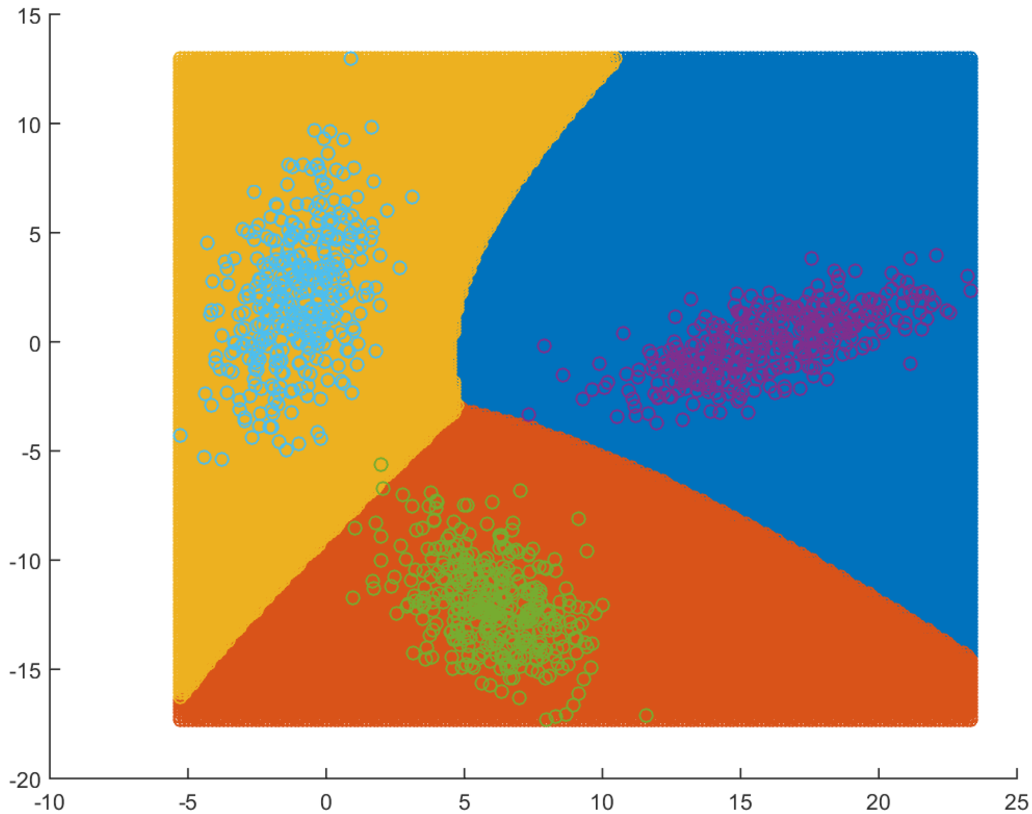
Class 2, Class 3



Class 1, Class 3



Decision Region Plot for All Classes



Observation

After making the covariance matrix different, the decision surface comes out to be hyperbolic.

The reason being that the covariance matrix is different for all the classes taken into consideration. The data could easily be distinguished with about 100% correct recognition rate.

There was one data point which could not be identified.
As compared to Bayes classifier, two decision surfaces are less curved and one is more curved.
This is due to the face that the characteristics are considered independent in this case.

2) Non-linearly Separable

2.1) Interlocking Classes

2.1.1) Bayes Classifier

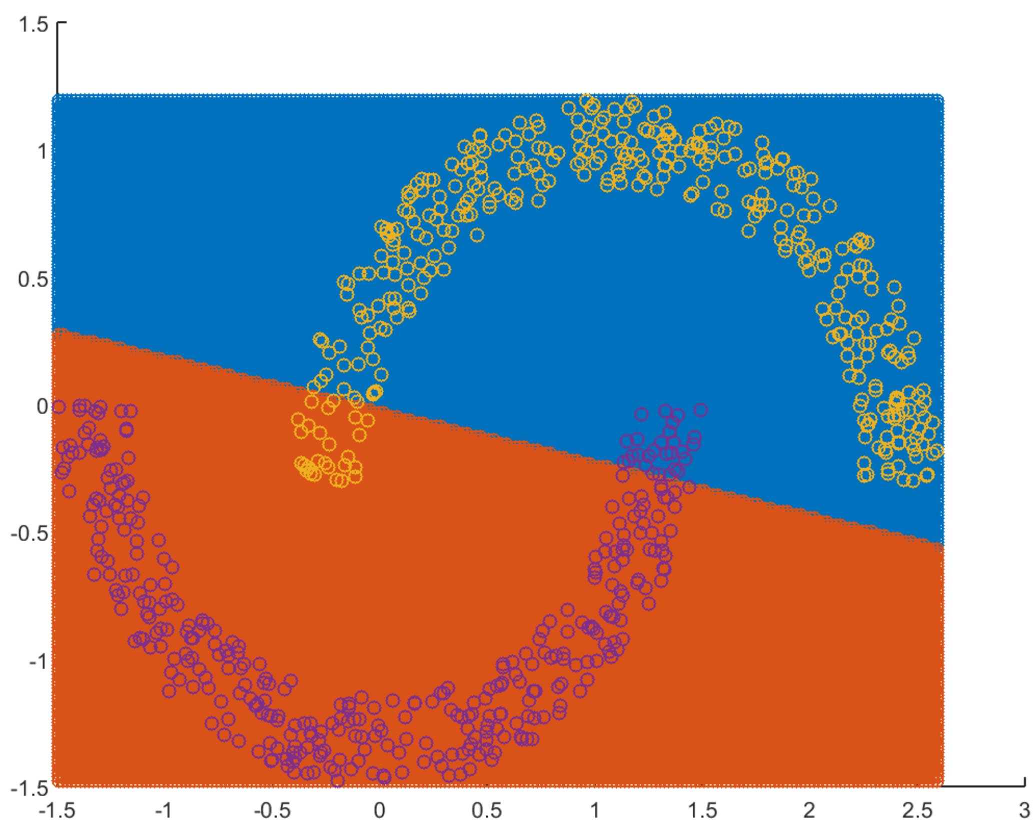
2.1.1.a.i) Average of Covariance Matrices of All Classes

Classification Accuracy :- 95.6 %

Confusion Matrix

	Class 1	Class 2
Class 1	118	7
Class 2	4	121

Decision Region Plot for All Classes



Observation

After making the covariance matrix same, the decision surface comes out to be linear. The line passes through the midpoint of the line joining the mean of the pair of classes.

The data is not linearly separable, there were some data points that could not be distinguished.

The accuracy rate comes out to be 95.6%.

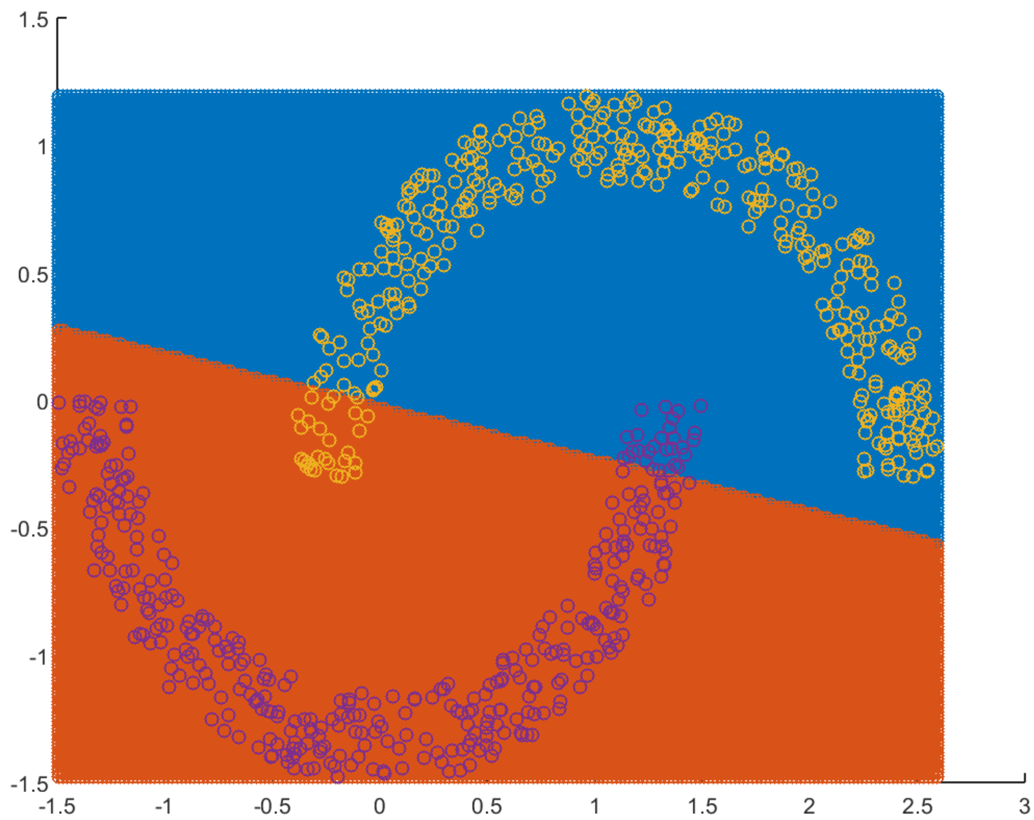
2.1.1.a.ii) Covariance Matrix from all Training Data Combined

Classification Accuracy :- 95.6 %

Confusion Matrix

	Class 1	**Class 2**
Class 1	**118**	**7**
Class 2	**4**	**121**

Decision Region Plot for All Classes



Observation

After making the covariance matrix same, the decision surface comes out to be linear. The line passes through the midpoint of the line joining the mean of the pair of classes.

The data is not linearly separable, there were some data points that could not be distinguished.

The accuracy rate comes out to be 95.6%.

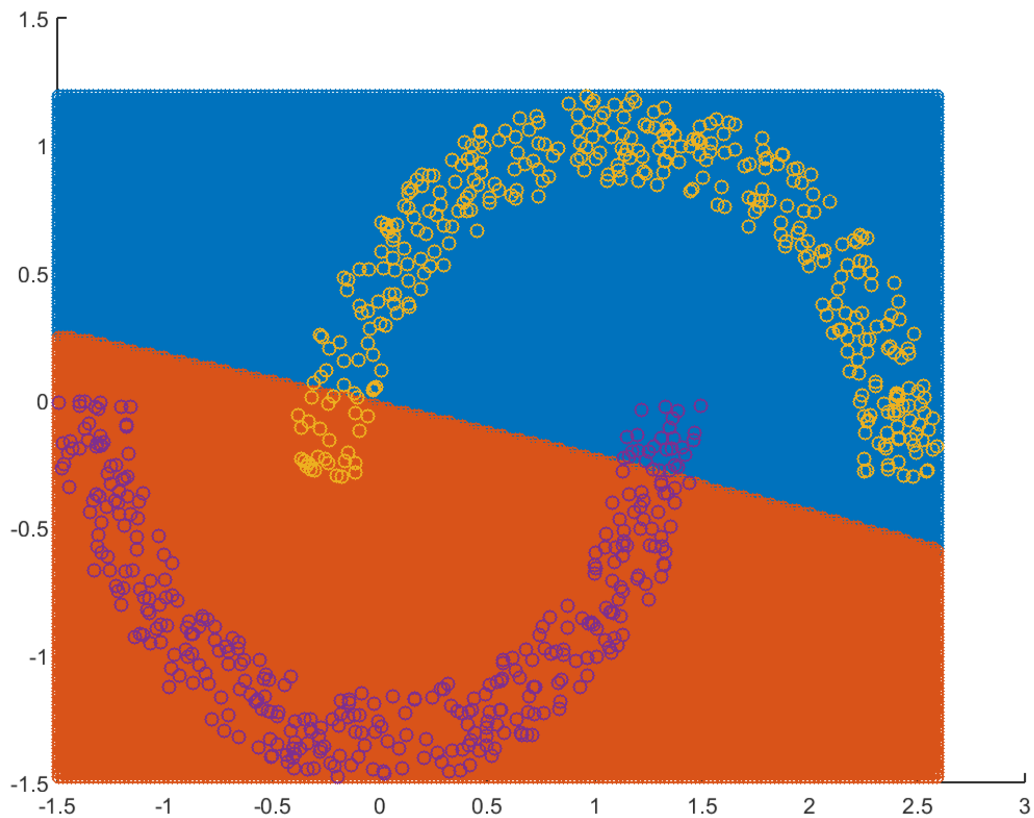
2.1.1.b) Covariance Matrix Different

Classification Accuracy :- 96.0 %

Confusion Matrix

	Class 1	Class 2
Class 1	**118**	**7**
Class 2	**3**	**122**

Decision Region Plot for All Classes



Observation

The covariance matric was different, the decision surface comes out to be hyperbolic but it does appear like a straight line in the given range of data.

The data is not linearly separable, there were some data points that could not be distinguished.

The accuracy rate comes out to be 96% so we can say that there was no significant increase in the accuracy of the classifier.

2.1.2) Naïve Bayes Classifier

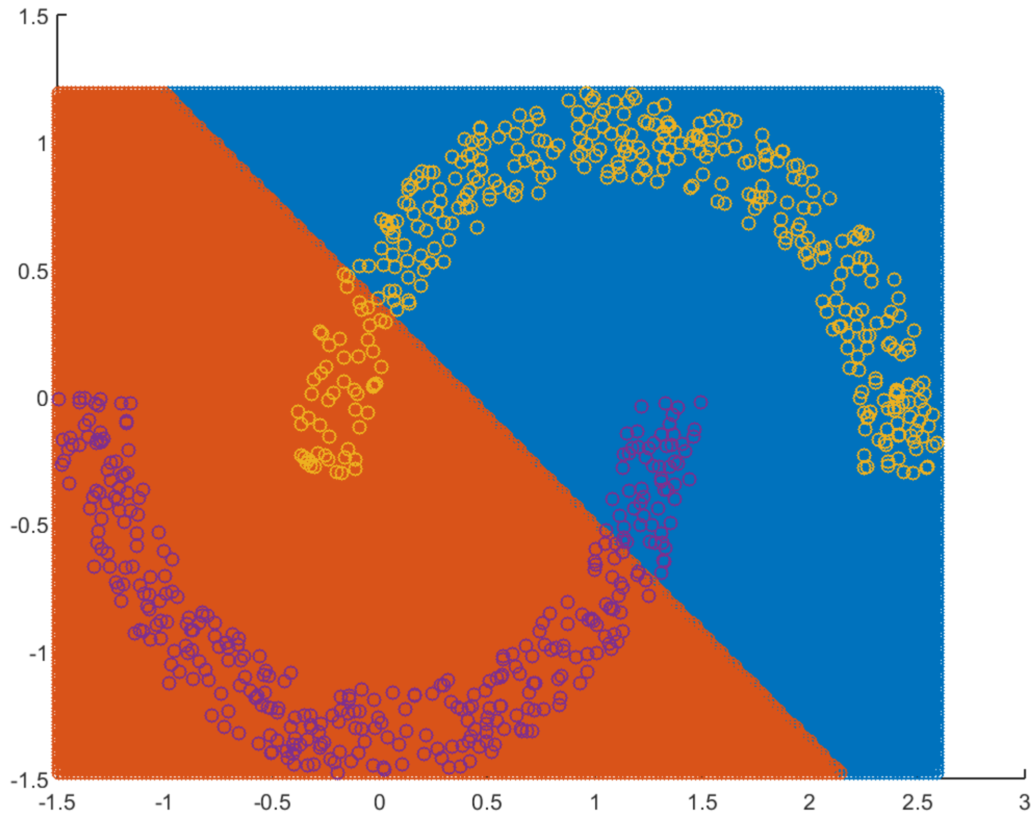
2.1.2.a) Covariance Matrix Same (σ^2I)

Classification Accuracy :- 87.6 %

Confusion Matrix

	Class 1	**Class 2**
Class 1	**106**	**19**
Class 2	**12**	**113**

Decision Region Plot for All Classes



Observation

After making the covariance matrix same, the decision surface comes out to be linear. The line passes through the midpoint of the line joining the mean of the pair of classes and is perpendicular to it.

The data is not linearly separable, there were some data points that could not be distinguished.

The accuracy rate comes out to be 87.6% which is worse than Bayes classifier as it is not taking cross covariance into account.

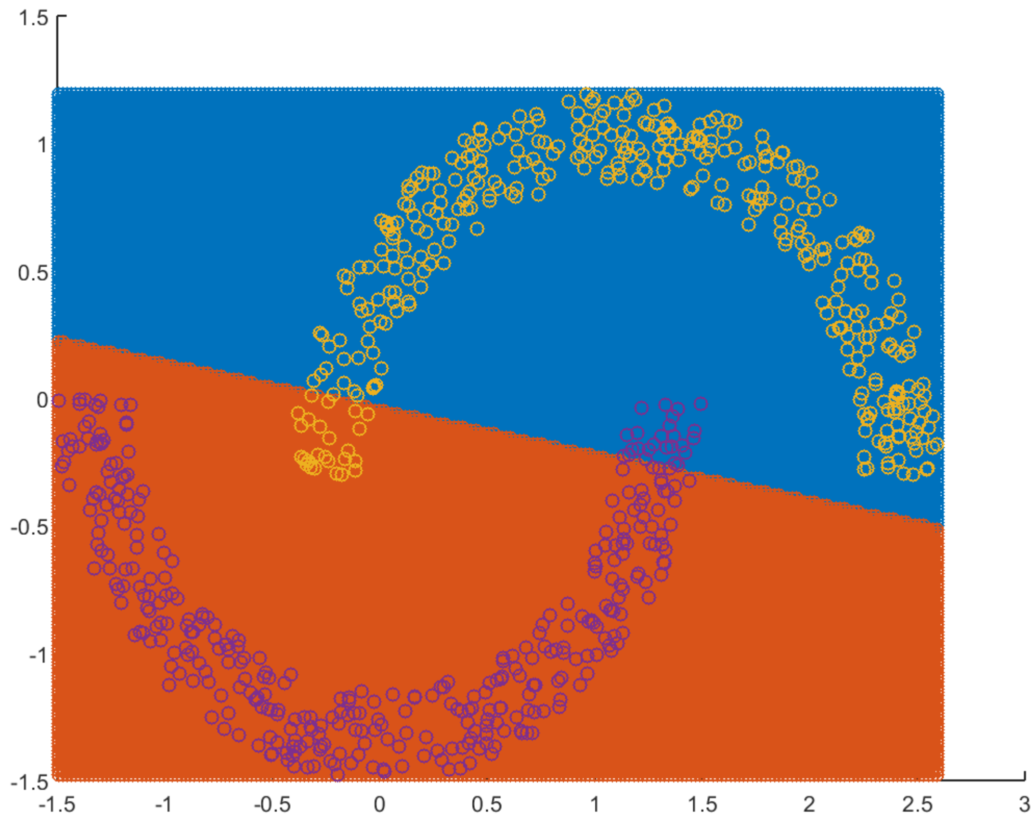
2.1.2.b) Covariance Matrix Same (C)

Classification Accuracy :- 96 %

Confusion Matrix

	Class 1	Class 2
Class 1	118	7
Class 2	3	122

Decision Region Plot for All Classes



Observation

After making the covariance matrix same, the decision surface comes out to be linear. The line passes through the midpoint of the line joining the mean of the pair of classes. The data is not linearly separable, there were some data points that could not be distinguished.

The accuracy rate comes out to be 96%.

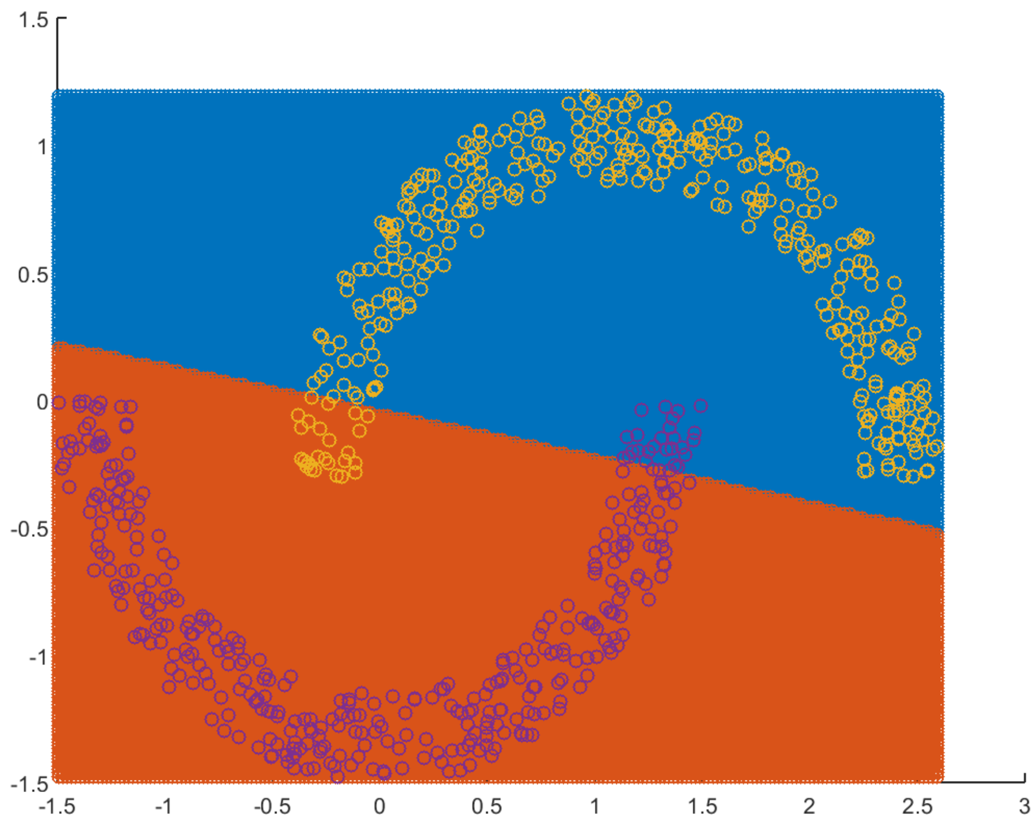
2.1.2.c) Covariance Matrix Different

Classification Accuracy :- 96 %

Confusion Matrix

	Class 1	**Class 2**
Class 1	**118**	**7**
Class 2	**3**	**122**

Decision Region Plot for All Classes



Observation

The covariance matric was different, the decision surface comes out to be hyperbolic but it does appear like a straight line in the given range of data.

The data is not linearly separable, there were some data points that could not be distinguished.

The accuracy rate comes out to be 96% so we can say that there was no significant increase in the accuracy of the classifier.

2.2) Ring (Central Mass)

2.2.1) Bayes Classifier

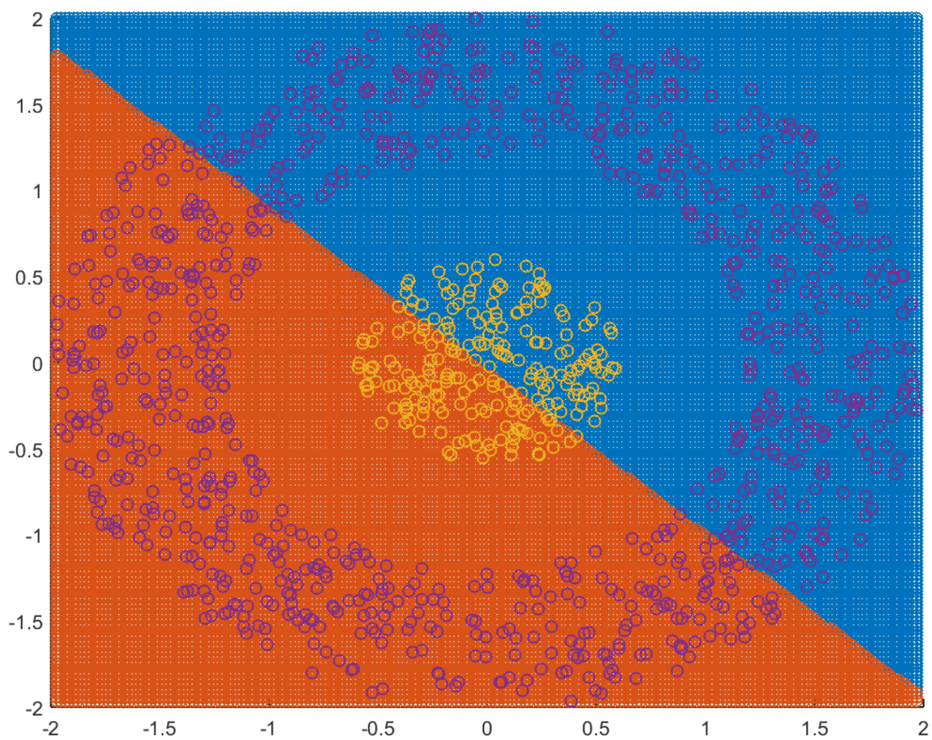
2.2.1.a.i) Average of Covariance Matrices of All Classes

Classification Accuracy :- 50.133 %

Confusion Matrix

	Class 1	Class 2
Class 1	46	29
Class 2	158	142

Decision Region Plot for All Classes



Observation

As the covariance matrix was same for all, the decision surface had to be linear. But the data being in the form of a ring could not be separated by such a surface. Hence the classifier performed very badly with only about 50% accuracy.

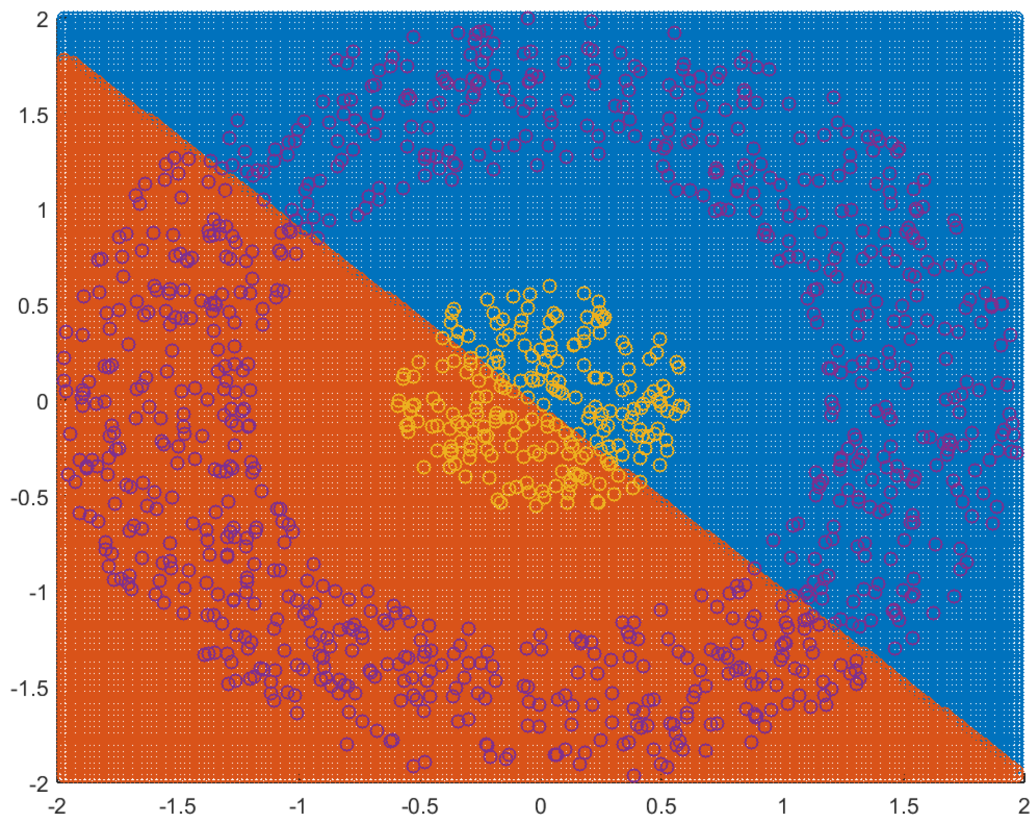
2.2.1.a.ii) Covariance Matrix from all Training Data Combined

Classification Accuracy :- 50.133 %

Confusion Matrix

	Class 1	**Class 2**
Class 1	**46**	**29**
Class 2	**158**	**142**

Decision Region Plot for All Classes



Observation

As the covariance matrix was same for all, the decision surface had to be linear. But the data being in the form of a ring could not be separated by such a surface. Hence the classifier performed very badly with only about 50% accuracy.

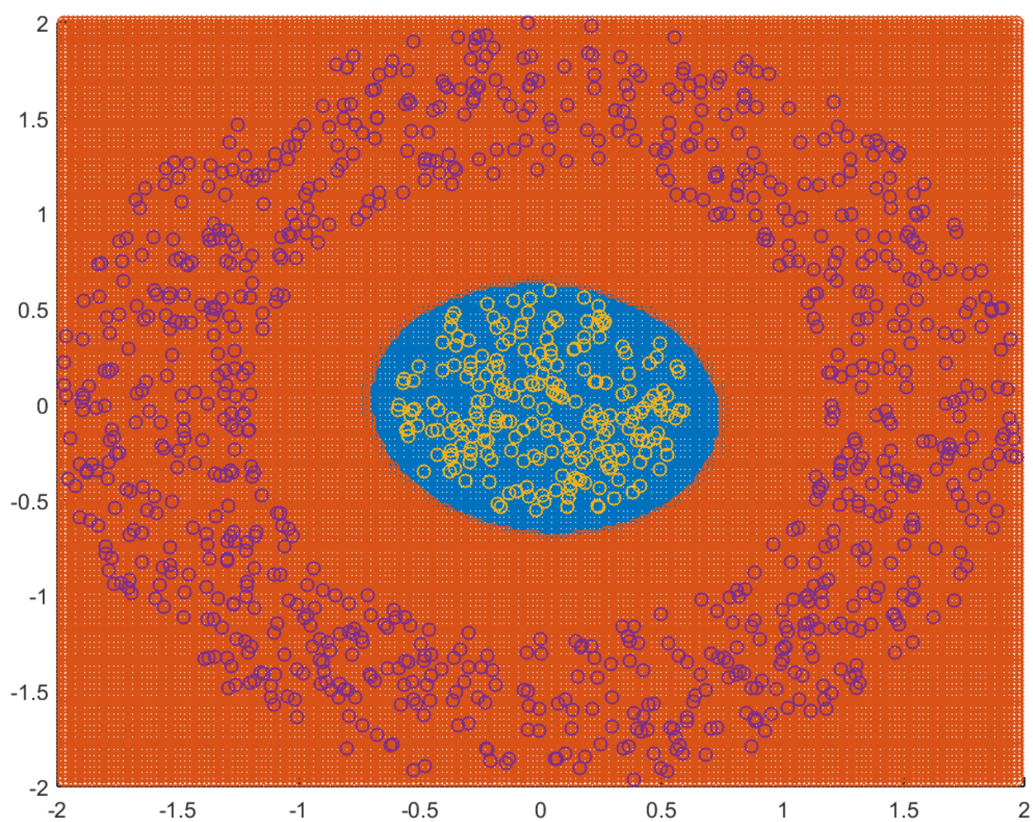
2.2.1.b) Covariance Matrix Different

Classification Accuracy :- 100 %

Confusion Matrix

	Class 1	**Class 2**
Class 1	**75**	**0**
Class 2	**0**	**300**

Decision Region Plot for All Classes



Observation

When we took different covariance matrix for both the classes, the classifier becomes 100% accurate due the elliptical decision surface it was able to produce.

2.2.2) Naïve Bayes Classifier

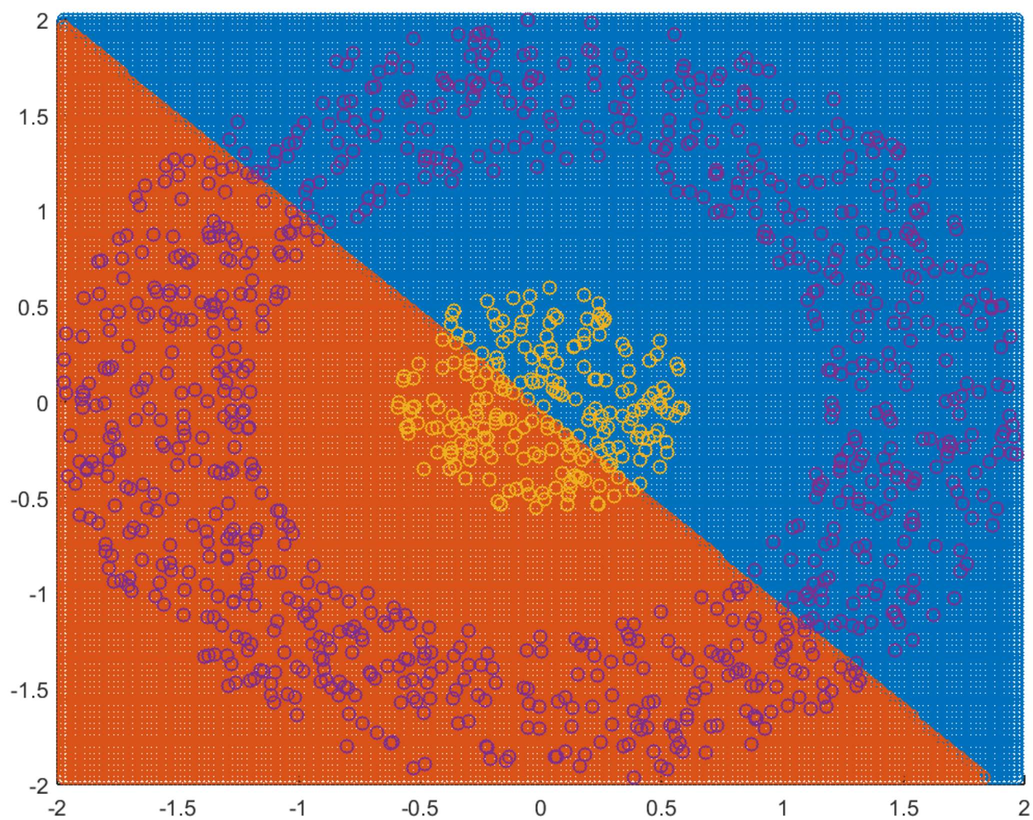
2.2.2.a) Covariance Matrix Same (σ^2I)

Classification Accuracy :- 51.2 %

Confusion Matrix

	Class 1	**Class 2**
Class 1	**47**	**28**
Class 2	**155**	**145**

Decision Region Plot for All Classes



Observation

As the covariance matrix was same for all, the decision surface had to be linear. But the data being in the form of a ring could not be separated by such a surface. Hence the classifier performed very badly with only about 50% accuracy

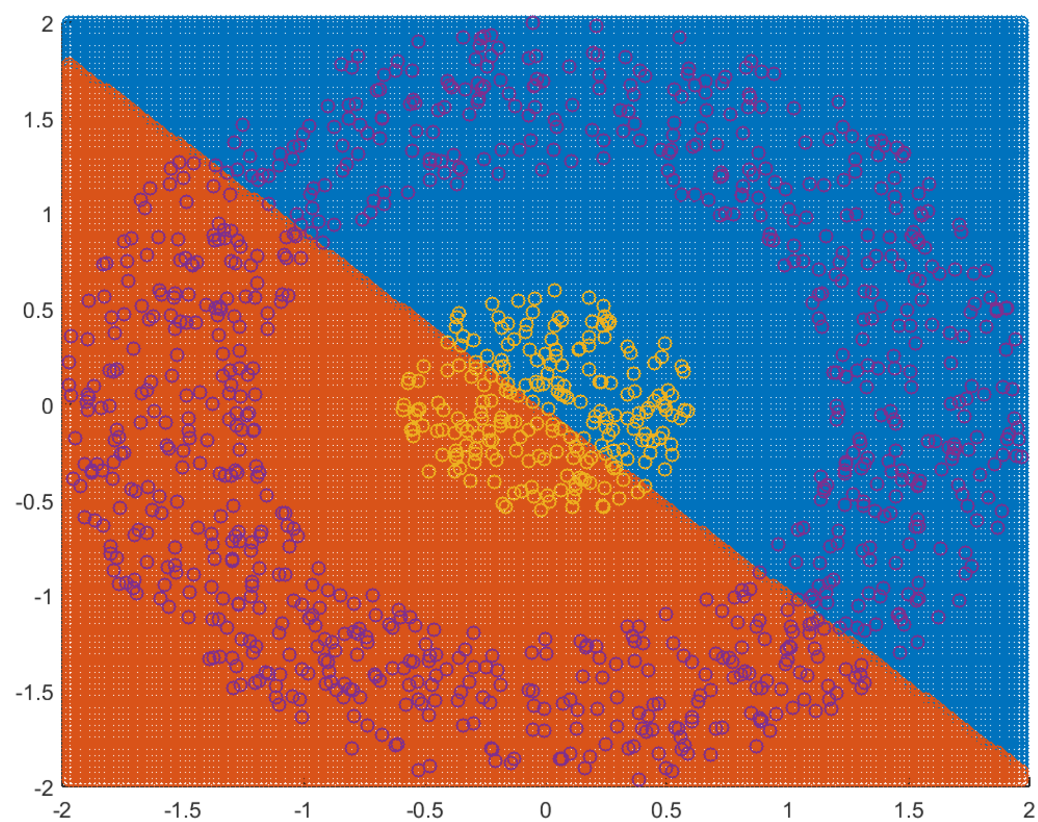
2.2.2.b) Covariance Matrix Same (C)

Classification Accuracy :- 50.133 %

Confusion Matrix

	Class 1	Class 2
Class 1	**46**	**29**
Class 2	**158**	**142**

Decision Region Plot for All Classes



Observation

As the covariance matrix was same for all, the decision surface had to be linear. But the data being in the form of a ring could not be separated by such a surface. Hence the classifier performed very badly with only about 50% accuracy

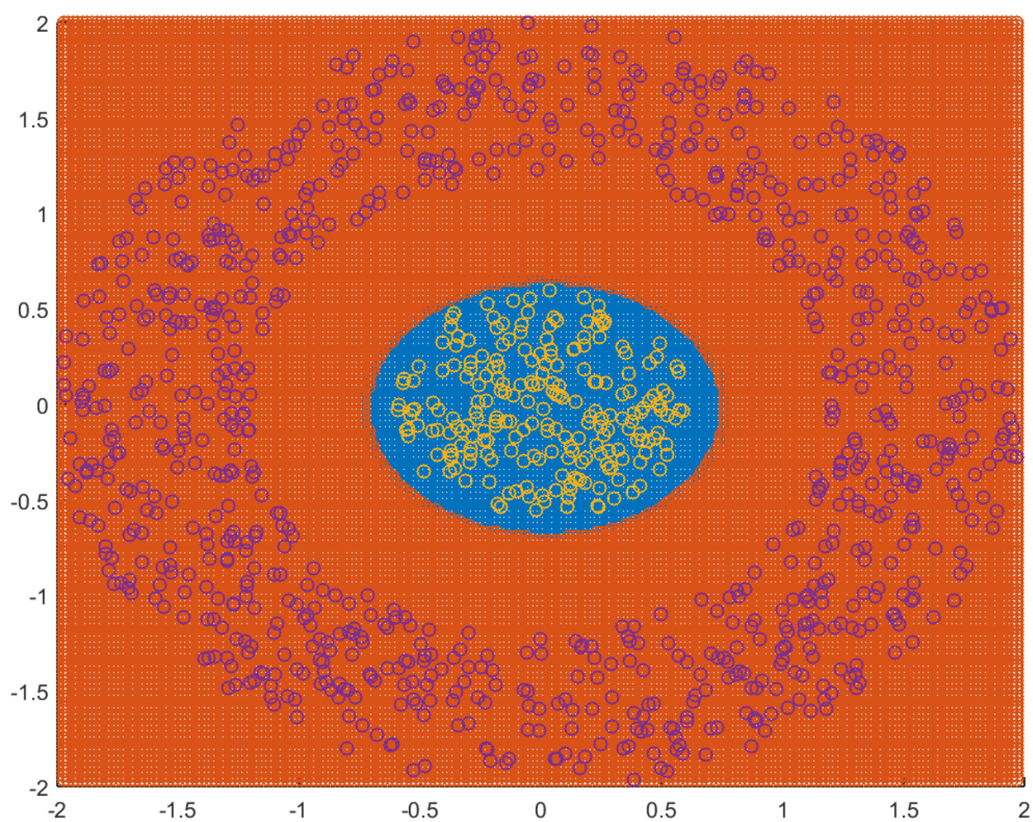
2.2.2.c) Covariance Matrix Different

Classification Accuracy :- 100 %

Confusion Matrix

	Class 1	**Class 2**
Class 1	**75**	**0**
Class 2	**0**	**300**

Decision Region Plot for All Classes



Observation

When we took different covariance matrix for both the classes, the classifier becomes 100% accurate due the elliptical decision surface it was able to produce.

2.3) Spiral Data Set

2.3.1) Bayes Classifier

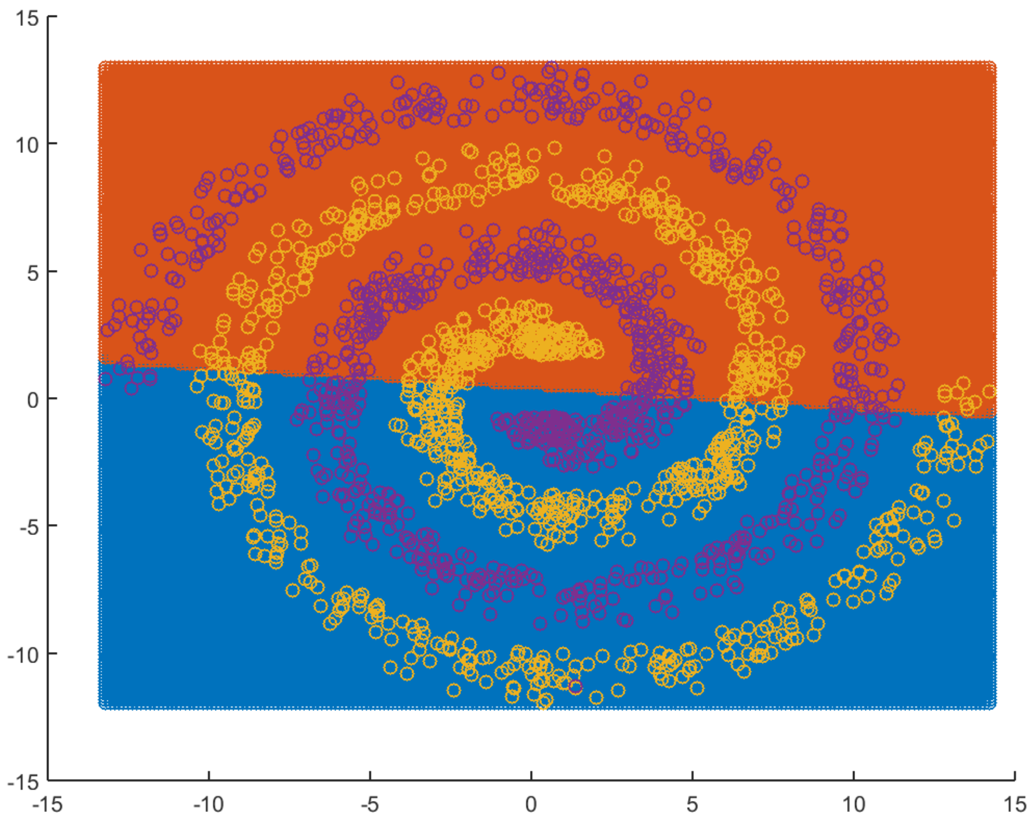
2.3.1.a.i) Average of Covariance Matrices of All Classes

Classification Accuracy :- 53.681 %

Confusion Matrix

	Class 1	Class 2
Class 1	175	151
Class 2	151	175

Decision Region Plot for All Classes



Observation

The data cannot be separated by any linear surface and hence the accuracy is pretty bad, merely crossing 50%.

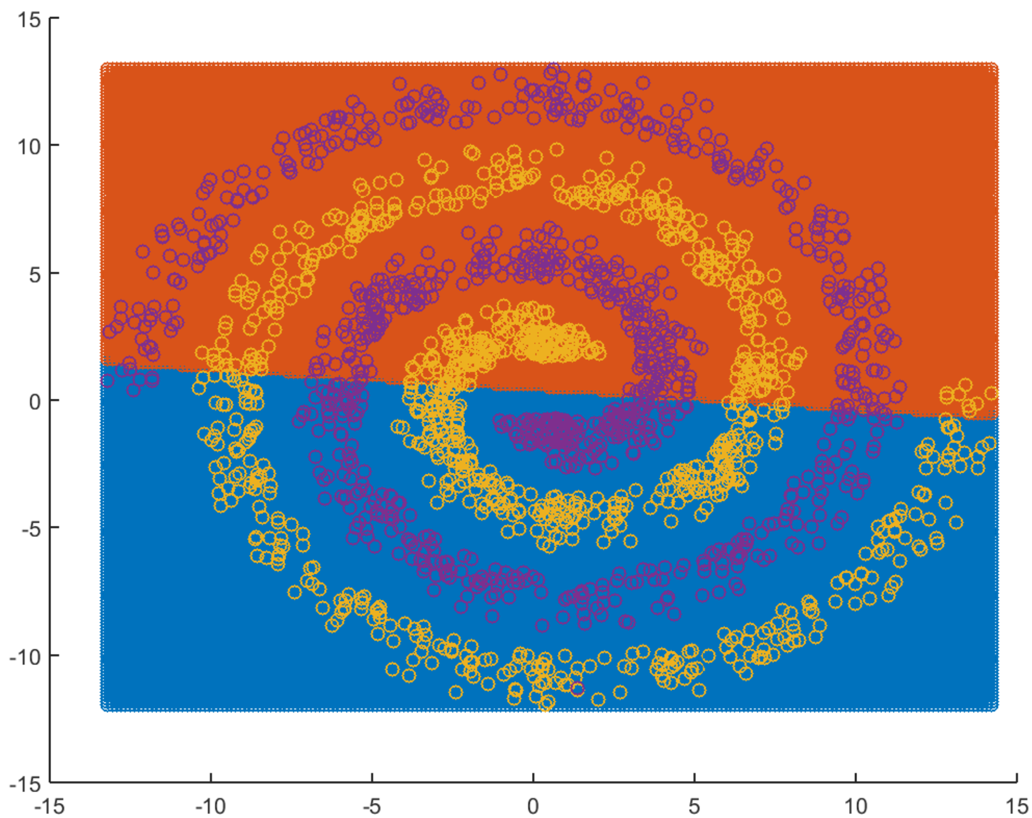
2.3.1.a.ii) Covariance Matrix from all Training Data Combined

Classification Accuracy :- 53.681 %

Confusion Matrix

	Class 1	**Class 2**
Class 1	**175**	**151**
Class 2	**151**	**175**

Decision Region Plot for All Classes



Observation

The data cannot be separated by any linear surface and hence the accuracy is pretty bad, merely crossing 50%.

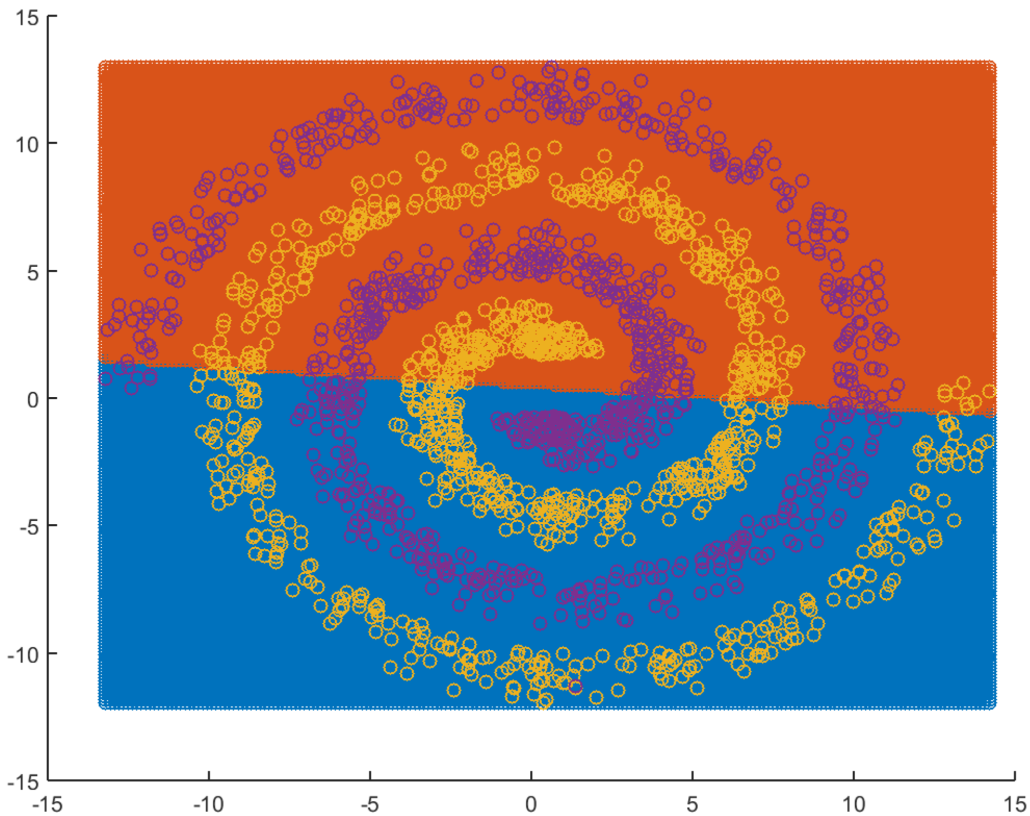
2.3.1.b) Covariance Matrix Different

Classification Accuracy :- 53.681 %

Confusion Matrix

	Class 1	Class 2
Class 1	**175**	**151**
Class 2	**151**	**175**

Decision Region Plot for All Classes



Observation

The decision surface comes out to be hyperbolic in nature. Data cannot be separated by a hyperbolic surface and hence the accuracy is pretty bad, merely crossing 50%.

2.3.2) Naïve Bayes Classifier

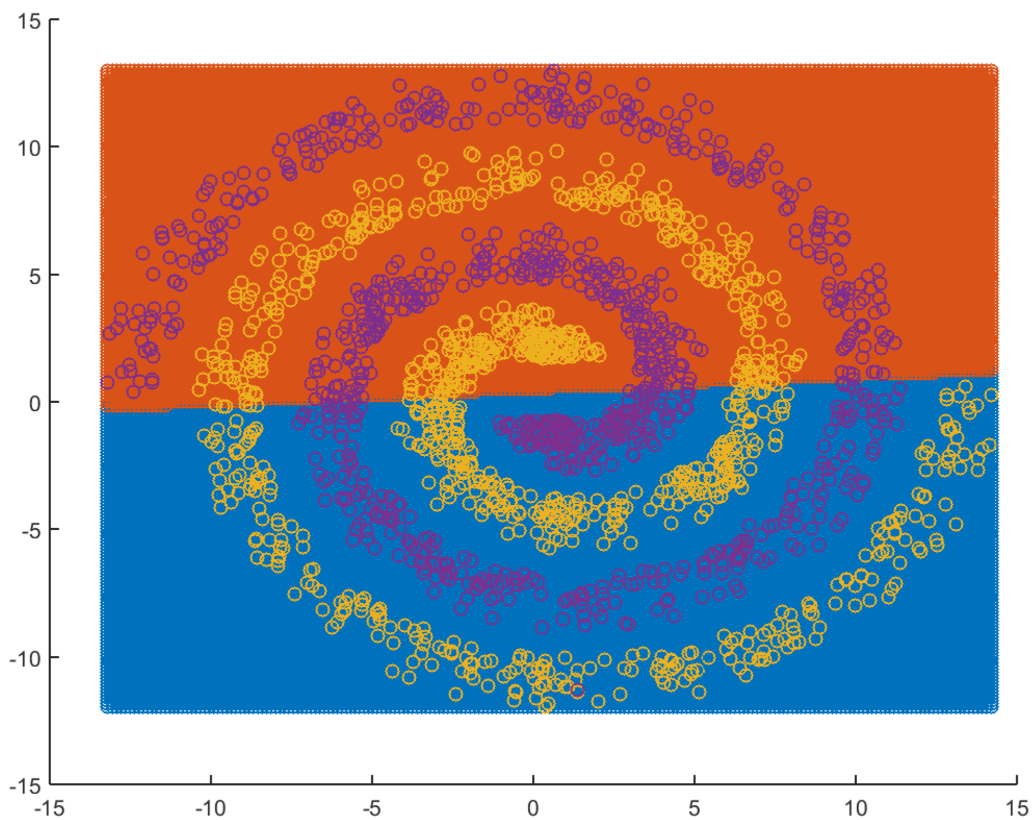
2.3.2.a) Covariance Matrix Same (σ^2I)

Classification Accuracy :- 54.141 %

Confusion Matrix

	Class 1	Class 2
Class 1	176	150
Class 2	149	177

Decision Region Plot for All Classes



Observation

The data cannot be separated by any linear surface and hence the accuracy is pretty bad, merely crossing 50%. The slope of the decision surface of this classifier is different as compared to the Bayesian case as the covariance matrix is of the form $\sigma^2 I$.

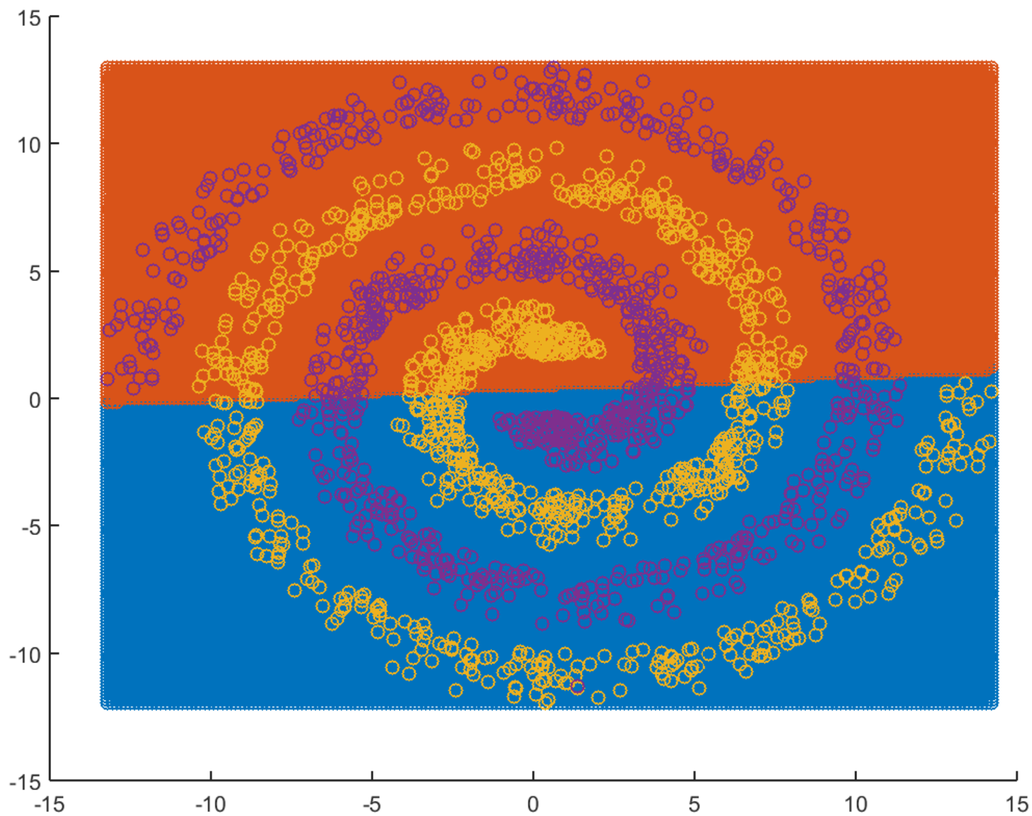
2.3.2.b) Covariance Matrix Same (C)

Classification Accuracy :- 54.294 %

Confusion Matrix

	Class 1	**Class 2**
Class 1	**177**	**149**
Class 2	**149**	**177**

Decision Region Plot for All Classes



Observation

The decision surface comes out to be hyperbolic in nature. Data cannot be separated by a hyperbolic surface and hence the accuracy is pretty bad, merely crossing 50%.

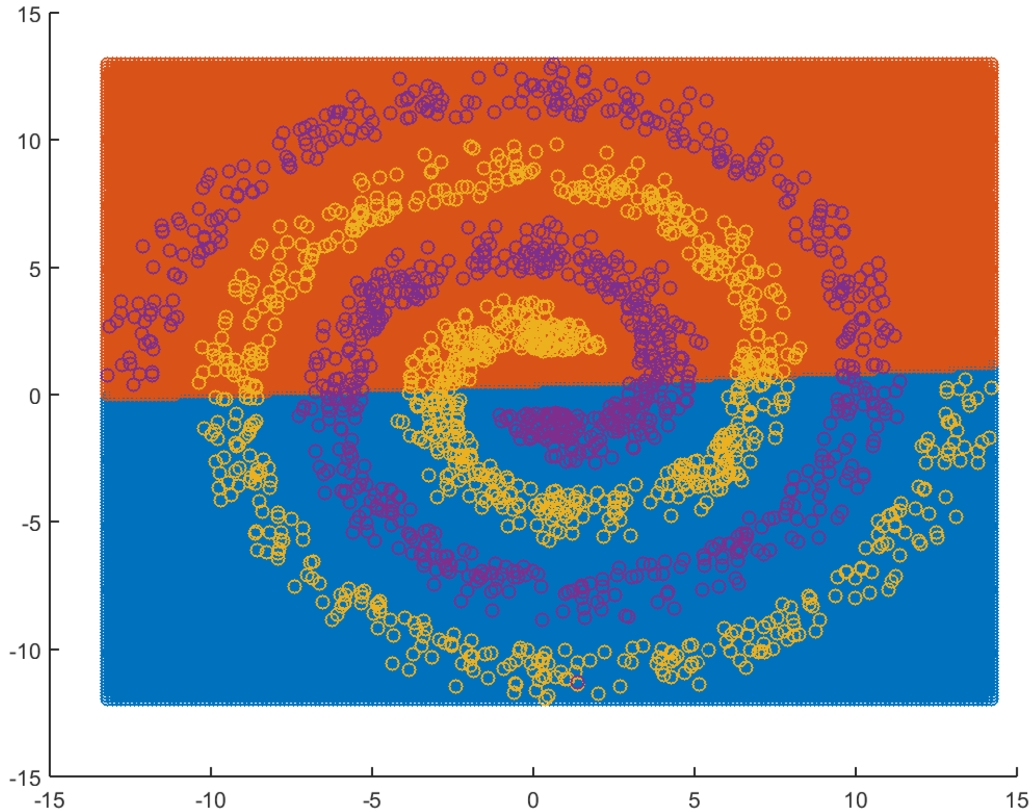
2.3.2.c) Covariance Matrix Different

Classification Accuracy :- 54.294 %

Confusion Matrix

	Class 1	Class 2
Class 1	177	149
Class 2	149	177

Decision Region Plot for All Classes



Observation

The decision surface comes out to be hyperbolic in nature. Data cannot be separated by a hyperbolic surface and hence the accuracy is pretty bad, merely crossing 50%.

3) Overlapping

3.1) Bayes Classifier

3.1.a) Covariance Matrix Same

3.1.a.i) Average of Covariance Matrices of All Classes

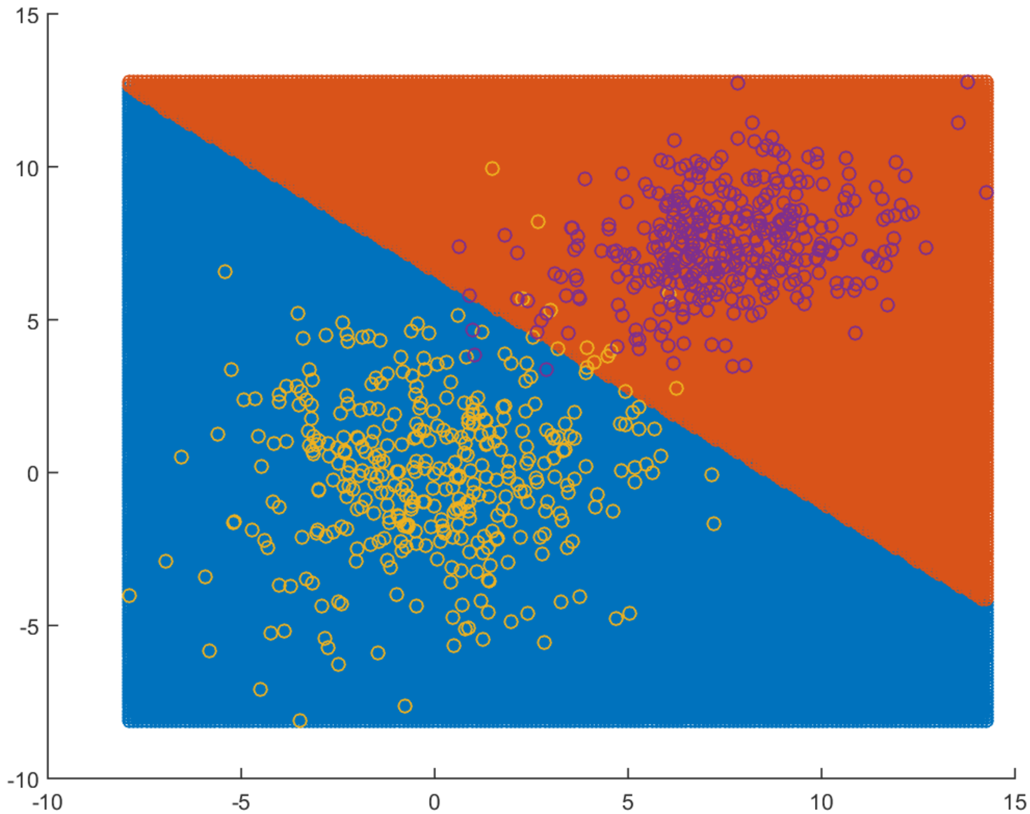
Classification Accuracy:- 97.867 %

Confusion Matrix

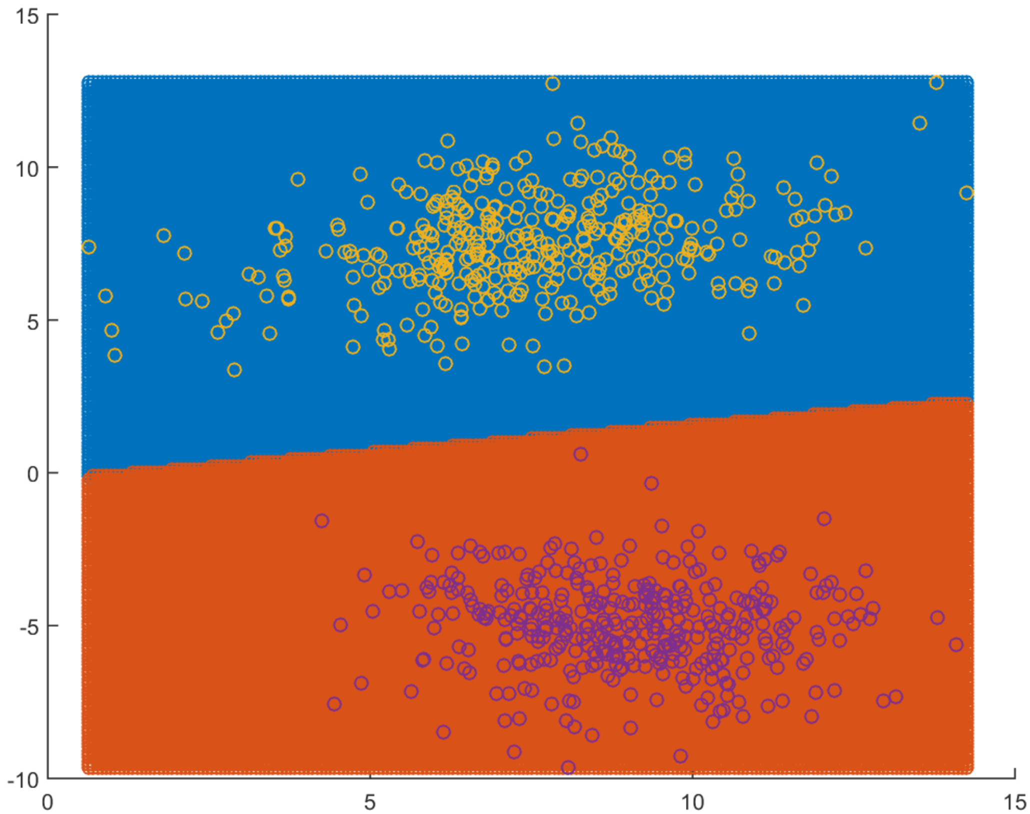
	Class 1	Class 2	Class 3
Class 1	119	3	3
Class 2	1	124	0
Class 3	1	0	124

Decision Region Plot for Pair

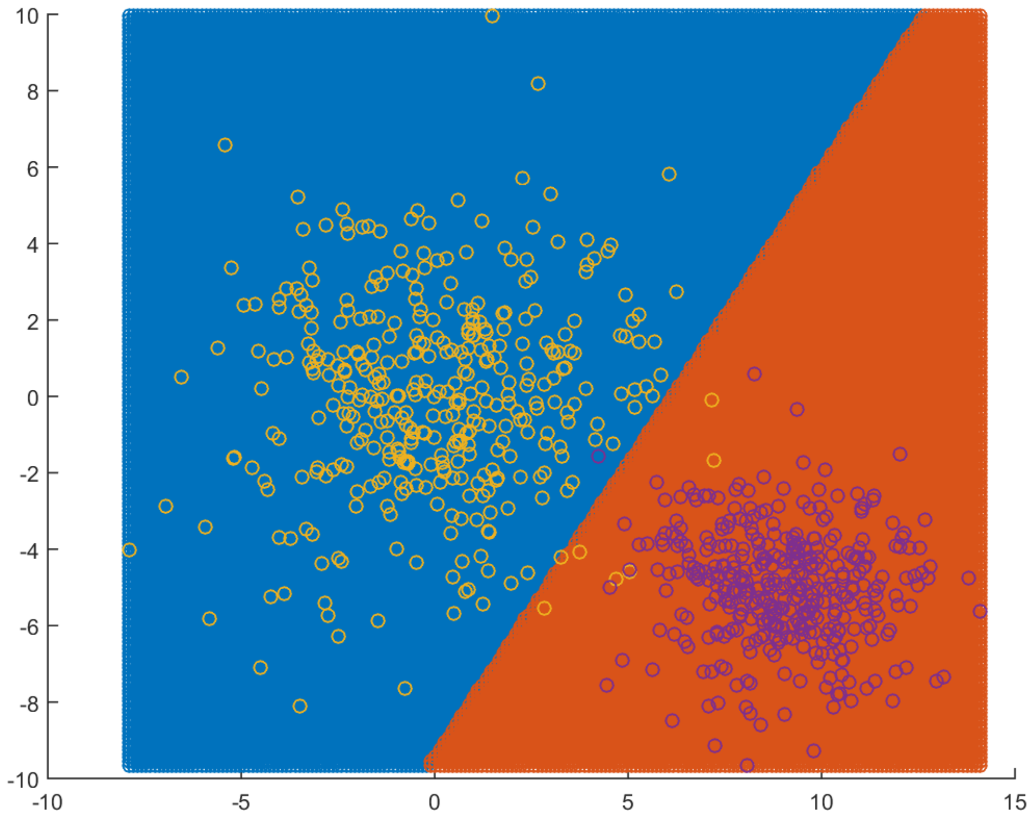
Class1, Class2



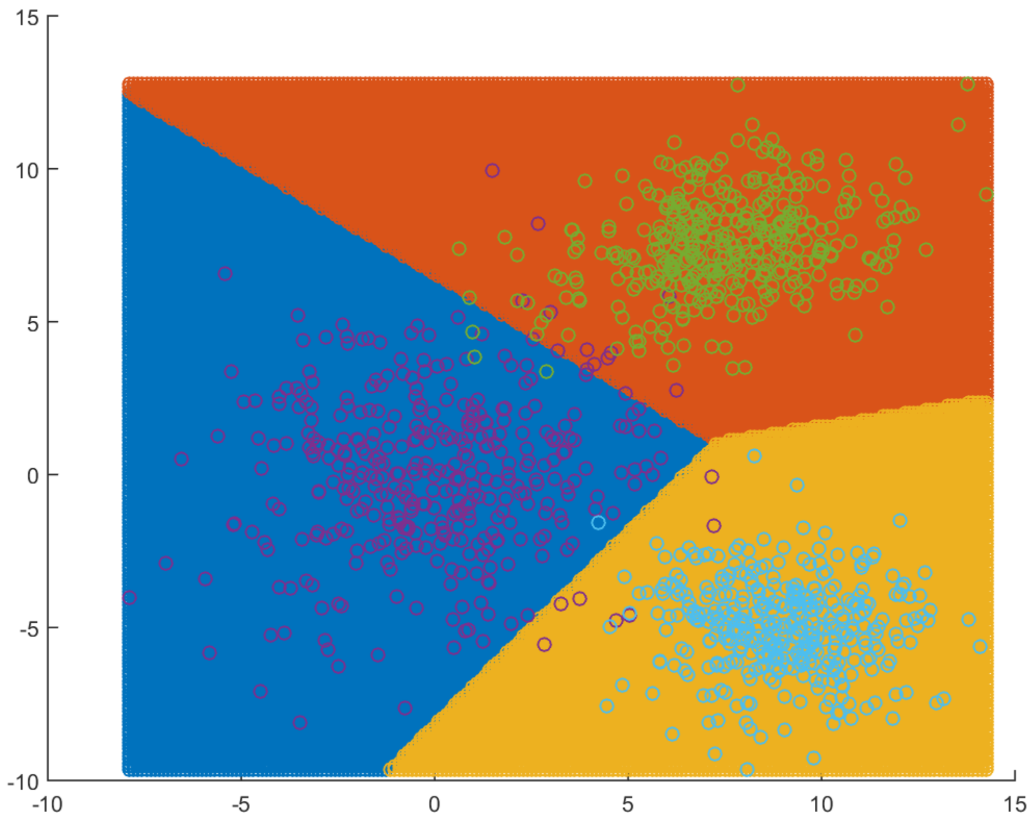
Class2, Class3



Class1, Class3



Decision Region Plot for All Classes



Observation

As the covariance matrix was same for all the classes therefore the decision surface is linear as can be observed from the graph. The decision surface lines passes through the average of the means of the classes. As the decision surface is linear therefore the accuracy is around 97.8 %.

3.1.a.ii) Covariance Matrix from all Training Data Combined

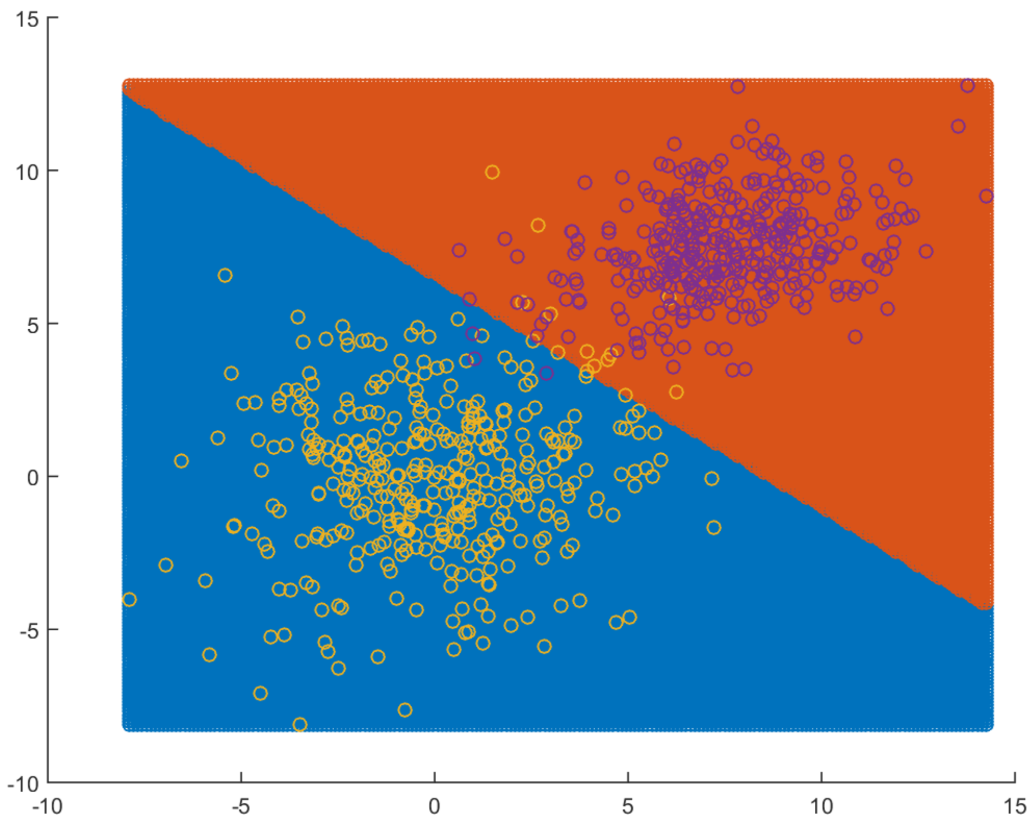
Classification Accuracy :- 98.133 %

Confusion Matrix

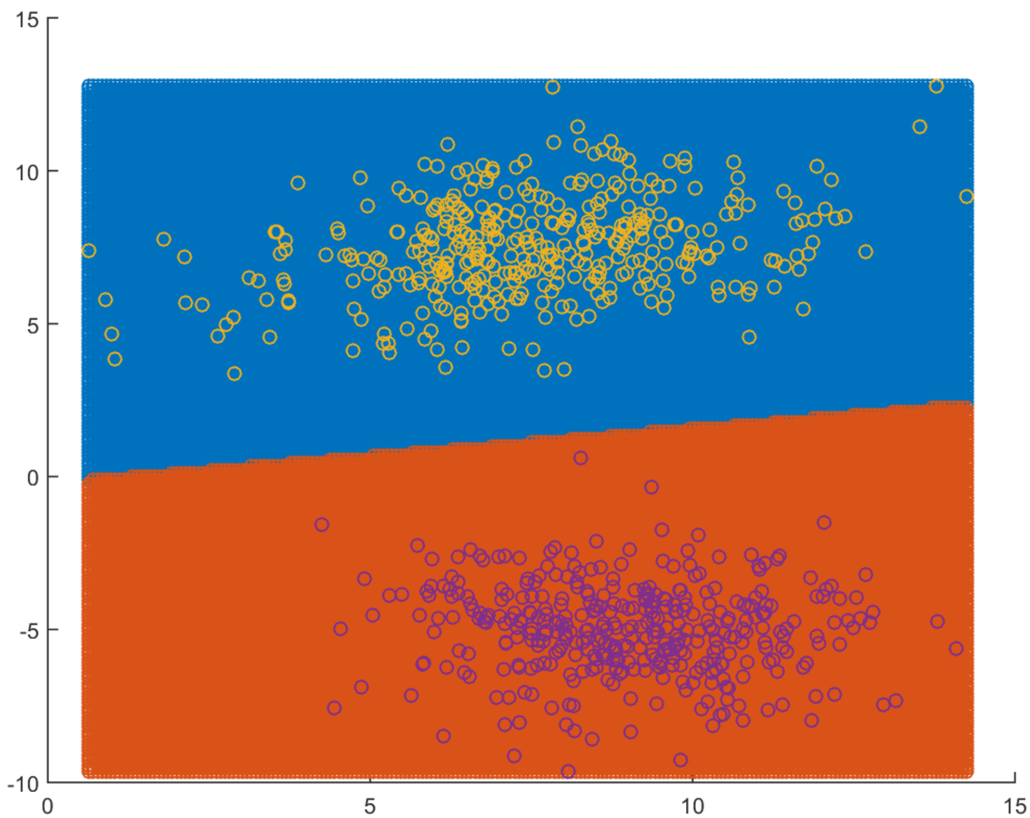
	Class 1	Class 2	Class 3
Class 1	120	3	2
Class 2	1	124	0
Class 3	1	0	124

Decision Region Plot for Pair

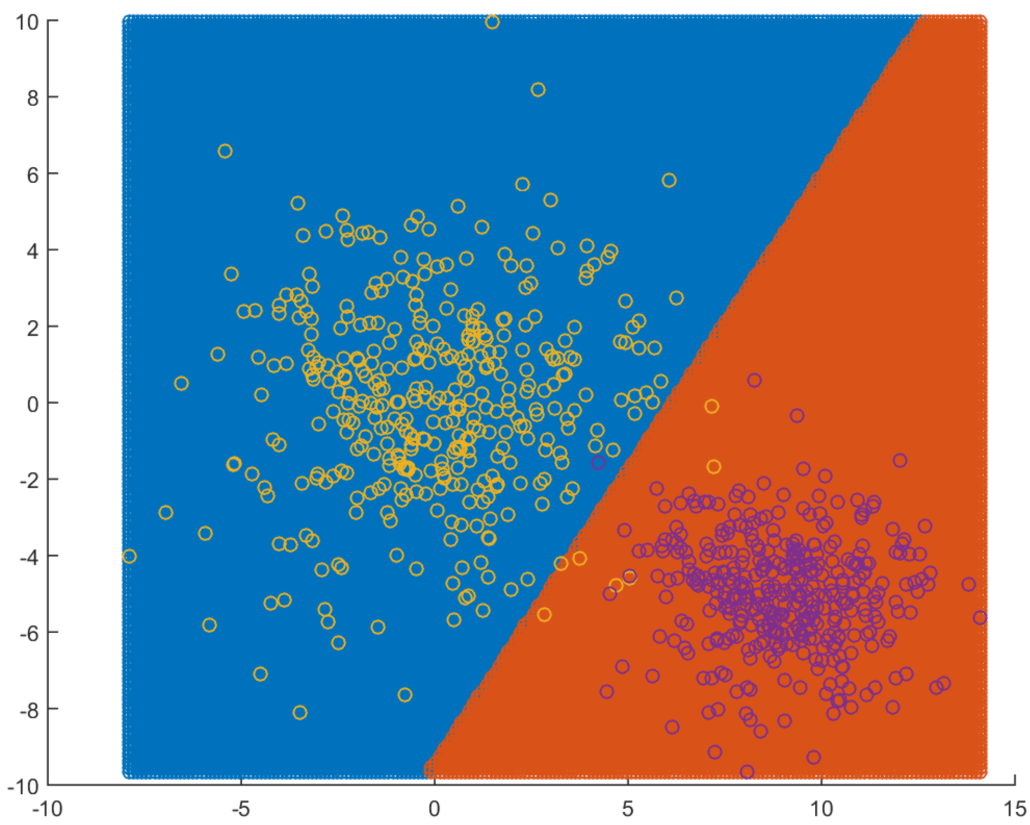
Class1, Class2



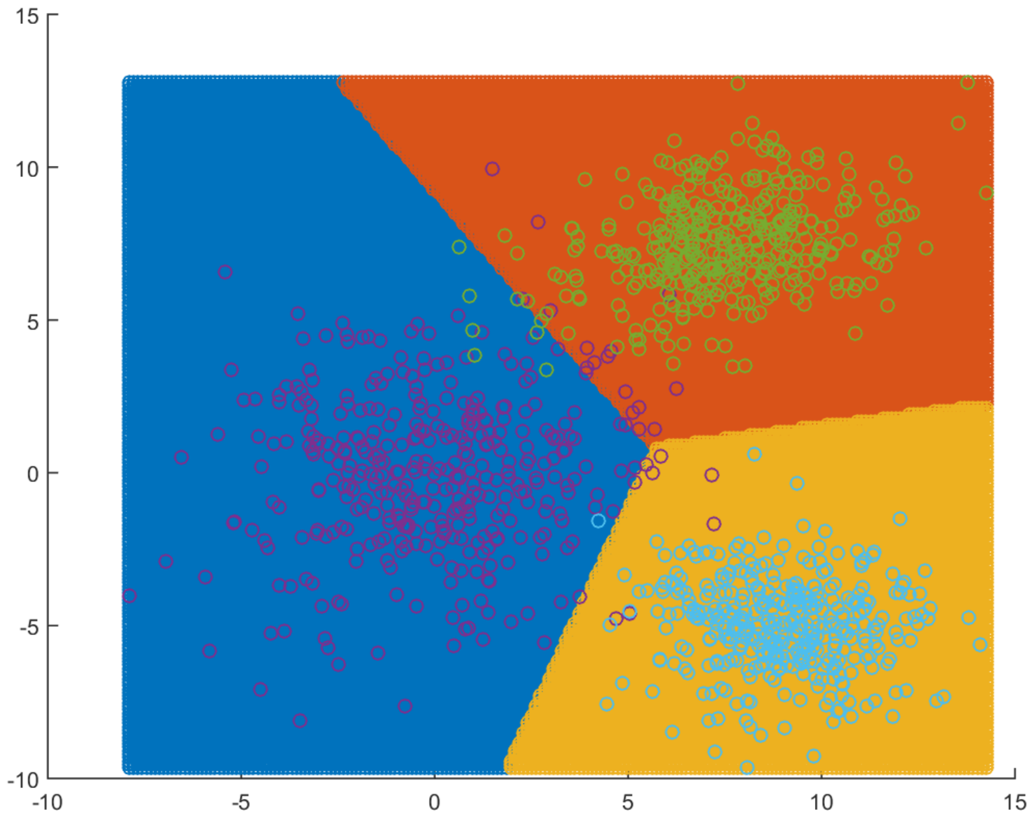
Class2, Class3



Class1, Class3



Decision Region Plot for All Classes



Observation

As the covariance matrix was same for all the classes therefore the decision surface is linear as can be observed from the graph. The decision surface lines passes through the average of the means of the classes. As the decision surface is linear therefore the accuracy is around 98.1 %.

3.1.b) Covariance Matrix Different

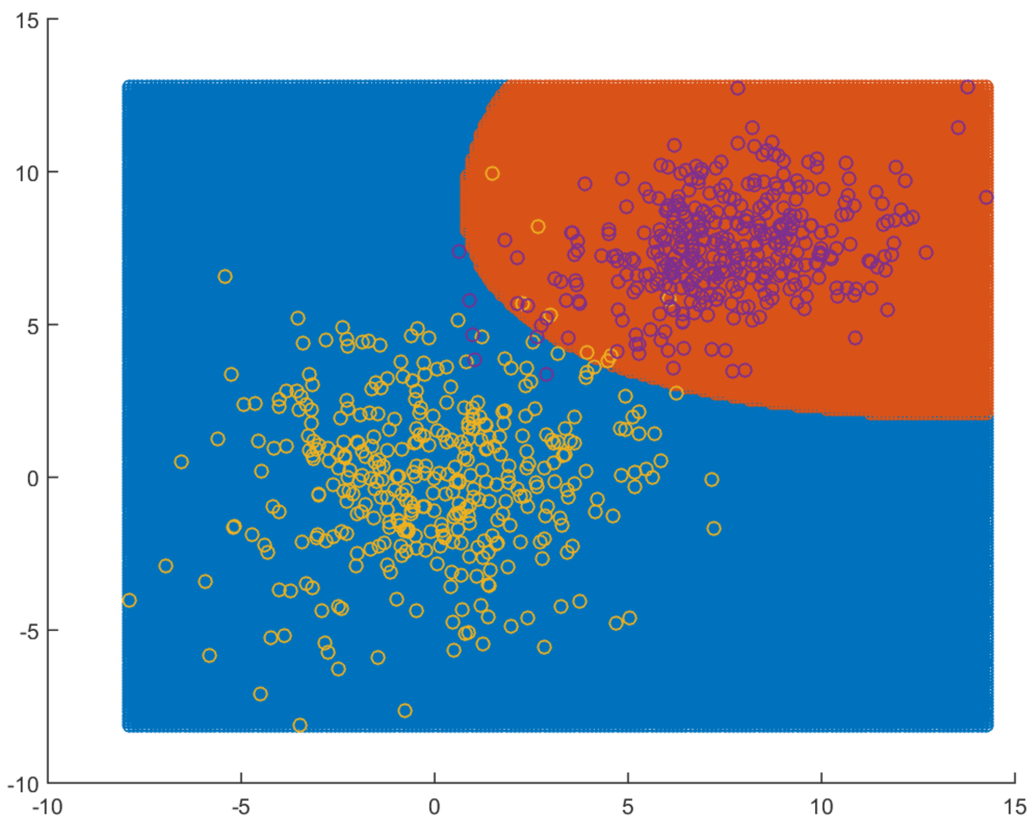
Classification Accuracy :- 98.4 %

Confusion Matrix

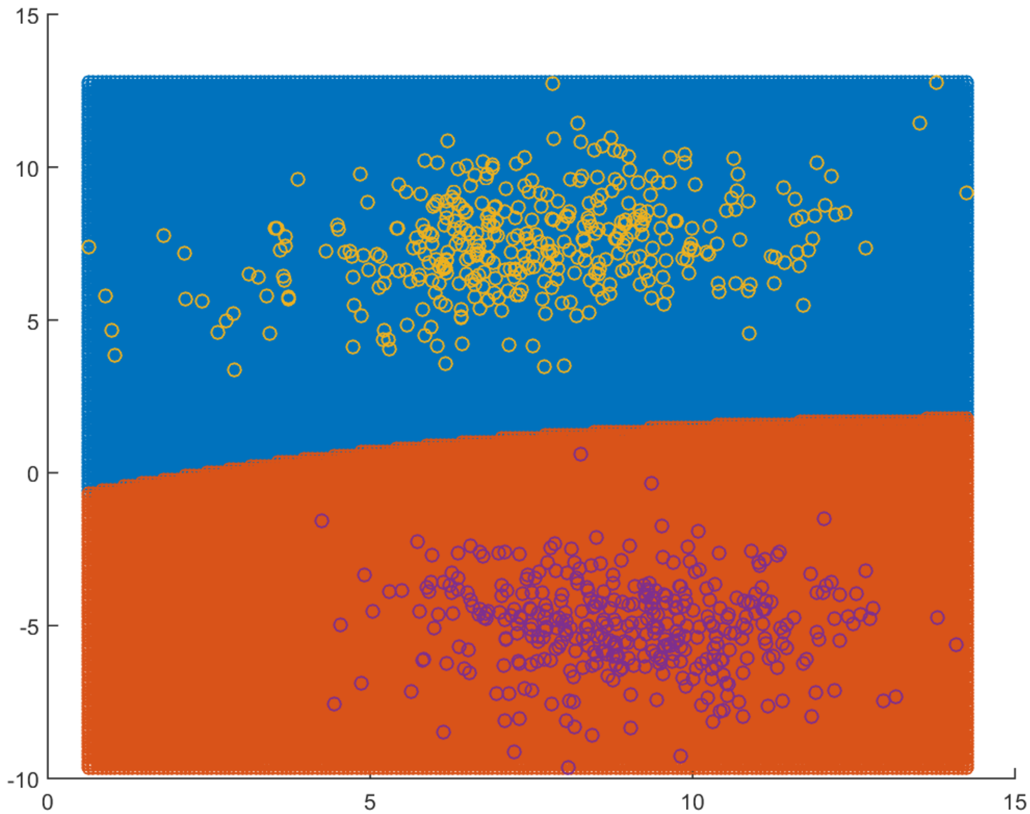
	Class 1	Class 2	Class 3
Class 1	122	1	2
Class 2	2	123	0
Class 3	1	0	124

Decision Region Plot for Pair

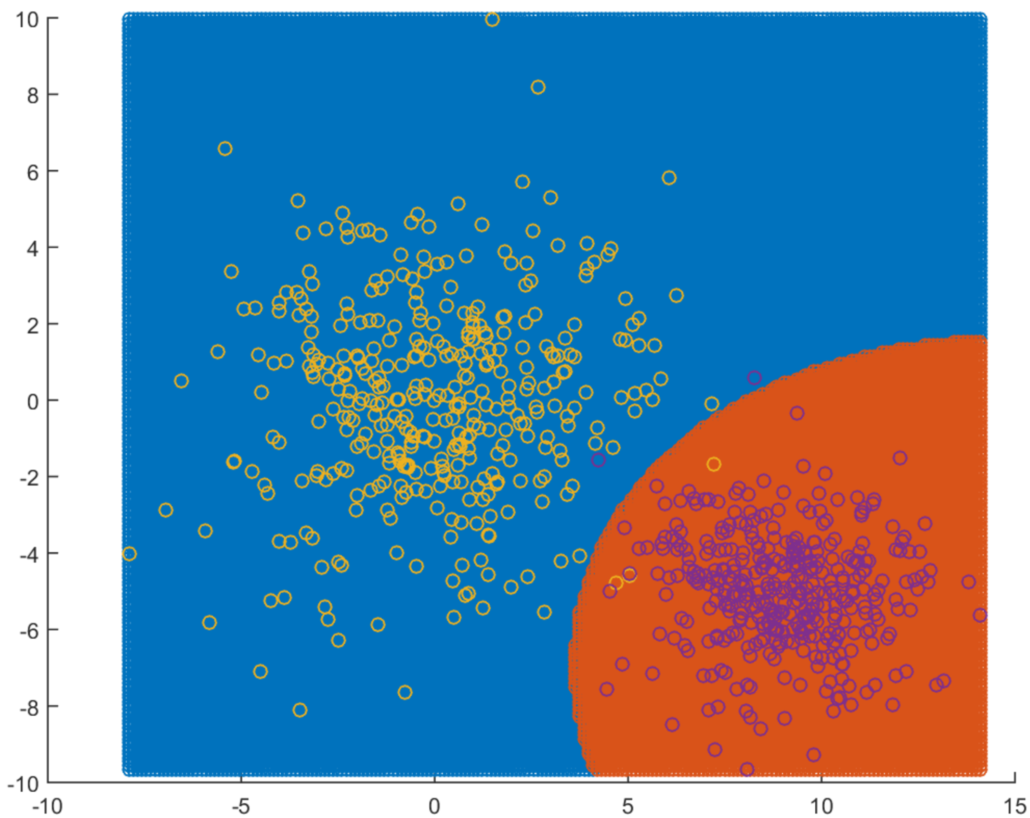
Class1, Class2



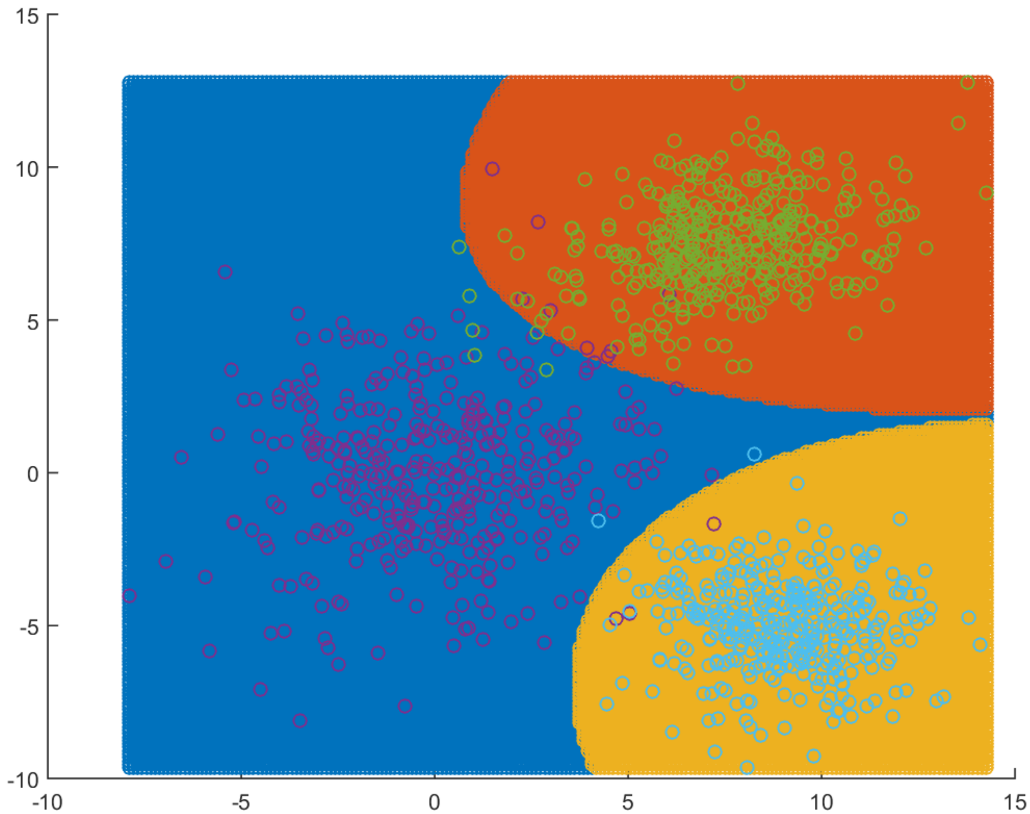
Class2, Class3



Class1, Class3



Decision Region Plot for All Classes



Observation

As the covariance matrices are different for different classes therefore the decision surface is a 2^{nd} order curve. Here, the yellow and the red classes are enclosed in two ellipses and the third class covers the rest of the area. Here, the accuracy is 98.4 % which is slightly better than the case of linear decision surface.

3.2) Naïve Bayes Classifier

3.2.a) Covariance Matrix Same (σ^2I)

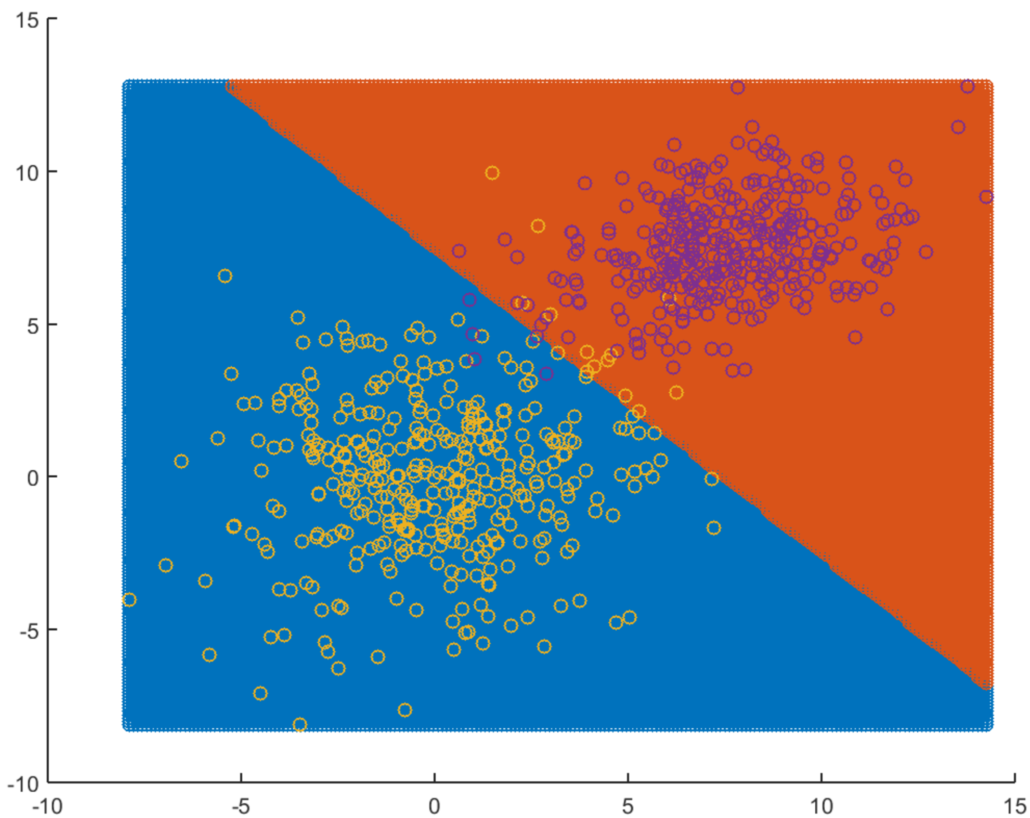
Classification Accuracy : - 98.133 %

Confusion Matrix

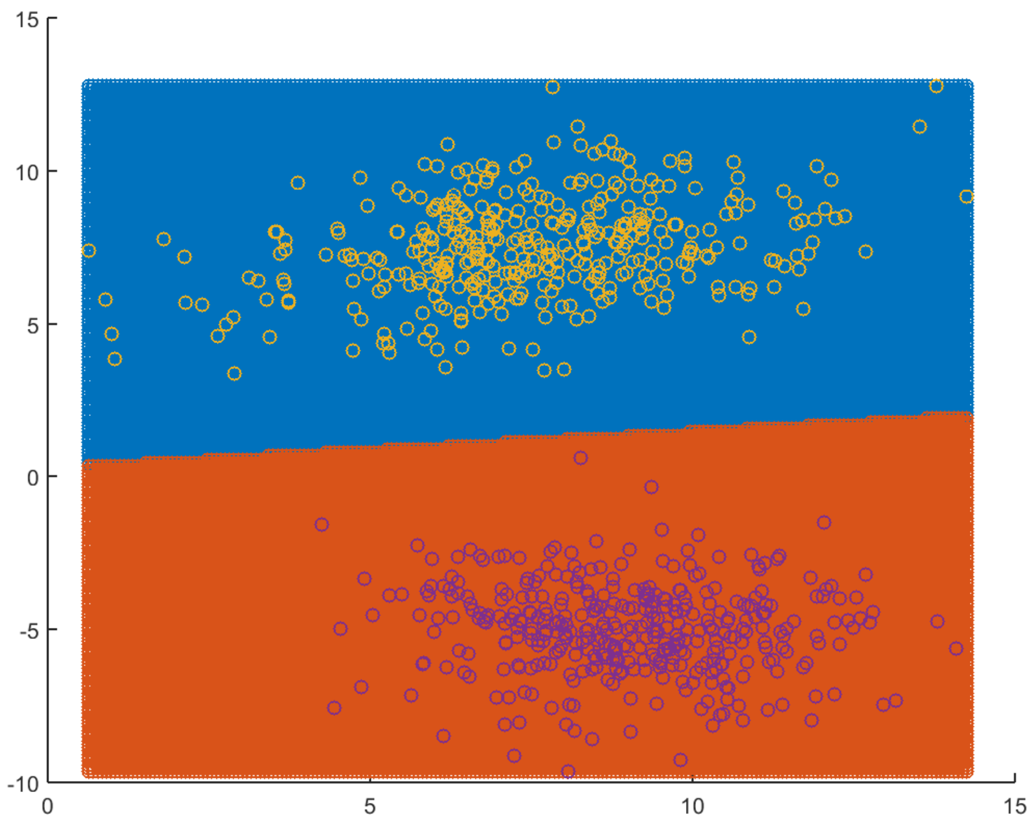
	Class 1	**Class 2**	**Class 3**
Class 1	121	2	2
Class 2	2	123	0
Class 3	1	0	124

Decision Region Plot for Pair

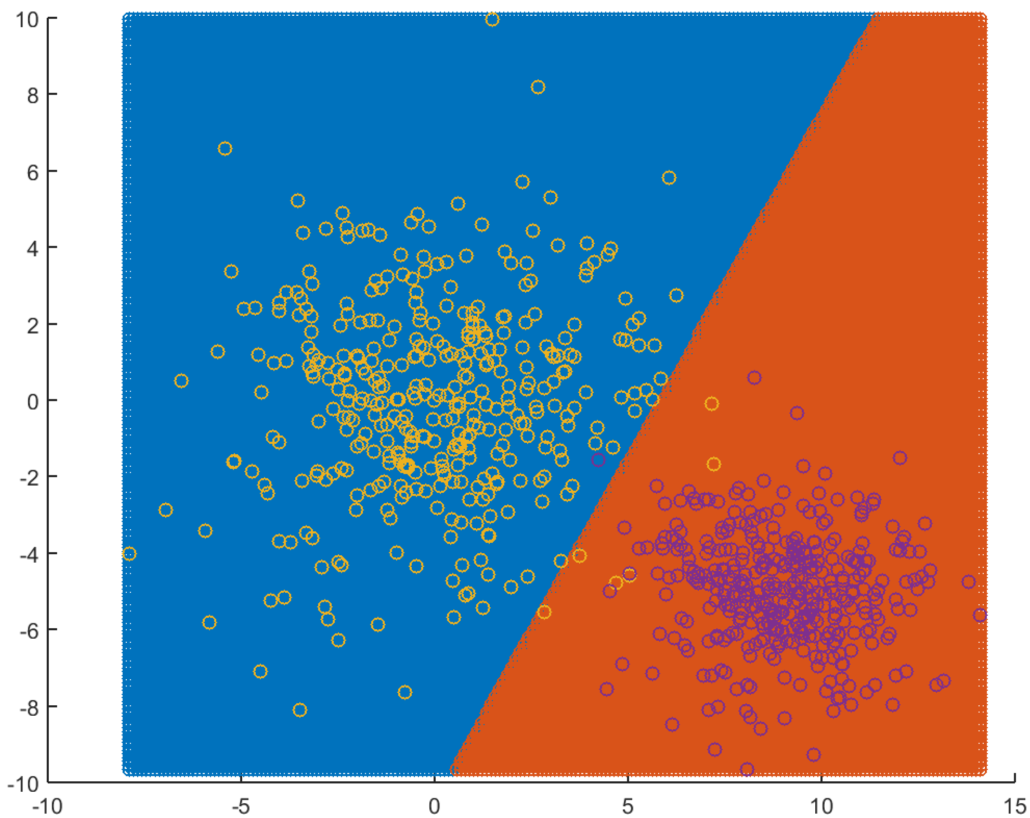
Class1, Class2



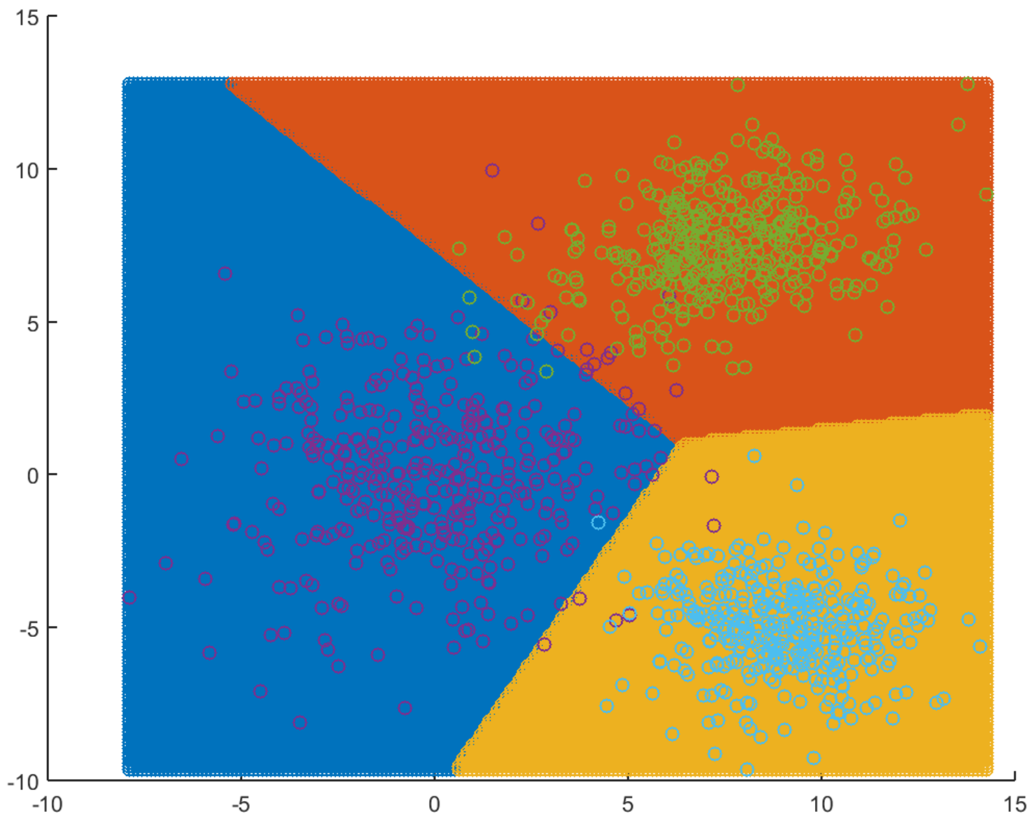
Class2, Class3



Class1, Class3



Decision Region Plot for All Classes



Observation

As the covariance matrix was same for all the classes therefore the decision surface is linear as can be observed from the graph. The decision surface lines passes through the average of the means of the classes. As the decision surface is linear therefore the accuracy is around 98.1 %. The accuracy is as for the Bayes.

3.2.b) Covariance Matrix Same (C)

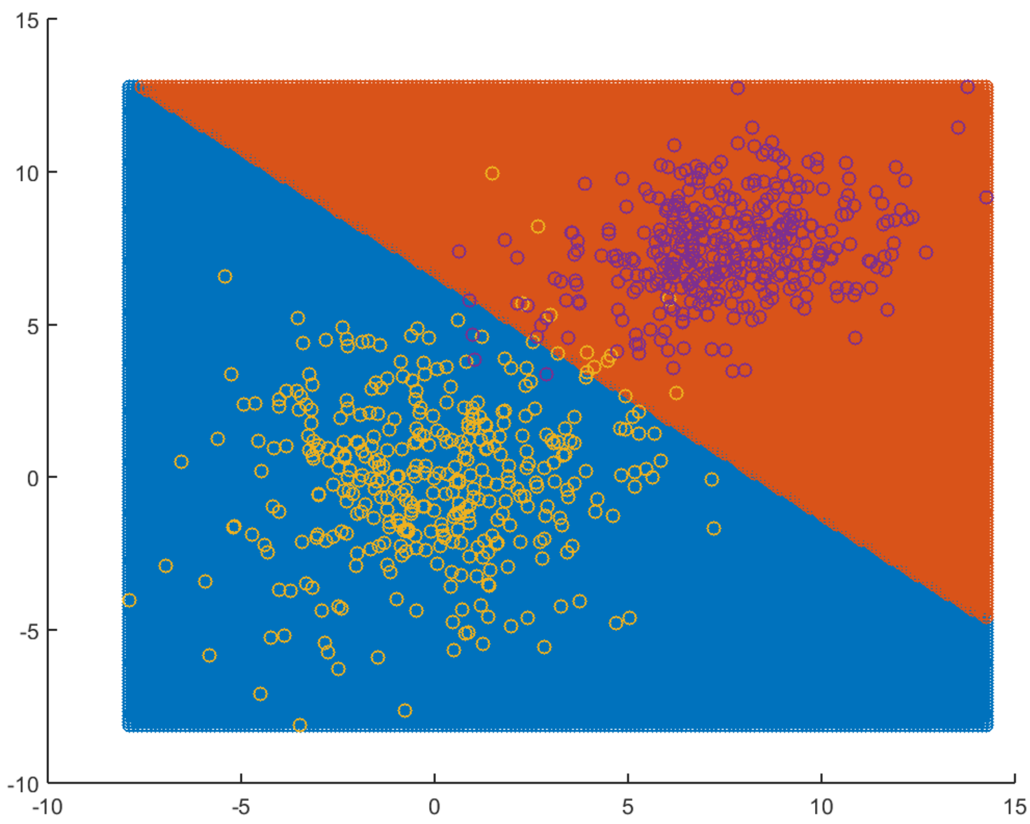
Classification Accuracy :- 98.133 %

Confusion Matrix

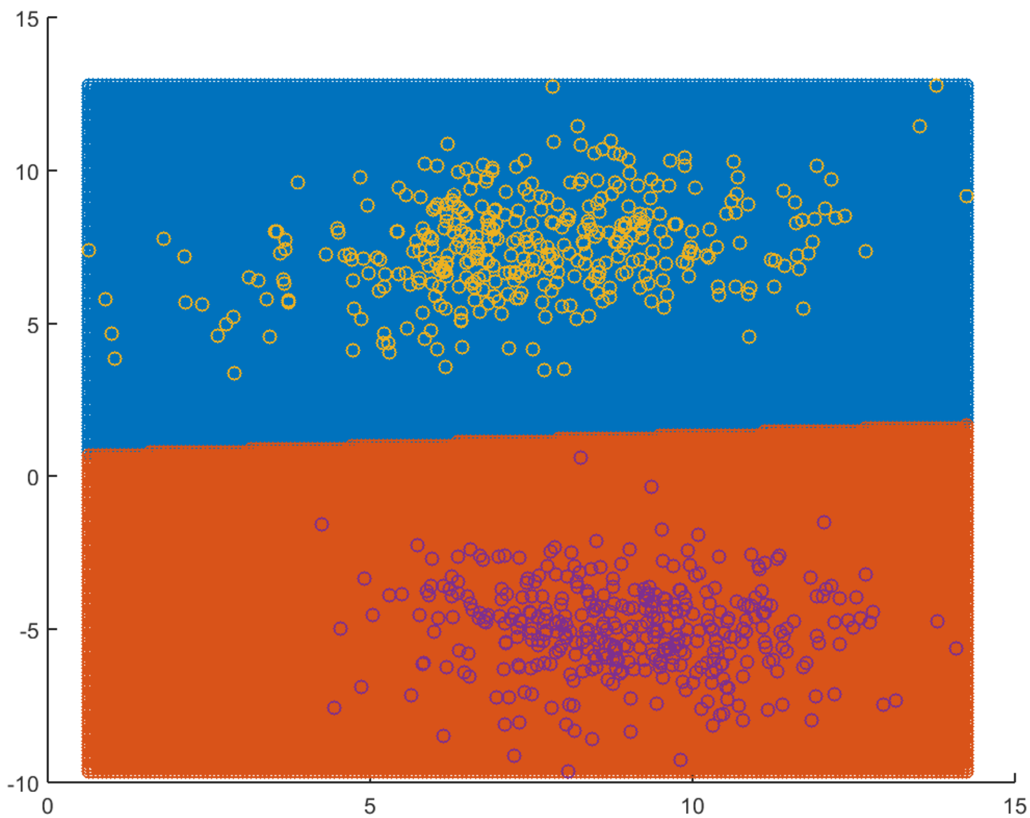
	Class 1	**Class 2**	**Class 3**
Class 1	120	3	2
Class 2	1	124	0
Class 3	1	0	124

Decision Region Plot for Pair

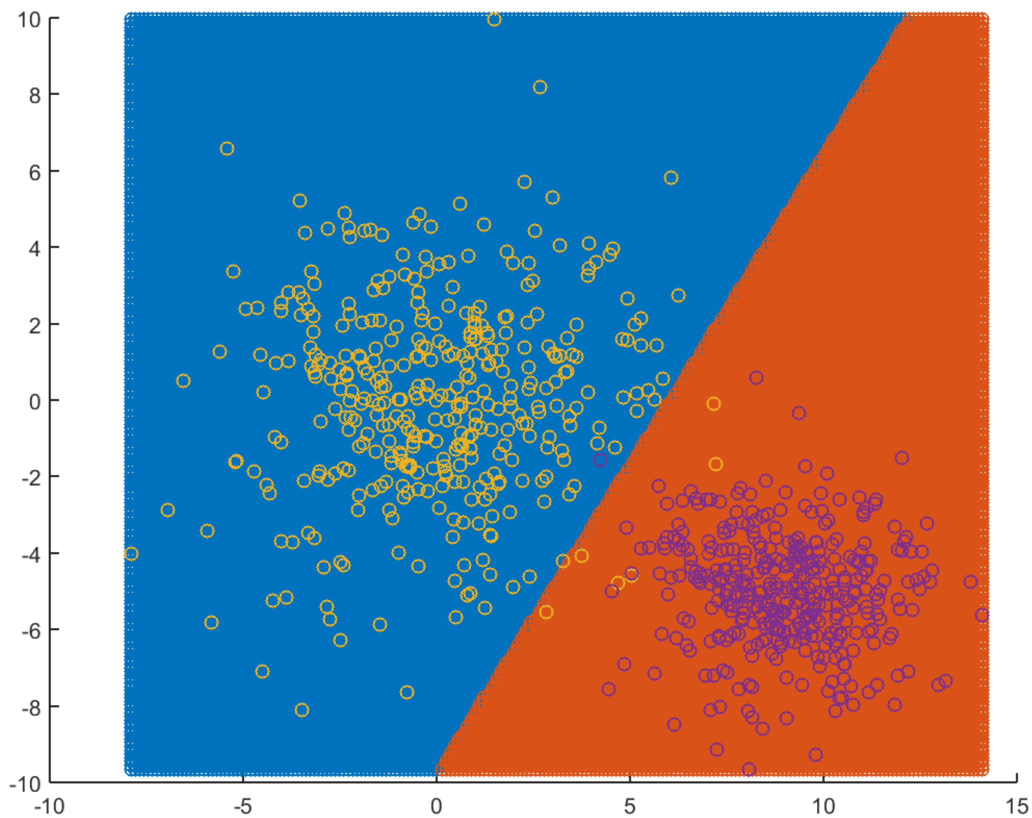
Class1, Class2



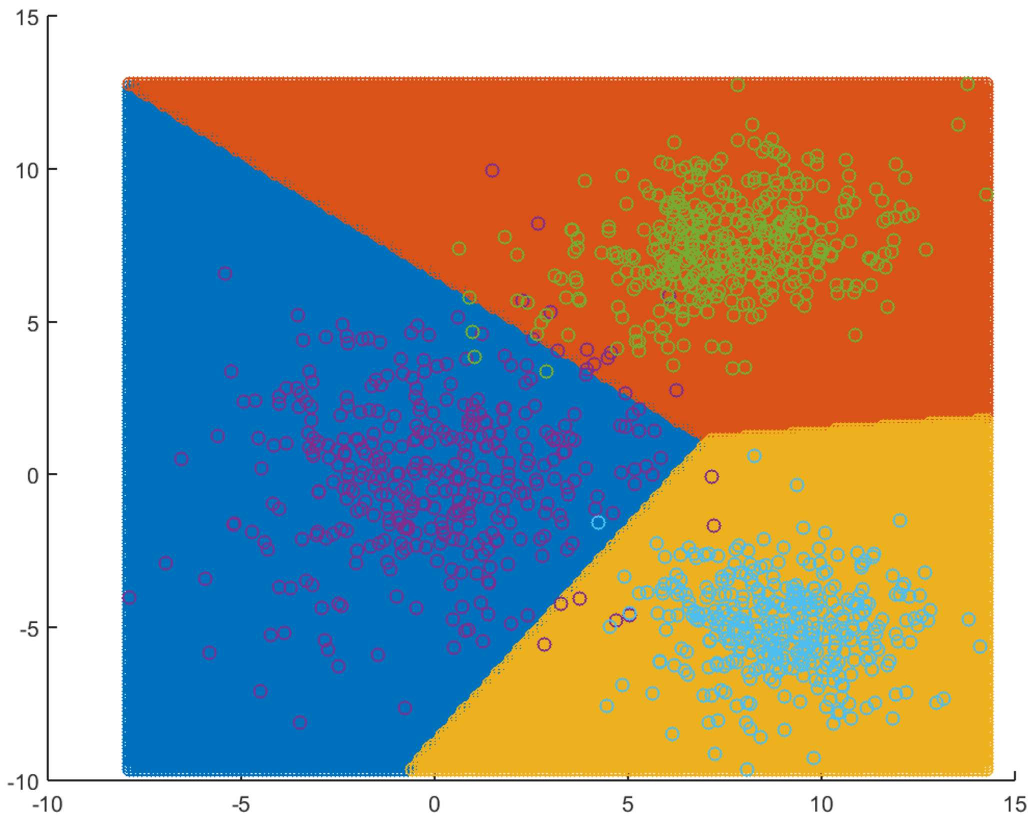
Class2, Class3



Class1, Class3



Decision Region Plot for All Classes



Observation

As the covariance matrix was same for all the classes therefore the decision surface is linear as can be observed from the graph. The decision surface lines passes through the average of the means of the classes. As the decision surface is linear therefore the accuracy is around 98.1 %. The accuracy is same as the Bayes classifier.

3.2.c) Covariance Matrix Different

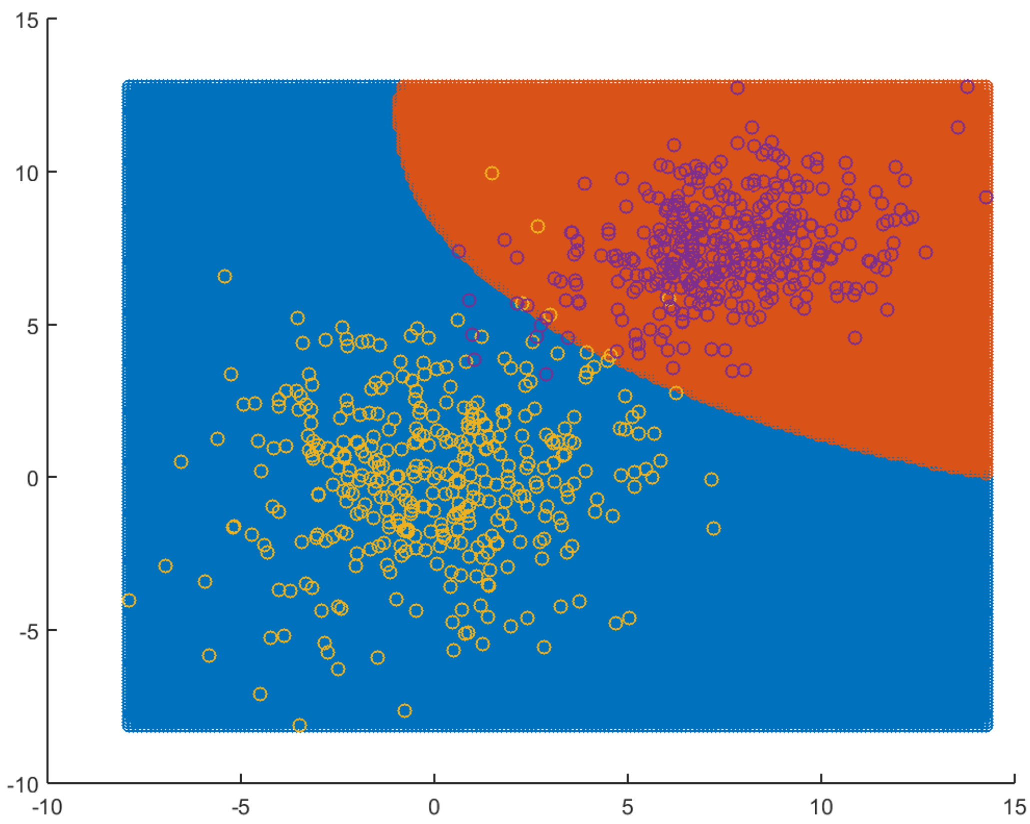
Classification Accuracy :- 98.933 %

Confusion Matrix

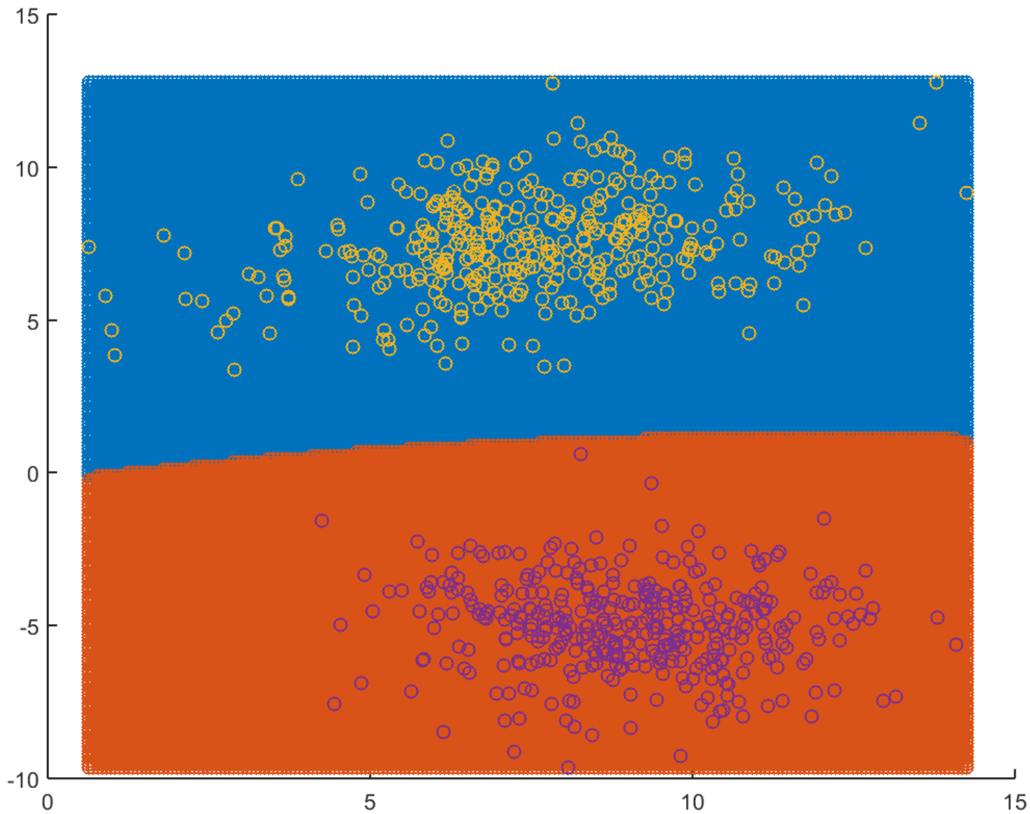
	Class 1	Class 2	Class 3
Class 1	124	1	0
Class 2	2	123	0
Class 3	1	0	124

Decision Region Plot for Pair

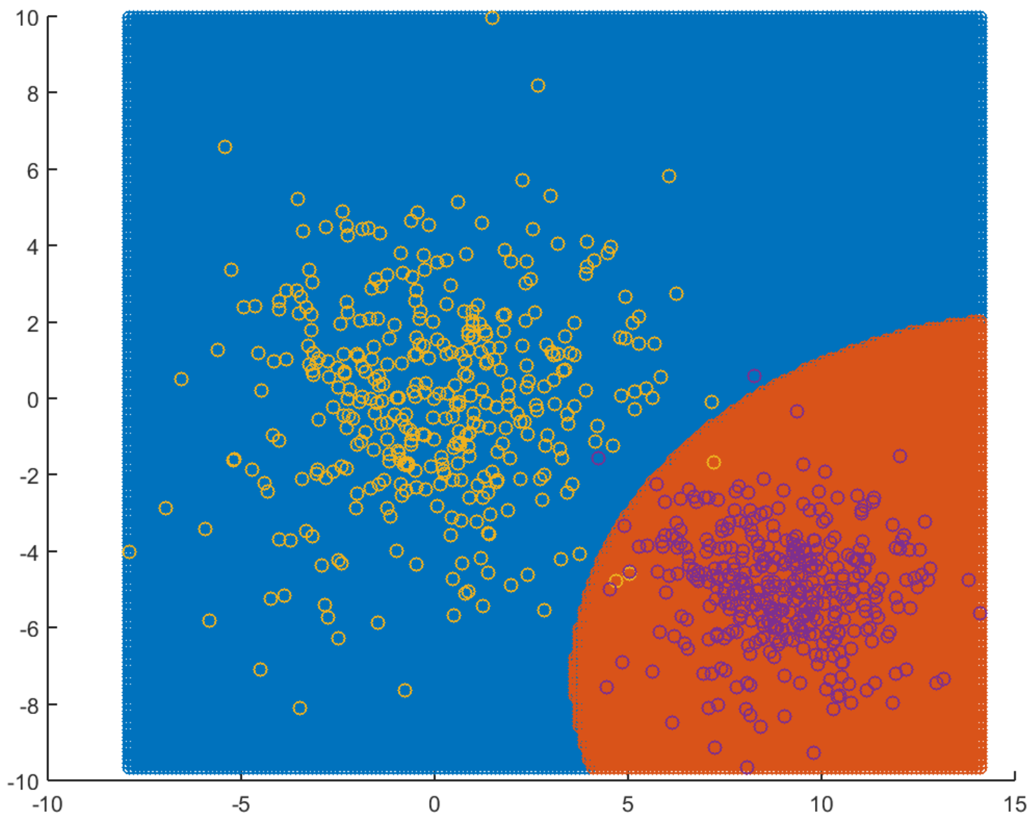
Class1, Class2



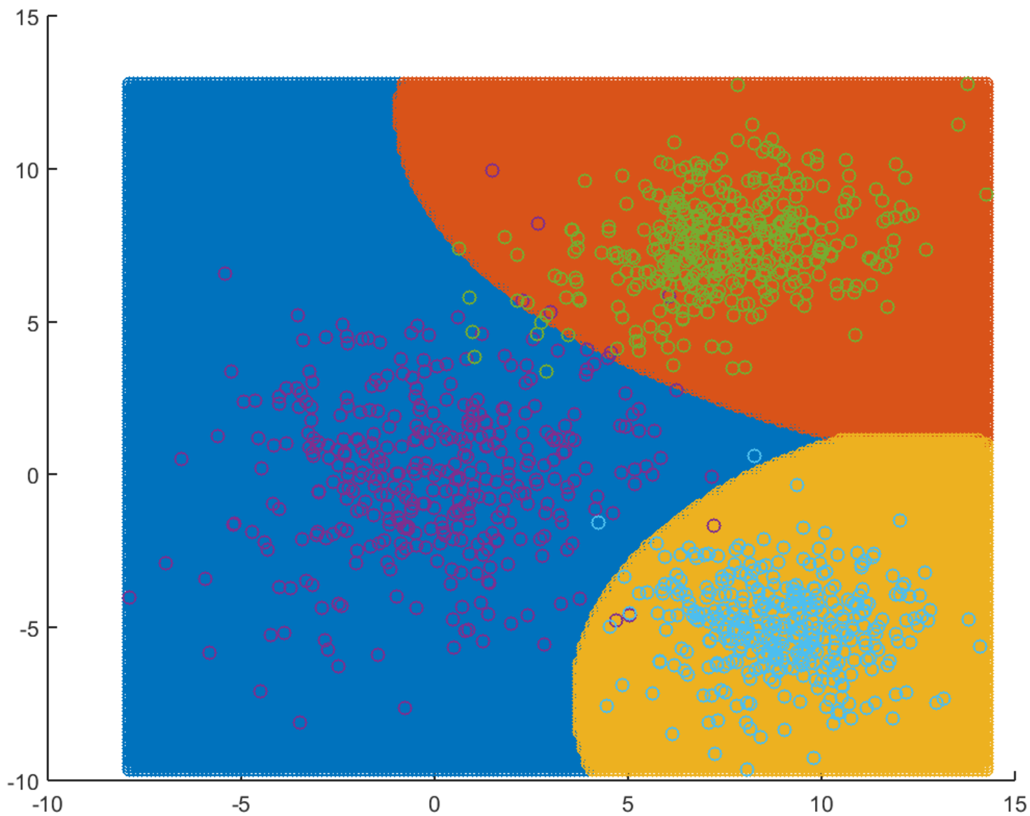
Class2, Class3



Class1, Class3



Decision Region Plot for All Classes



Observation

As the covariance matrices are different for different classes therefore the decision surface is a 2^{nd} order curve. Here, the yellow and the red classes are enclosed in two ellipses and the third class covers the rest of the area. Here, the accuracy is 98.9 % which is better than the Bayes classifier for the different covariance matrices. This indicates that the classes are indeed showing good independence characteristics.

Data-Set II (Real World Data)

1) Bayes Classifier

1.a) Covariance Matrix Same

1.a.i) Average of Covariance Matrices of All Classes

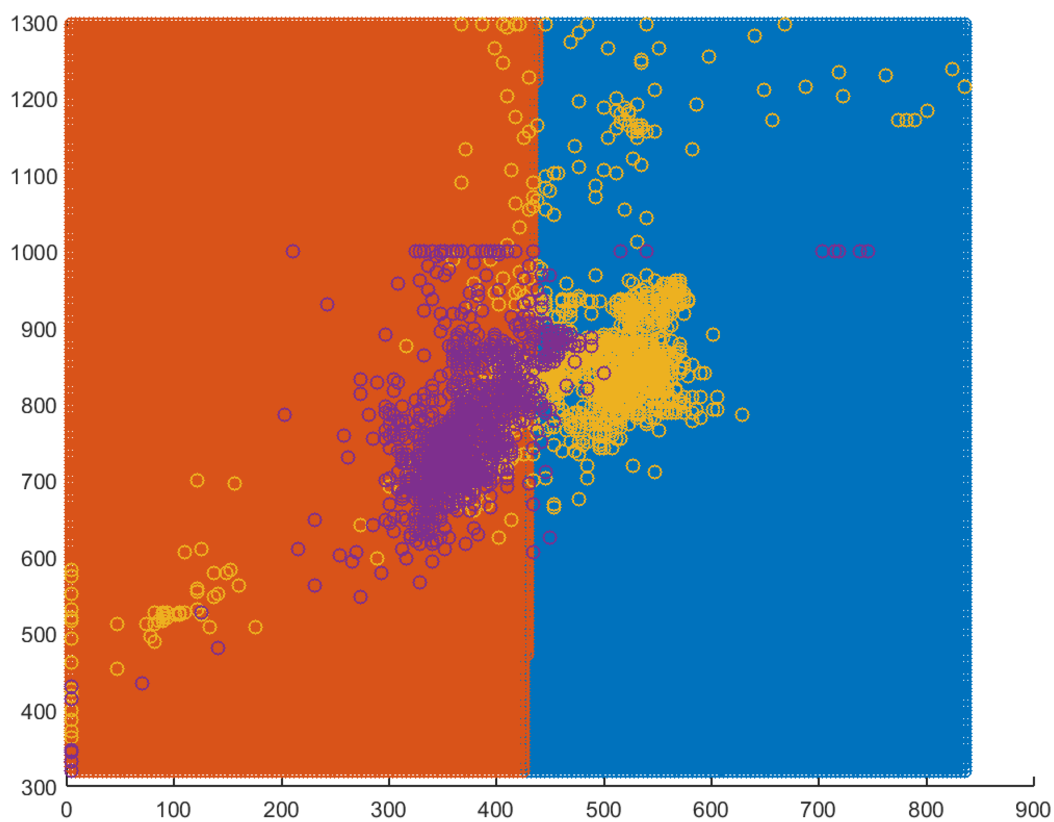
Classification Accuracy: - 87.009%

Confusion Matrix

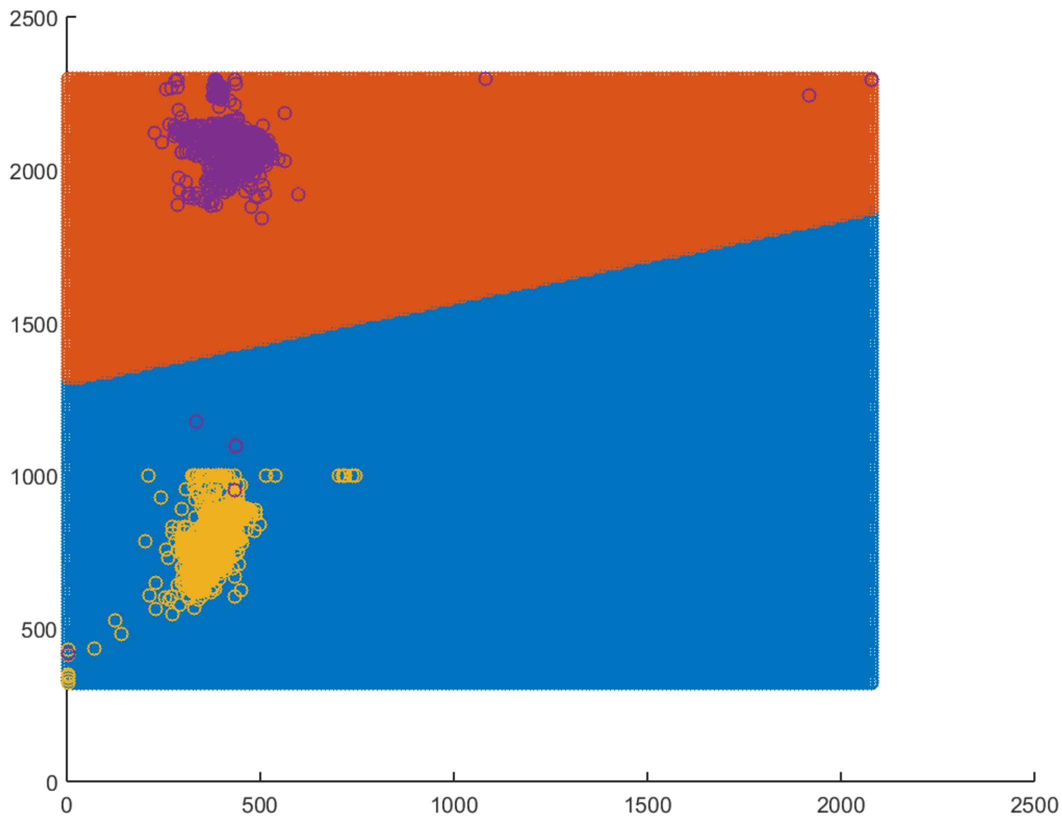
	Class1	Class2	Class3
Class1	426	188	0
Class2	29	593	0
Class3	3	15	555

Decision Region Plot for Pair

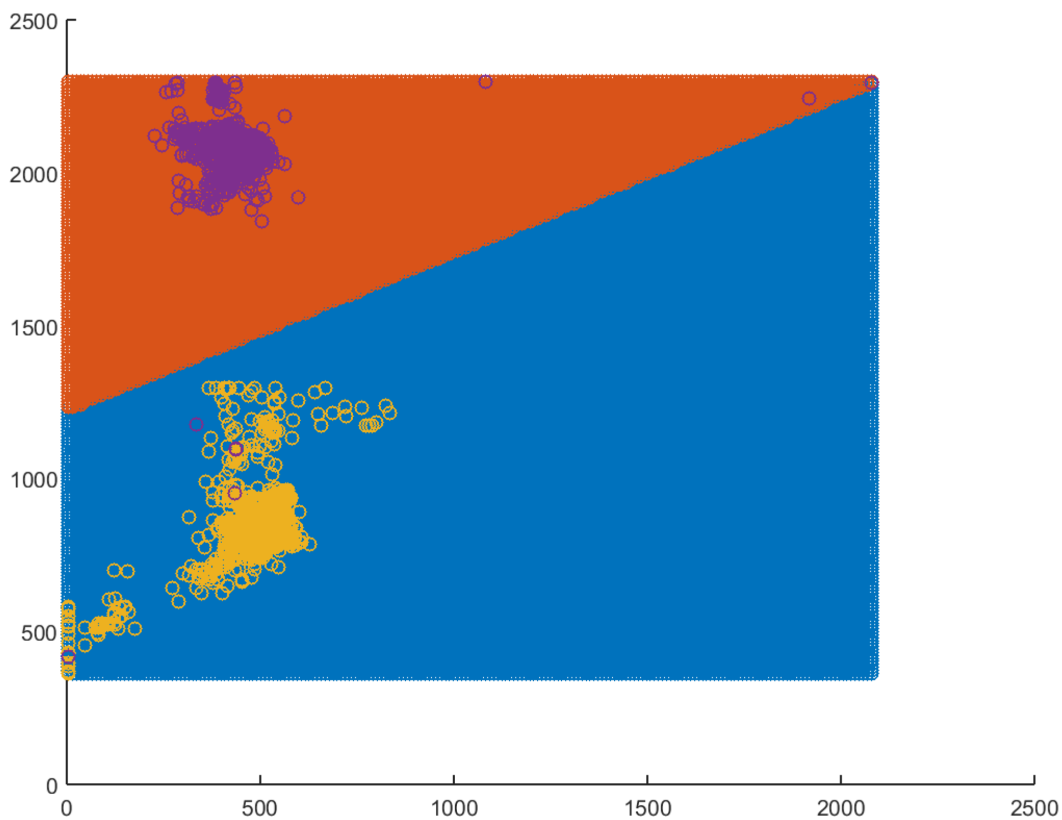
Class 1, Class 2



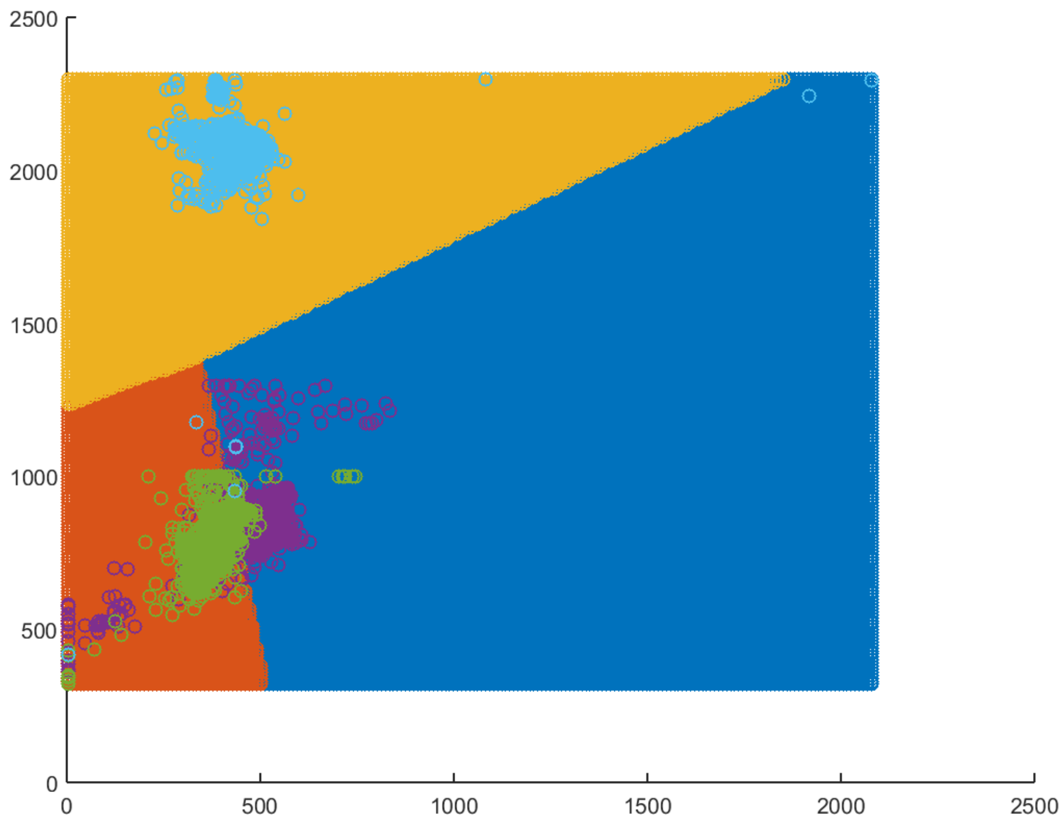
Class 2, Class 3



Class 1, Class 3



Decision Region Plot for All Classes



Observation

After making the covariance matrix same, the decision surface comes out to be linear. The line passes through the midpoint of the line joining the mean of the pair of classes.

As classes 1 and 2 highly overlap each other, the decision boundary distinguishes them with an accuracy of 87%.But class 3 can be well distinguished, because it does not overlap well with other 2 classes.

1.a.ii) Covariance Matrix from all Training Data Combined

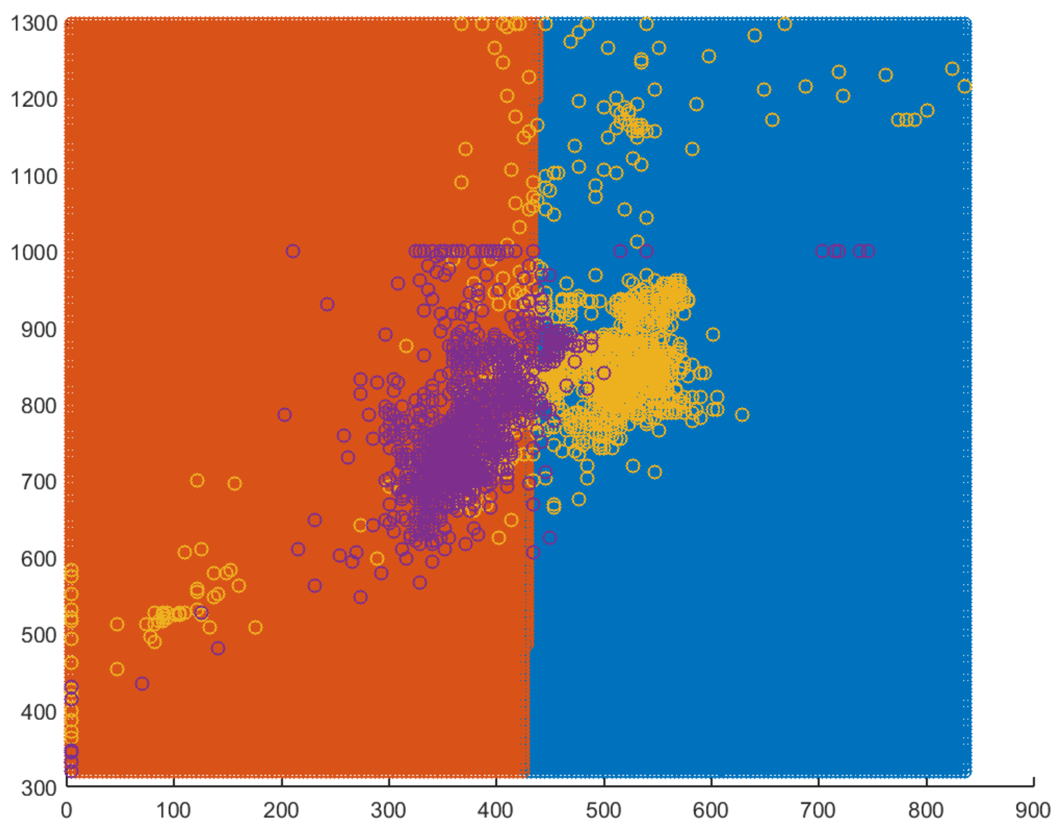
Classification Accuracy: - 86.07%

Confusion Matrix

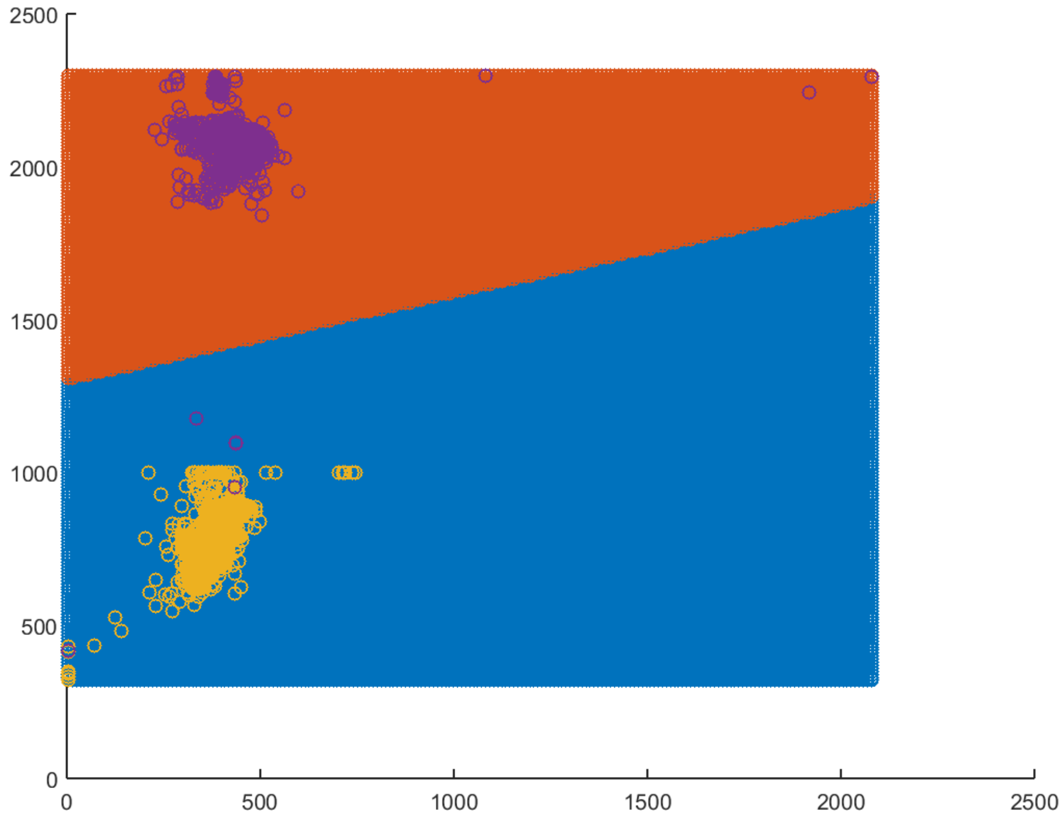
	Class1	Class2	Class3
Class1	409	205	0
Class2	21	601	0
Class3	5	21	547

Decision Region Plot for Pair

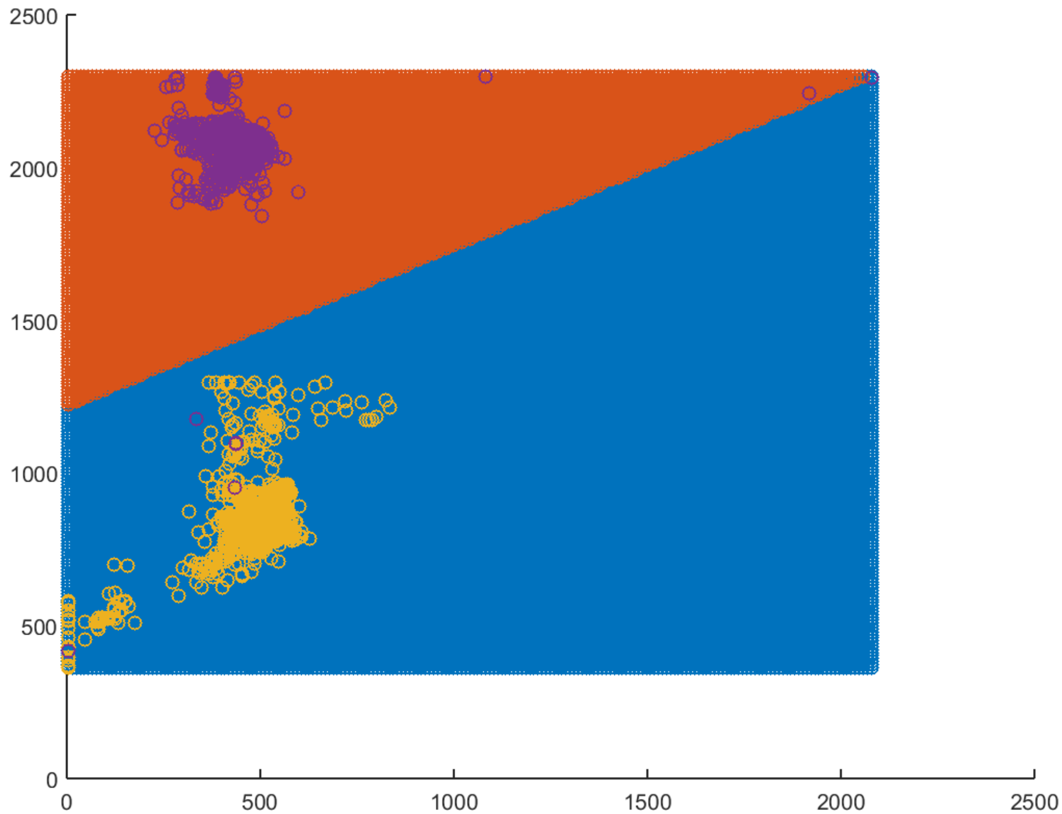
Class 1, Class 2



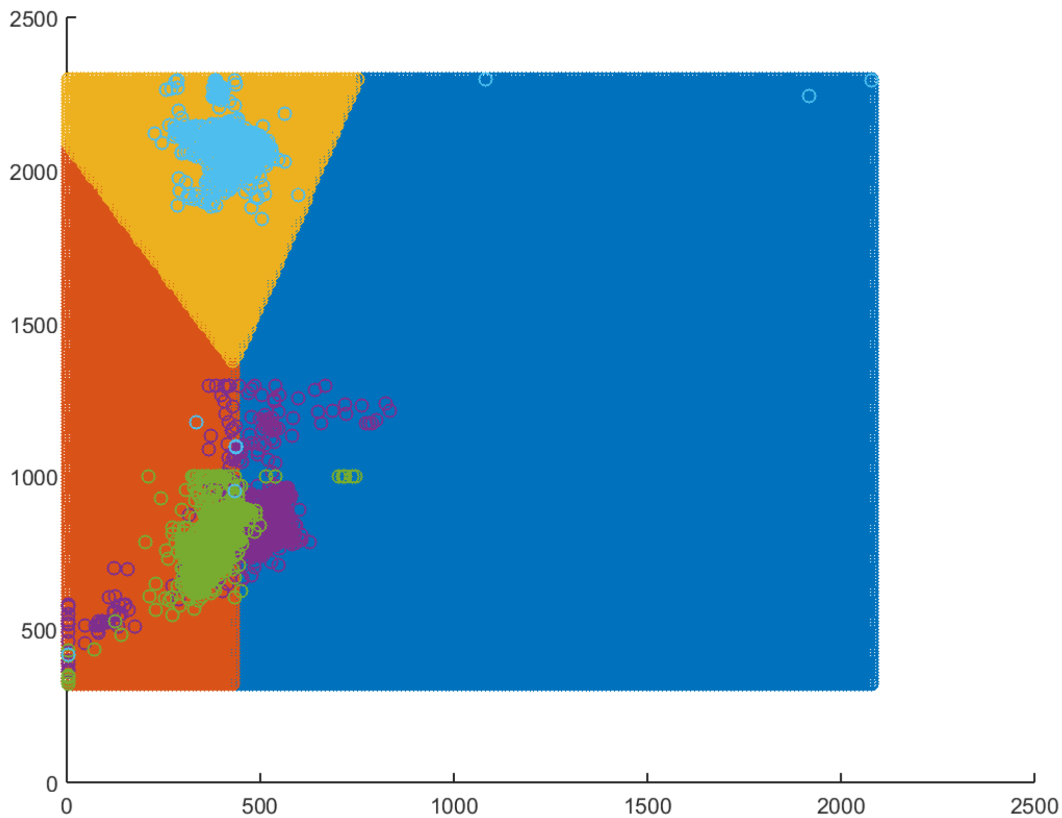
Class 2, Class 3



Class 1, Class 3



Decision Region Plot for All Classes



Observation

After making the covariance matrix same, the decision surface comes out to be linear. The slope for the lines changes as compared to previous case as the method for calculating the covariance is different for both.

The data cannot be distinguished into 1st and 2nd classes similar to the previous case, and the classifier accuracy turns out to be 86%, nearly same in the previous method.

1.b) Covariance Matrix Different

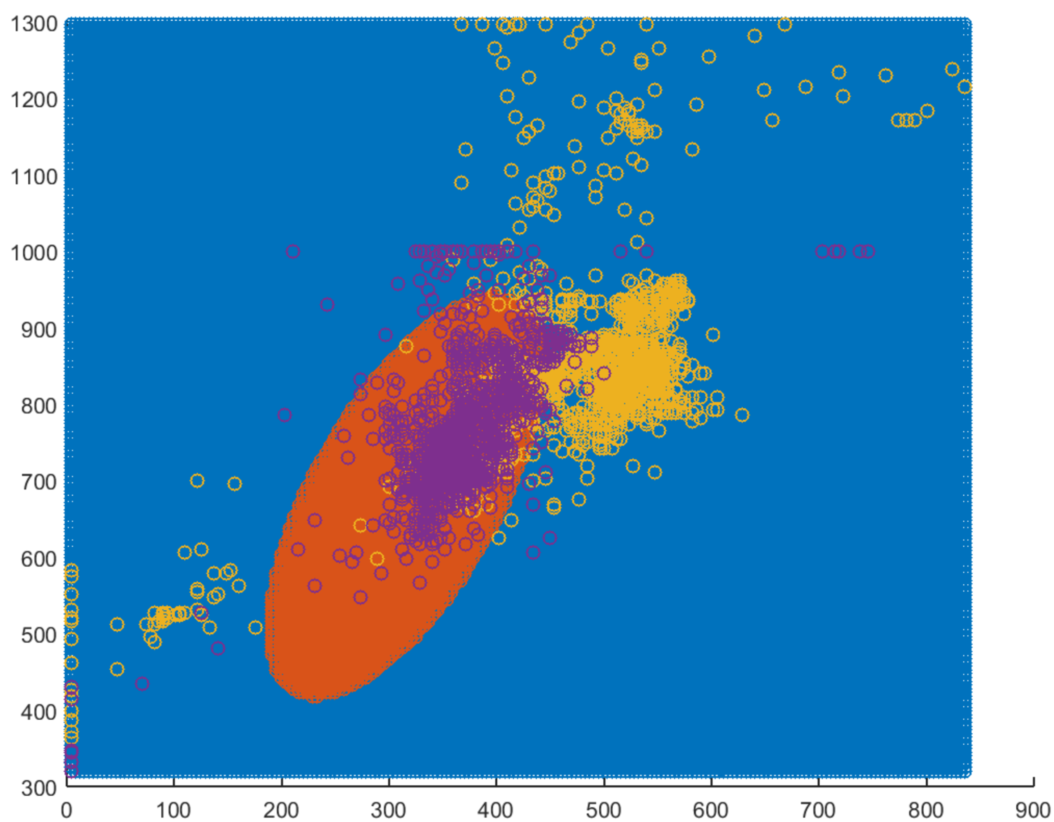
Classification Accuracy: - 82.698%

Confusion Matrix

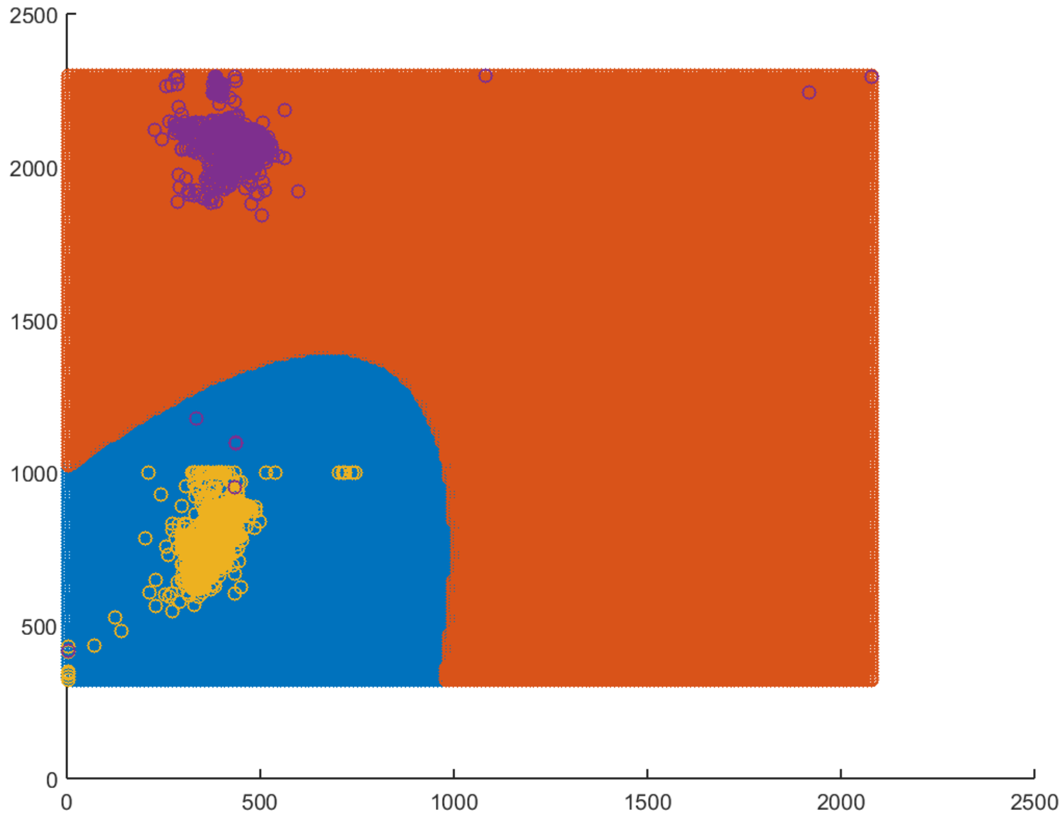
	Class1	Class2	Class3
Class1	411	203	0
Class2	91	531	0
Class3	11	8	554

Decision Region Plot for Pair

Class 1, Class 2

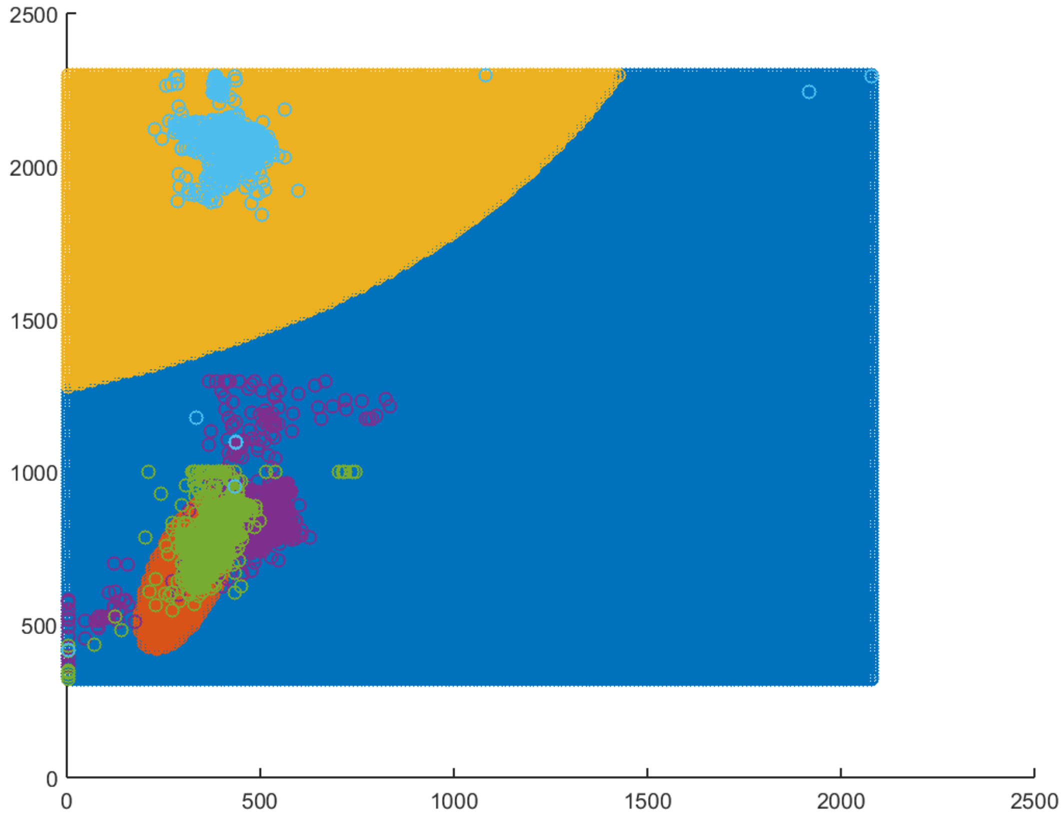


Class 2, Class 3



Class 1, Class 3

Decision Region Plot for All Classes



Observation

Since the covariance matrix is different, the decision surface between classes 1 and 2 turns out to be elliptical, and hyperbolic between other classes.

Since the distribution of each class deviates from Gaussian distribution, the accuracy of this classifier when covariance are different i.ie. 82.7% is lesser as compared to the previous cases, where their covariances were averaged. Hence its performance accuracy drops.

2) Naïve Bayes Classifier

2.a) Covariance Matrix Same (σ^2I)

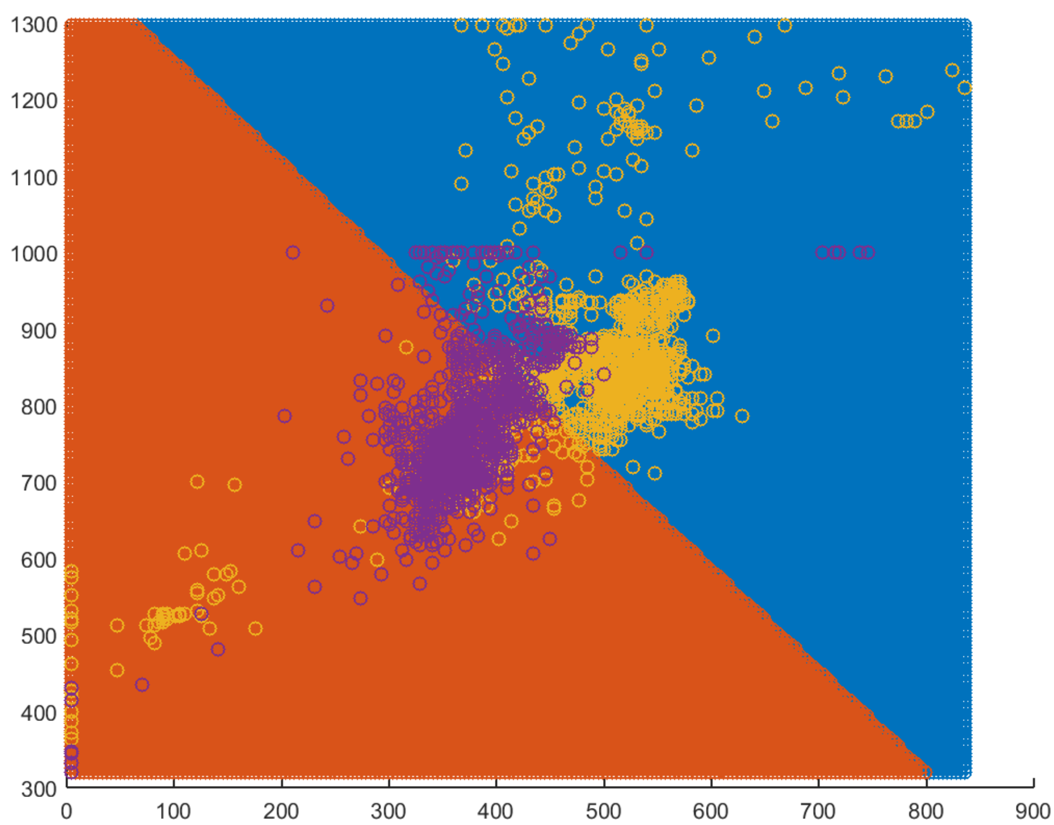
Classification Accuracy: - 86.733%

Confusion Matrix

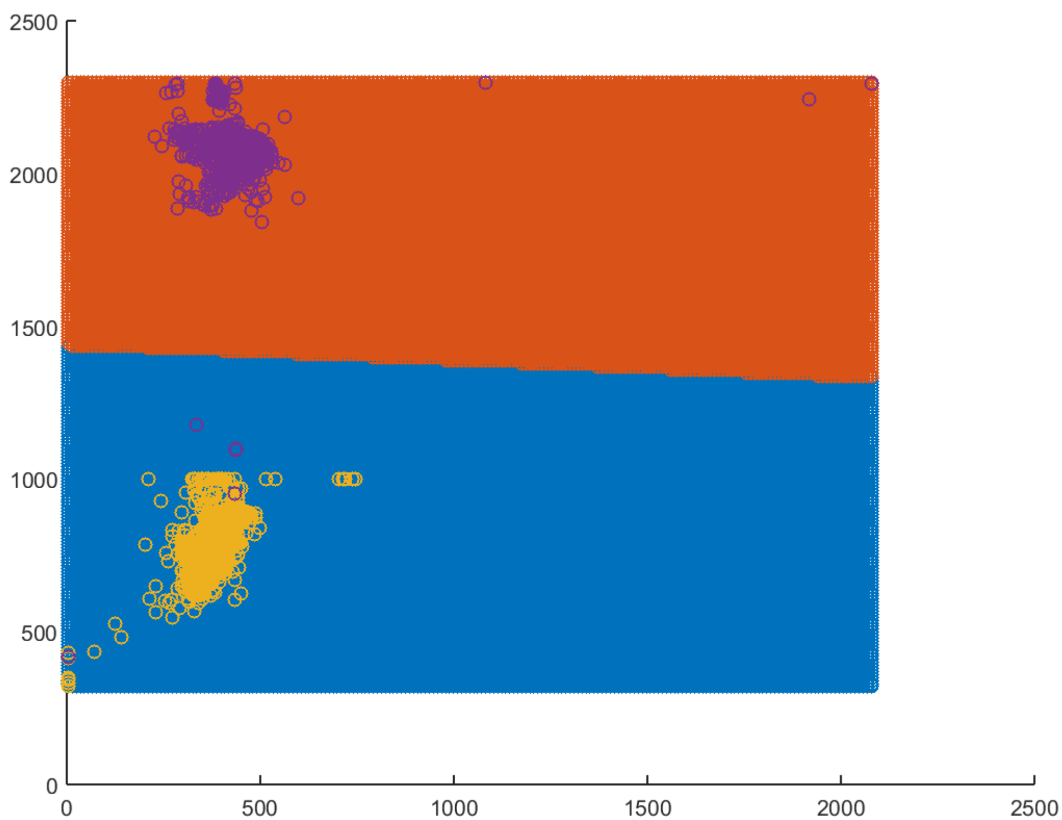
	Class1	Class2	Class3
Class1	442	172	0
Class2	50	572	0
Class3	5	13	555

Decision Region Plot for Pair

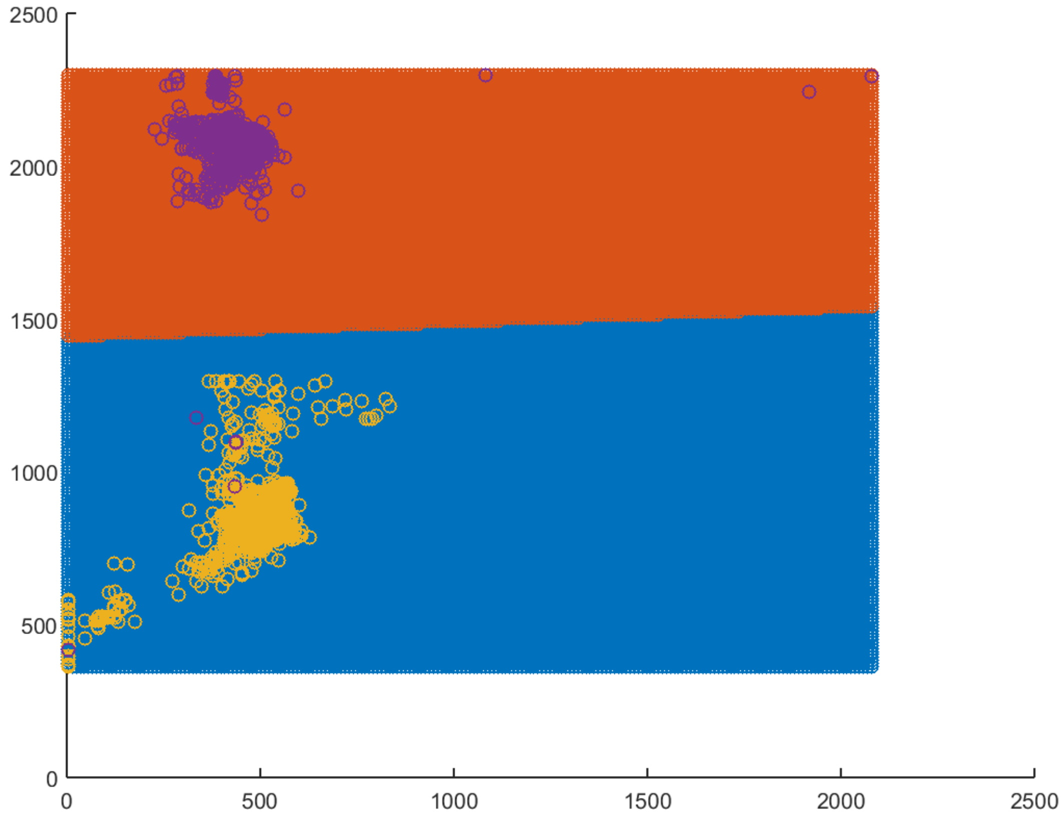
Class 1, Class 2



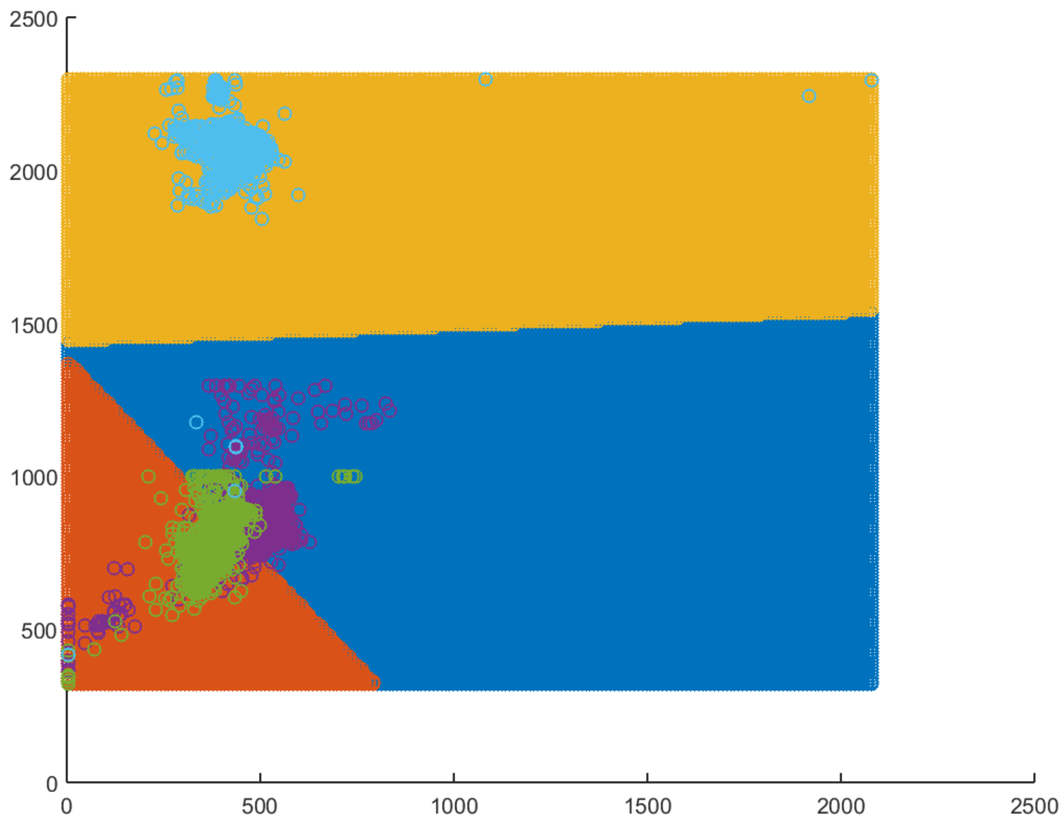
Class 2, Class 3



Class 1, Class 3



Decision Region Plot for All Classes



Observation

After making the covariance matrix same, the decision surface comes out to be linear. The line passes through the midpoint of the line joining the mean of the pair of classes and is perpendicular to it as we have taken covariance matrix of the form σ^2I.

As classes 1 and 2 highly overlap each other, the decision boundary distinguishes them with an accuracy of 87%.But class 3 can be well distinguished, because it does not overlap well with other 2 classes.

2.b) Covariance Matrix Same (C)

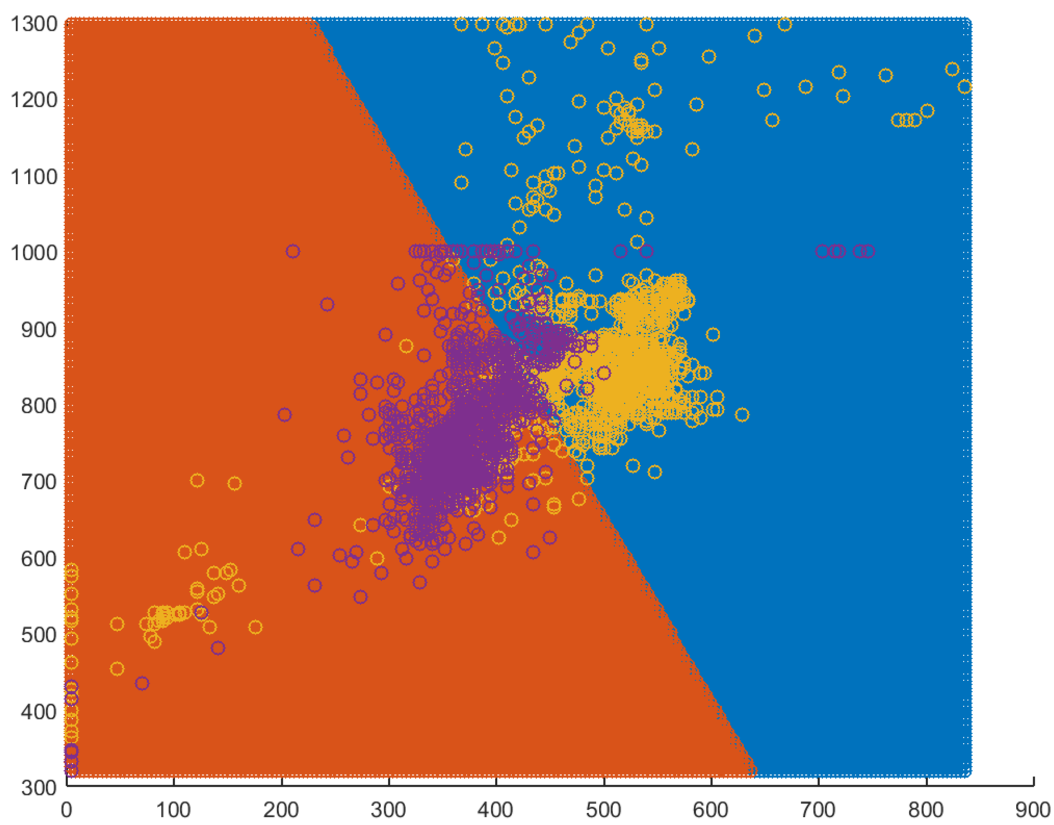
Classification Accuracy: - 86.844%

Confusion Matrix

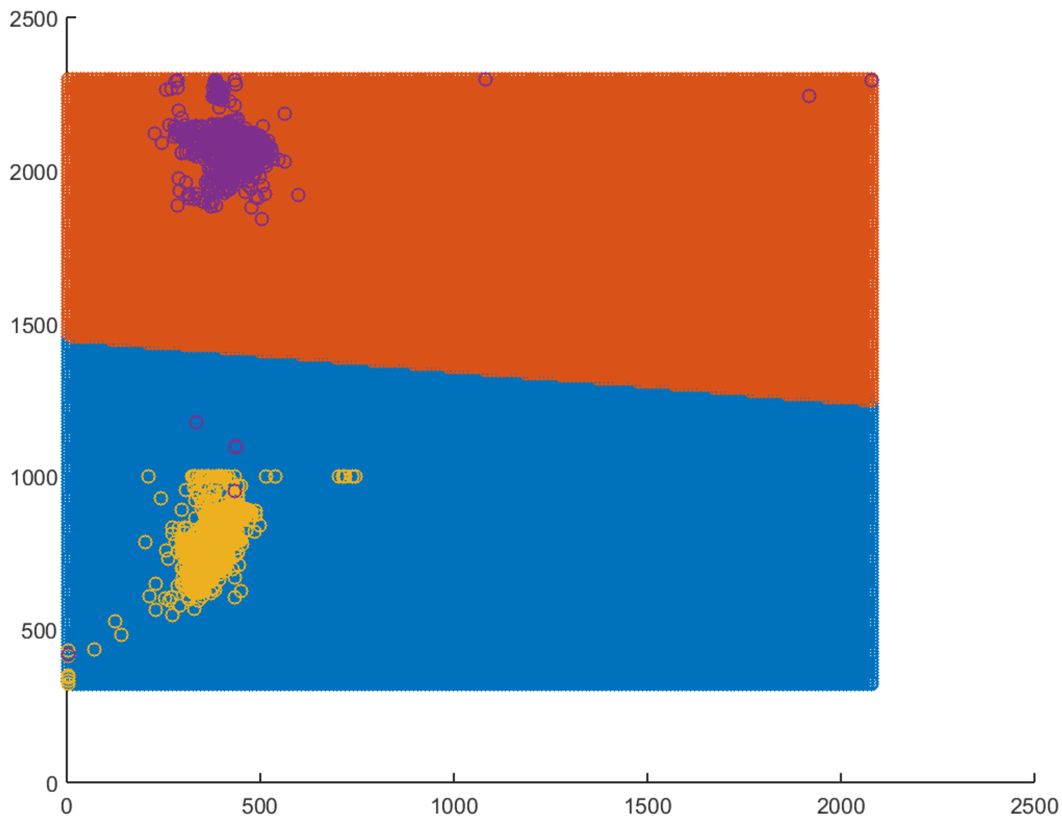
	Class1	Class2	Class3
Class1	434	180	0
Class2	40	582	0
Class3	4	14	555

Decision Region Plot for Pair

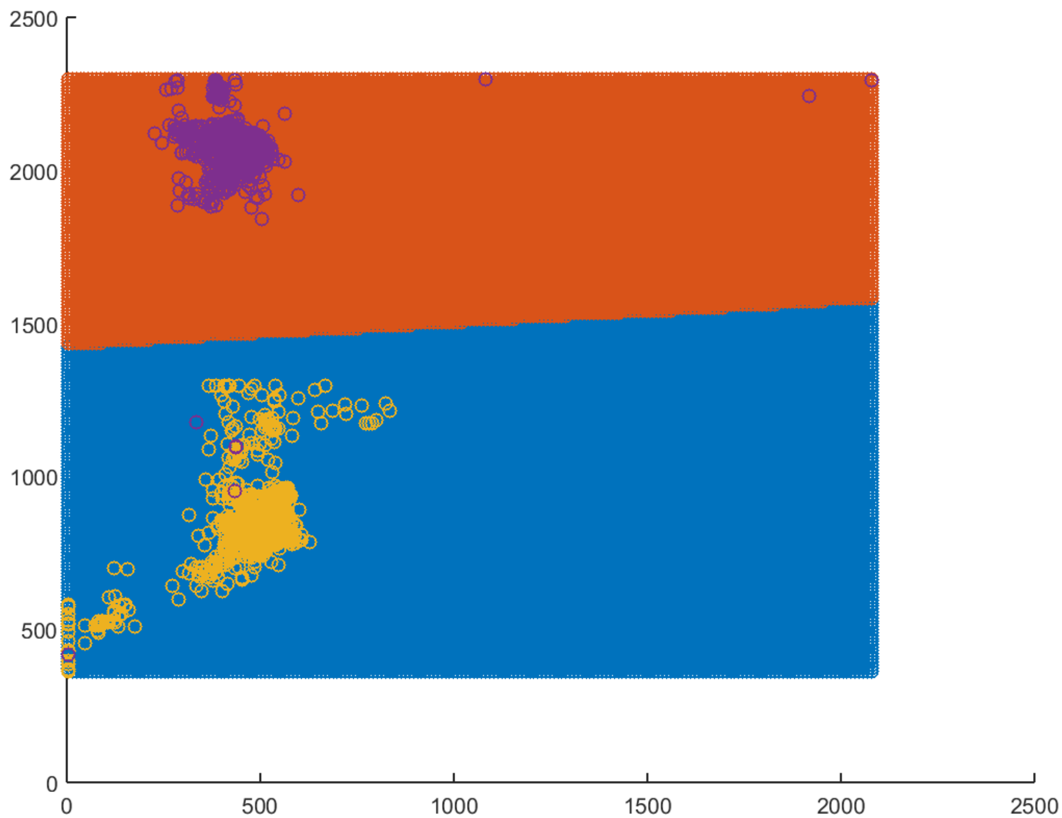
Class 1, Class 2



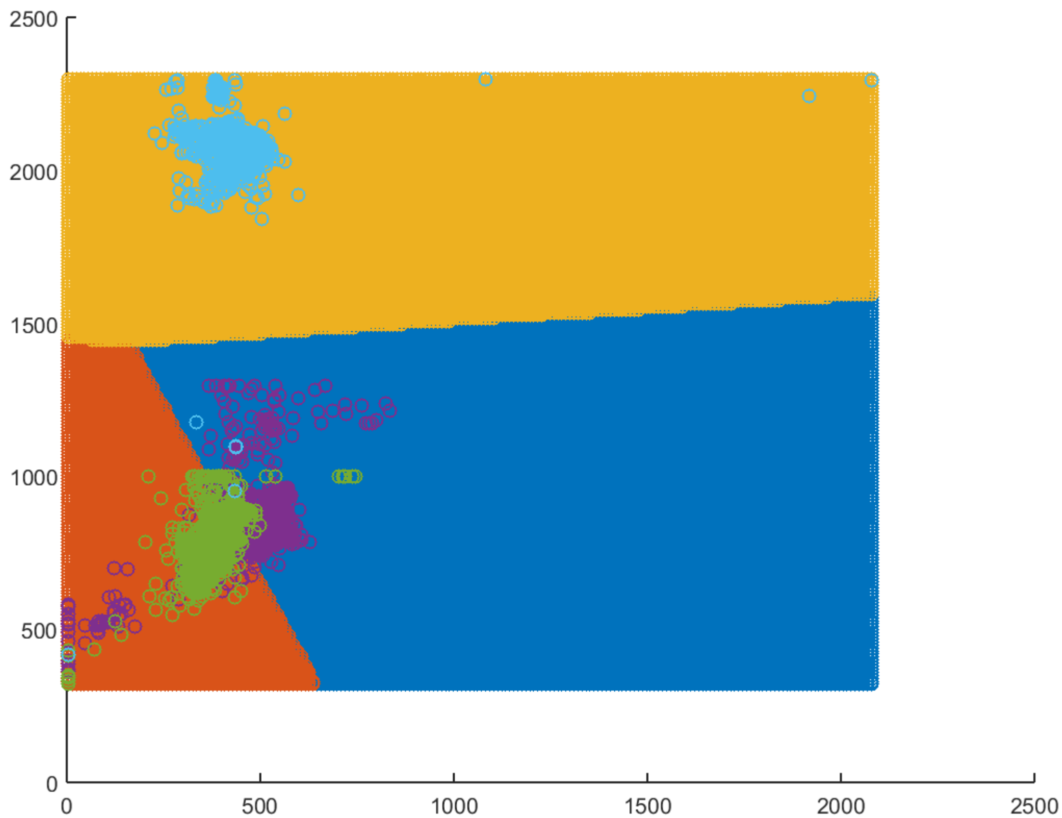
Class 2, Class 3



Class 1, Class 3



Decision Region Plot for All Classes



Observation

After making the covariance matrix same, the decision surface comes out to be linear. The line passes through the midpoint of the line joining the mean of the pair of classes as we have taken covariance matrix same.

As classes 1 and 2 highly overlap each other, the decision boundary distinguishes them with an accuracy of 87%.But class 3 can be well distinguished, because it does not overlap well with other 2 classes.

2.c) Covariance Matrix Different

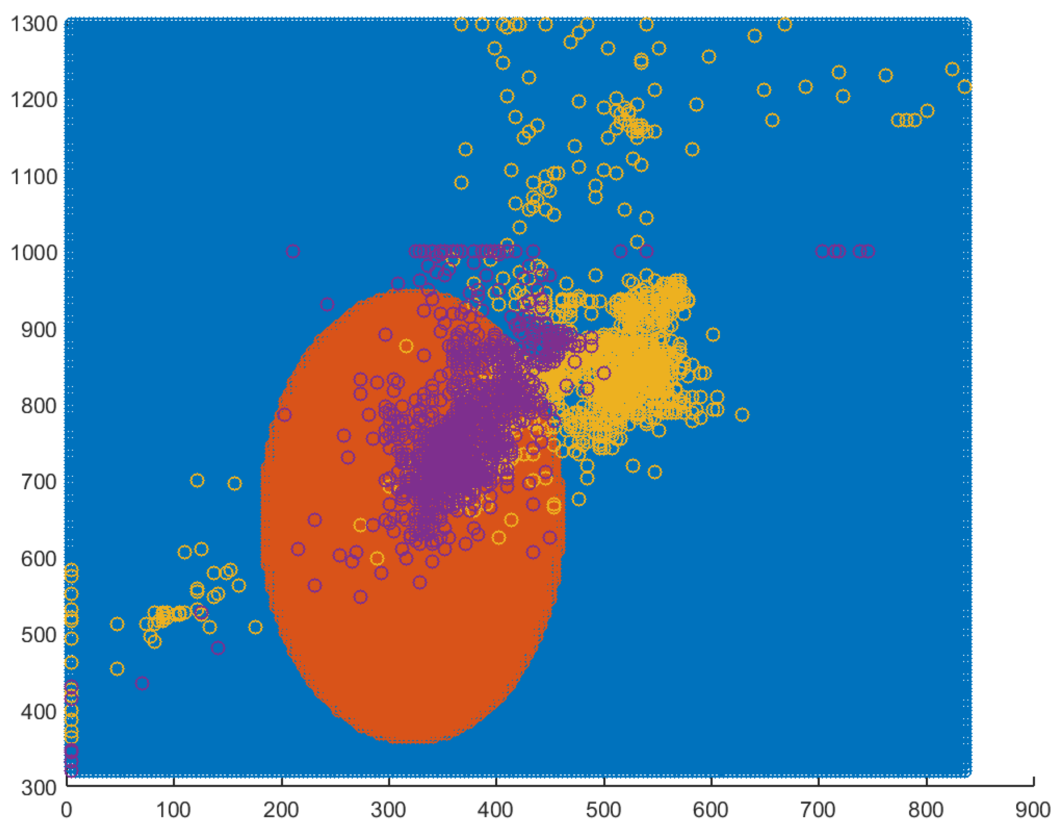
Classification Accuracy: - 84.522%

Confusion Matrix

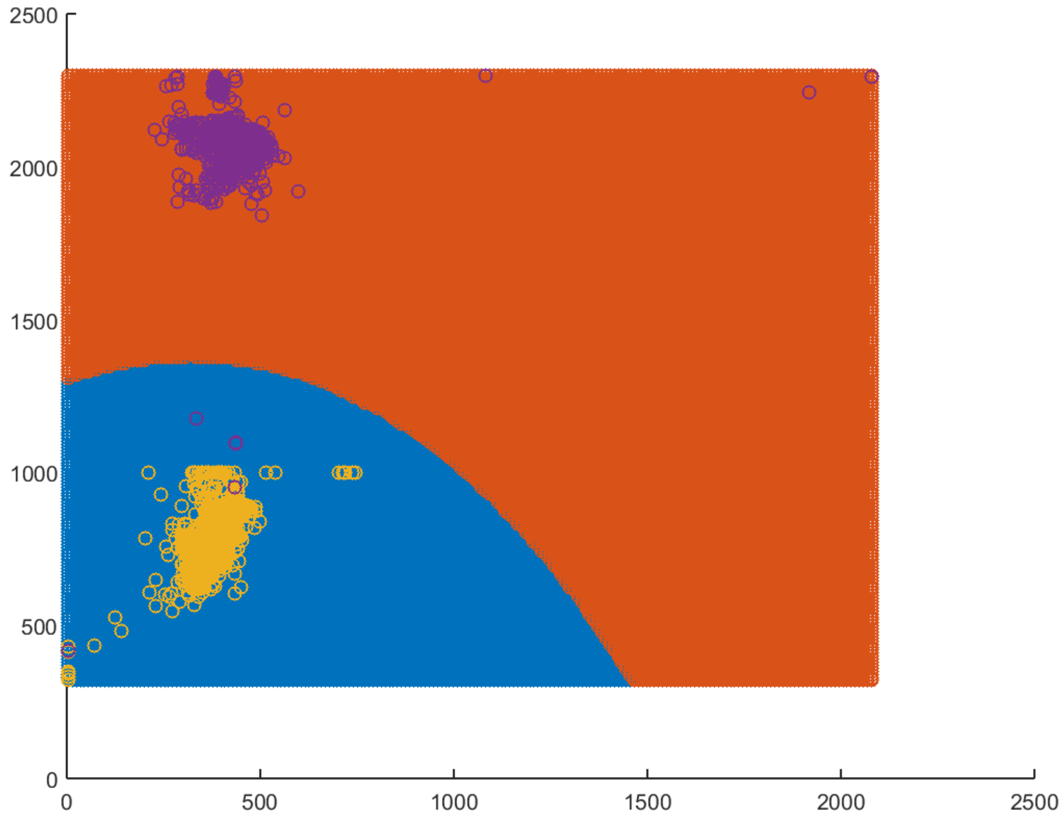
	Class1	Class2	Class3
Class1	429	185	0
Class2	76	546	0
Class3	11	8	554

Decision Region Plot for Pair

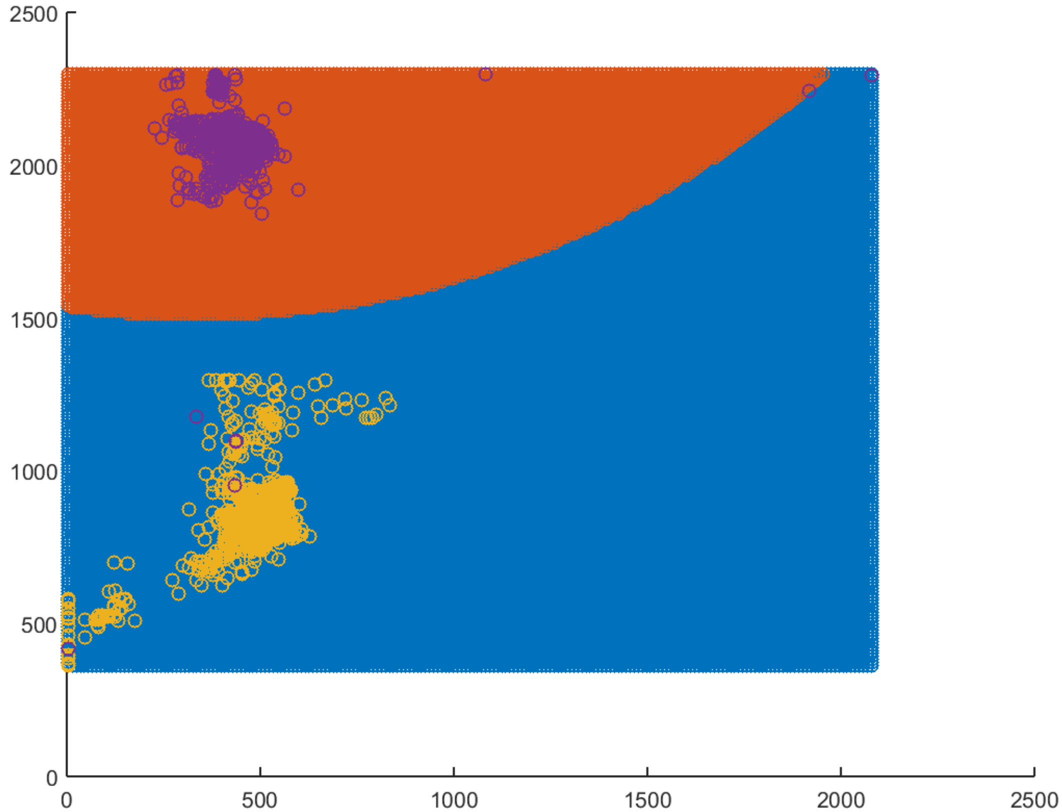
Class 1, Class 2



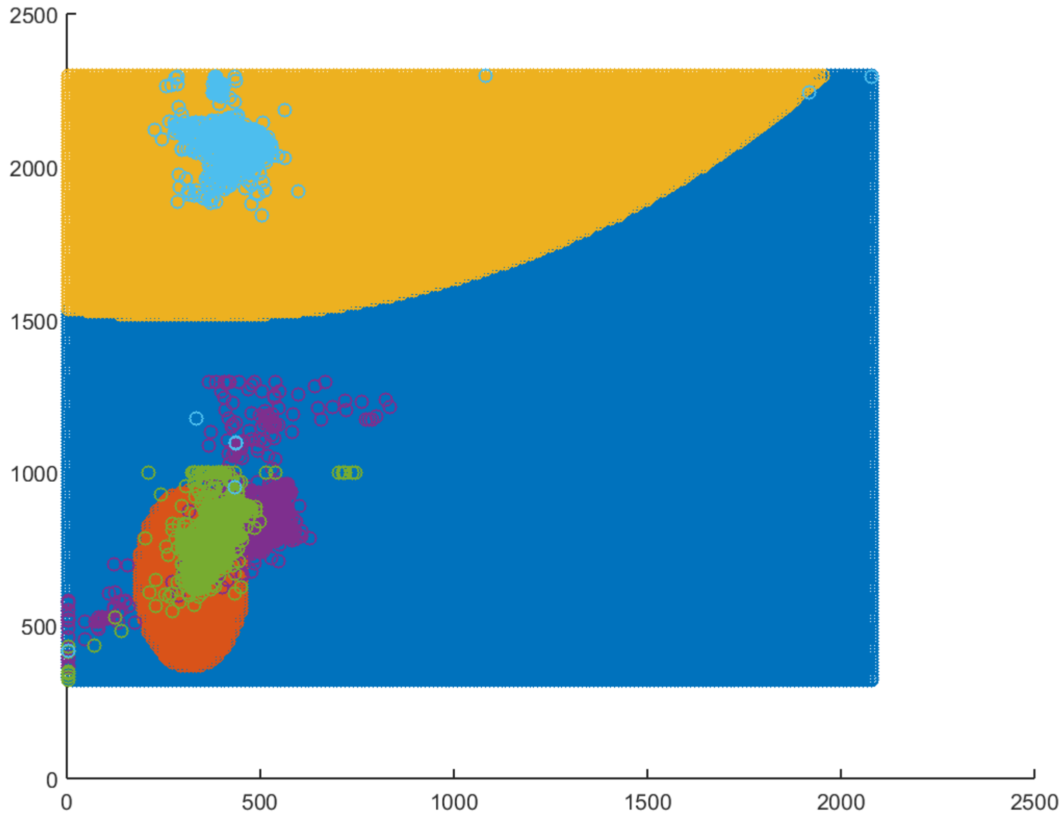
Class 2, Class 3



Class 1, Class 3



Decision Region Plot for All Classes



Observation

After making the decision matrix different, they appear to be same as in bayes classifier but rotated, as the matrix is diagonalized and the cross-covariance terms are zero.

Class 1 and class 2 are distinguished by decision surface of ellipse, which encloses class 2. Their accuracy is worse than naïve bayes, with covariance matrix same, because this distribution deviates from Gaussian and the average covariance distinguishes the classes better than that of individually.

Inference

After studying all the cases, we have come to a conclusion that Bayes and Naïve Bayes where covariance matrices not forced to be same, provide similar and accurate results for almost all types of data sets. But the Naïve Bayes classifiers reduce the complexity of the problem to a great extent, and hence can be used in many of the daily life applications.