CS6690- Pattern Recognition

Report on

Programming Assignment 2

Submitted By

Group IV

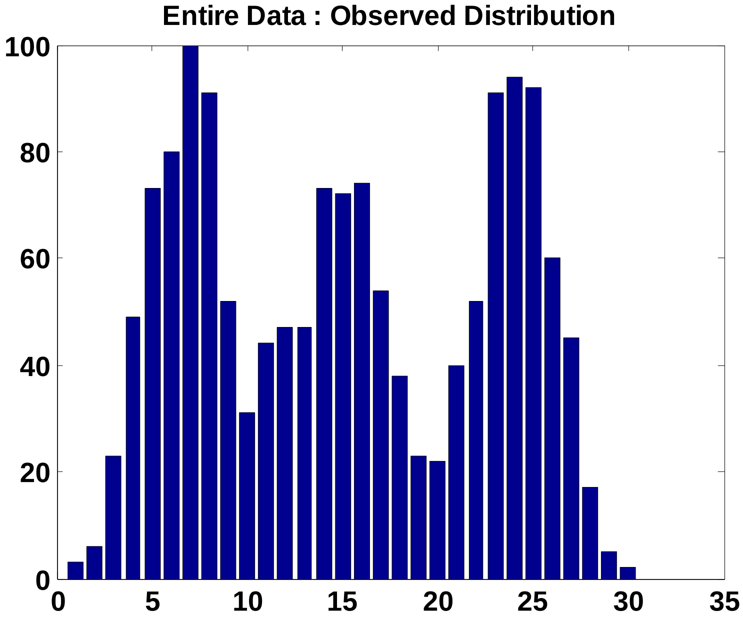
Divya (ED11S002)
Swarna (EE11S019)
Archana (AM11S027)

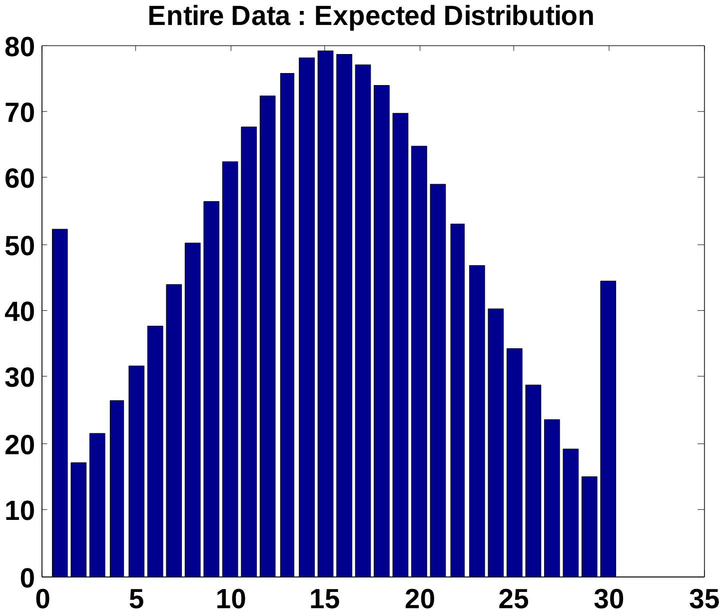
6th October 2011

Dataset 1: 1 Dimensional Data - 3 overlapping classes

Goodness of fit using Chi Square test was performed on the data. First when the entire given data was considered, the data did not fit into a distribution properly as the given data is multimodal. But when considered as separate classes the data fits well into the distribution with 0.01% significance.

Entire data taken for Chi square test:





Chi square Value = 726.6498

Number of bins = 30

Number of Parameters estimated = 2

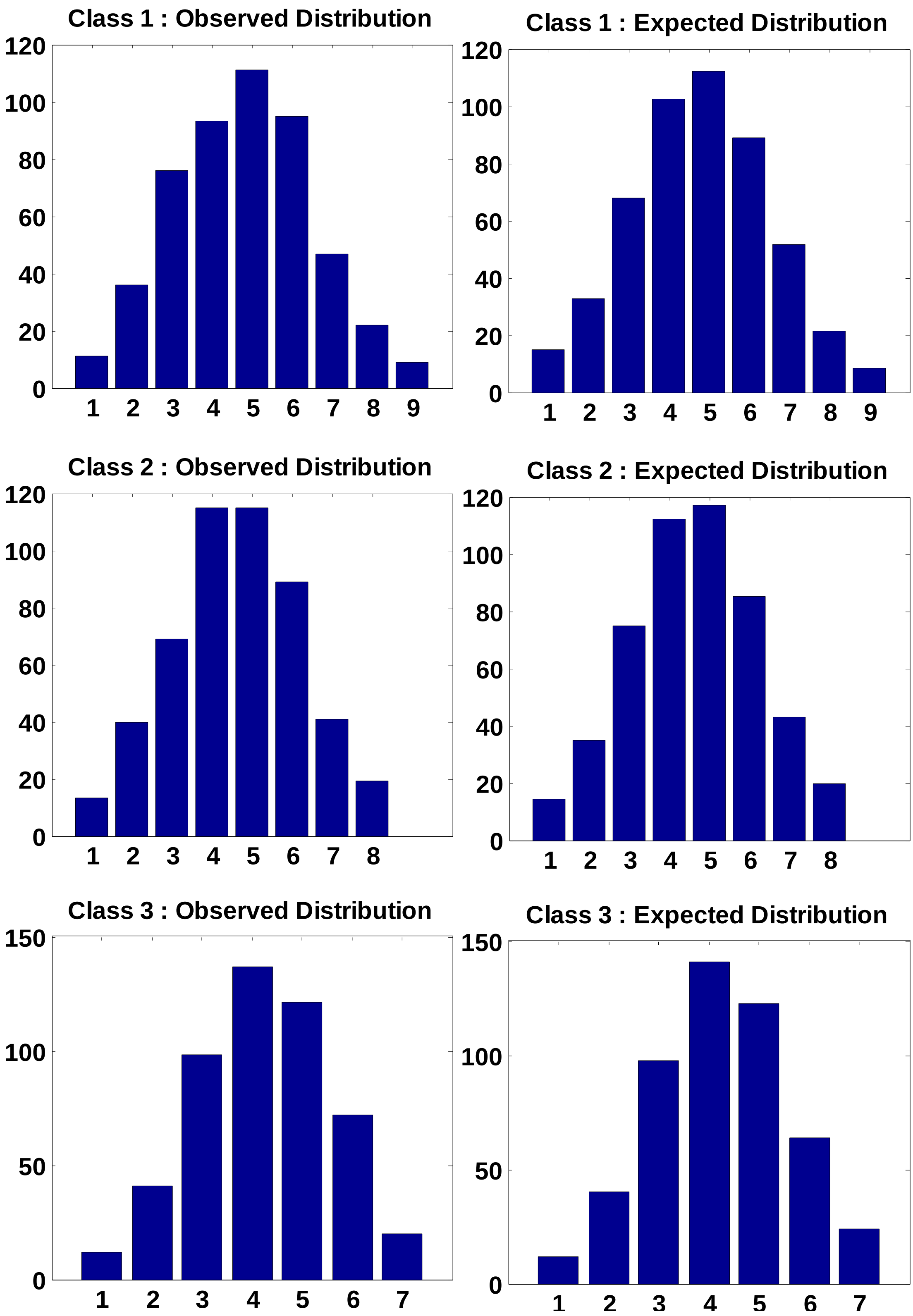
Degrees of freedom = 30 – 1 – 2 = 27

For this distribution, the critical value for the 0.05 significance level is 40.113. Since 726.6498 > 40.113, the null hypothesis that the data are normally distributed is rejected.

Three classes taken separately with 10 bins:

	Chi Square Value	Degrees of freedom	Critical Value for 0.01 Significance level	Null Hypothesis Accepted/ Rejected
Entire Data	726.6498	27	46.963	**Rejected**
Class 1	4.0579	6	16.812	**Accepted**
Class 2	1.7575	5	15.086	**Accepted**
Class 3	1.9071	4	13.277	**Accepted**

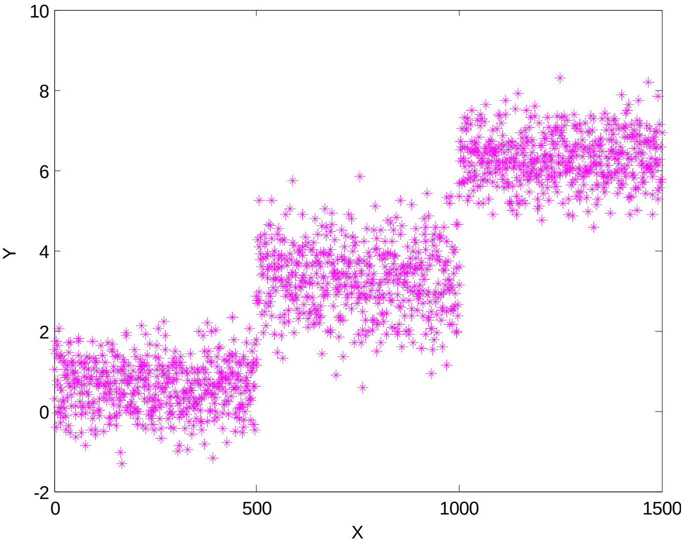
Observed and Expected Distributions of the three classes:



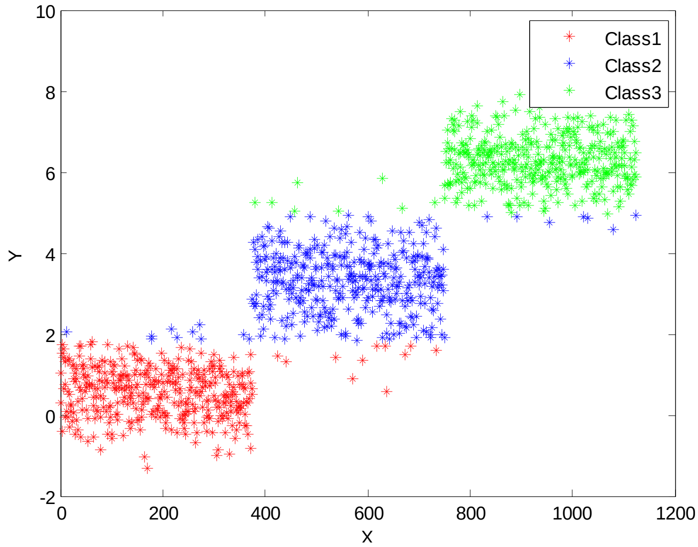
Classification using Bayes classifier:

Training data:

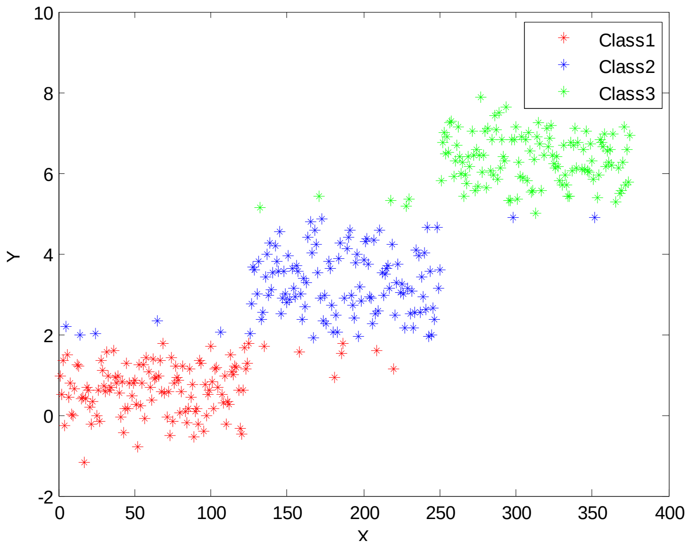
Before classification



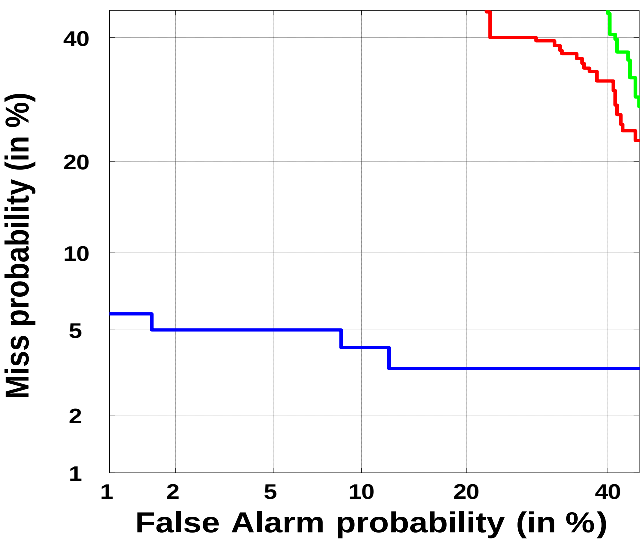
After classification



Test Data after classification



DET Curve



Classification Accuracy = 94.93%

Class wise Accuracy = [96 90.4 98.4]

Confusion Matrix = $\begin{bmatrix} 120 & 5 & 0 \\ 7 & 113 & 5 \\ 0 & 2 & 123 \end{bmatrix}$

The data is well classified using the Bayes classifier, except for a few points near the boundary which are classified wrongly.

Dataset 2: 2 Dimensional Data

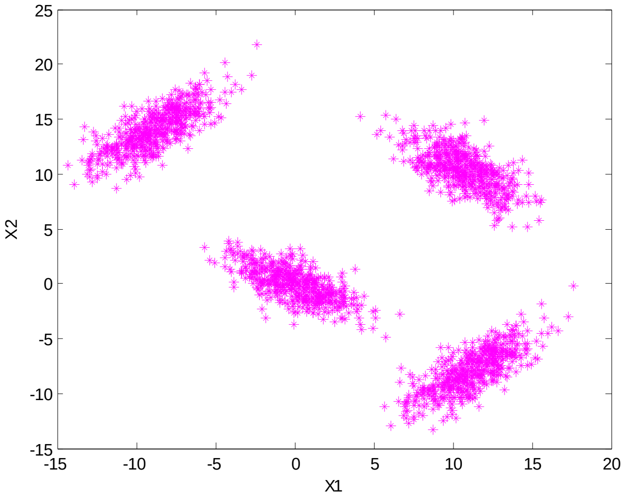
In these exercises, the training and testing data sets are given for each class. From the given data set the 75% of the data is used as training set to design the classifier and the remaining 25% data is used to test the classifier. Depending upon the assumption of the covariance matrix the decision region will change.

In the Bayes classifier, the classifier is designed with the assumption that the feature vectors are dependent. Here the feature vector is two dimensional vector with two features x1 and x2.

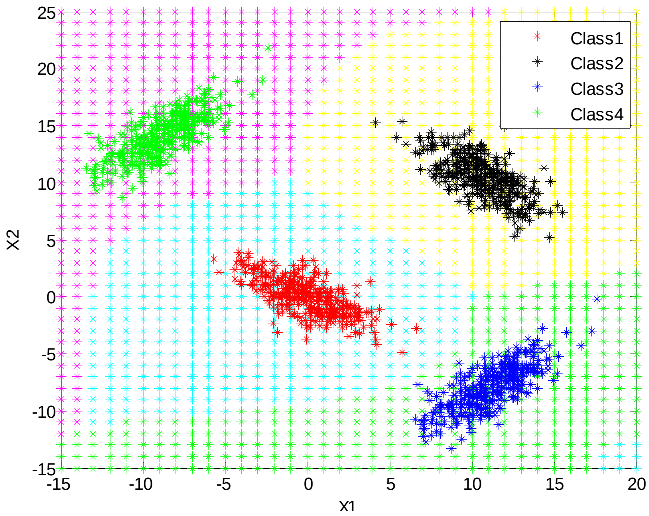
In Naive Bayes classifier the classifier is designed based on the assumption that all the features are independent. But in real world this is not the case, all the features are dependent each other.

A) Linearly separable classes

Training data before classification :

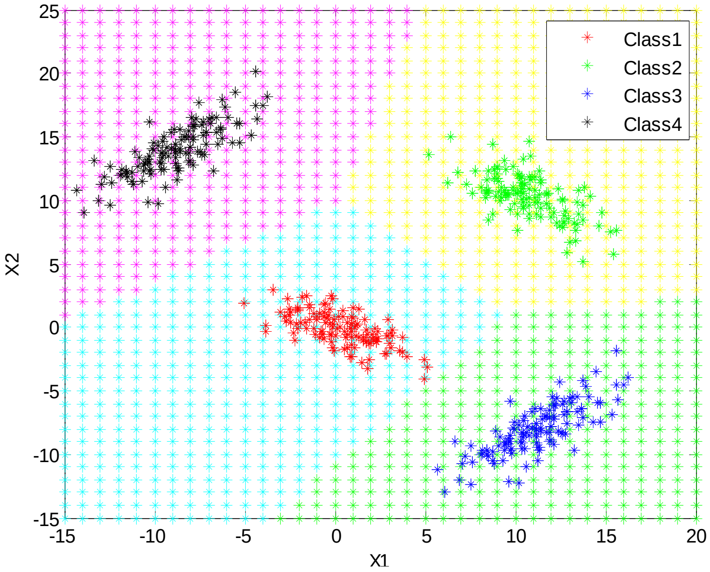


Training data after classification :

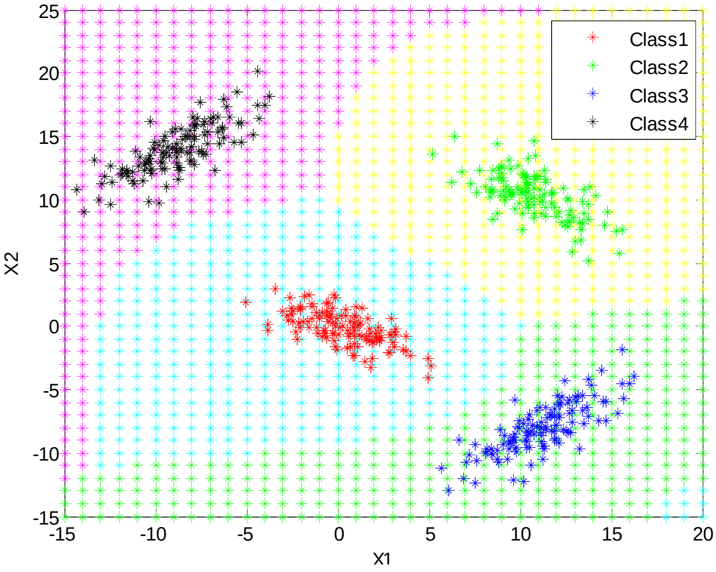


(i) Bayes classifier

$\sum_i = \sum$



$\sum$ = Different



Performance of Bayes classifier:

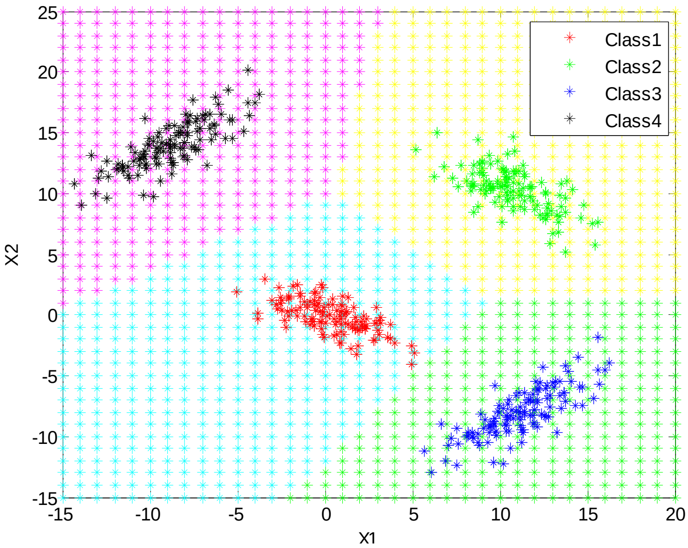
Covariance Matrix	$\sum_i = \sum$	$\sum$ = Different
Accuracy	100	100
Class-wise Accuracy	100 100 100 100	100 100 100 100
Confusion Matrix	125 0 0 0 0 125 0 0 0 0 125 0 0 0 0 125	125 0 0 0 0 125 0 0 0 0 125 0 0 0 0 125

Observations:

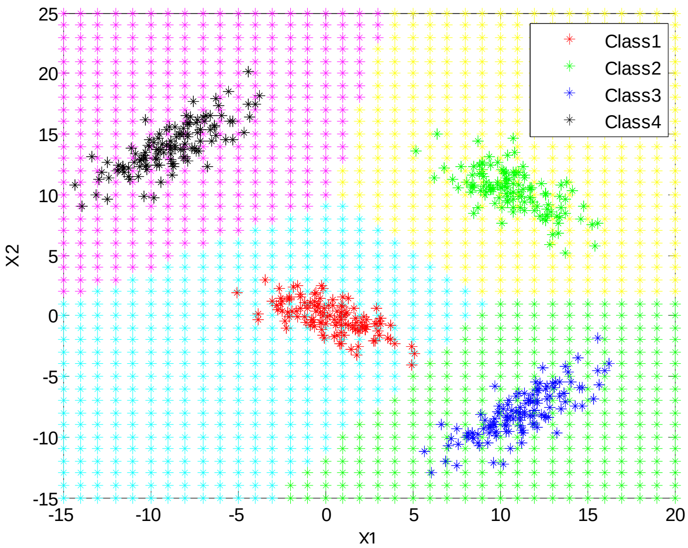
- The given data is well separated in to 4 classes.
- In different covariance case the quadratic term are present in the discriminant function, so the shape of the decision region will be quadratic in nature.
- But when we consider the same covariance matrix for all classes, the decision boundary becomes linear.
- Here as all the classes are well separated, an accuracy of 100% is observed in all cases.

Naïve Bayes classifier

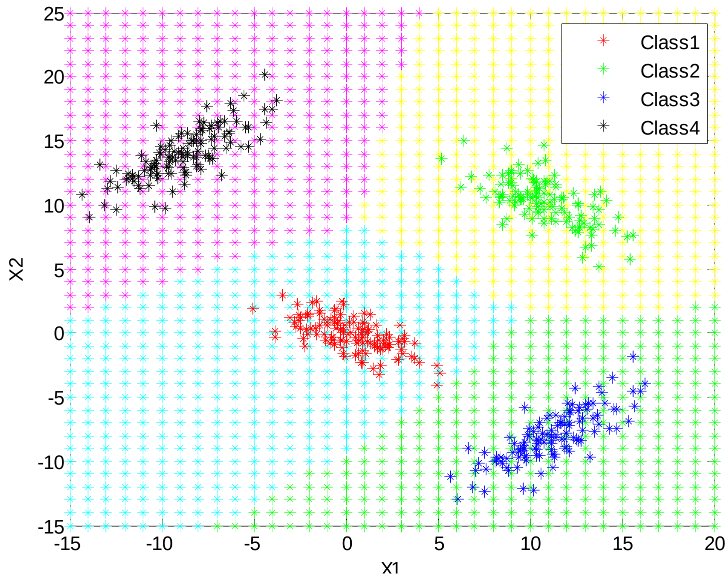
$\sum = \sigma^2 I$



$\sum_i = \sum$



$\sum$ = Different

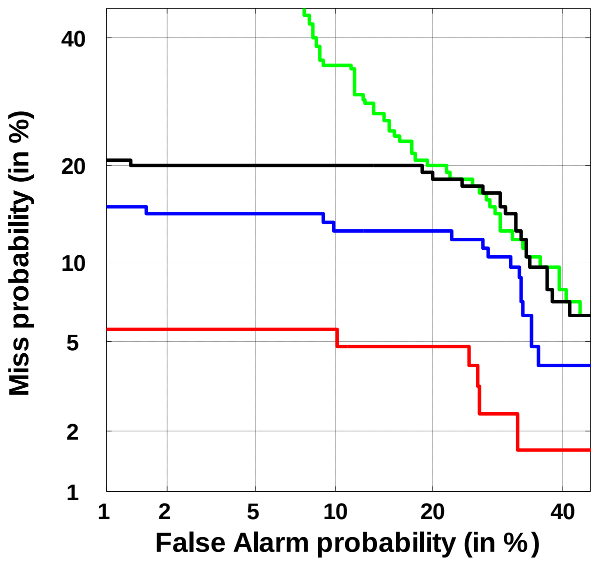


Performance of Naive Bayes classifier:

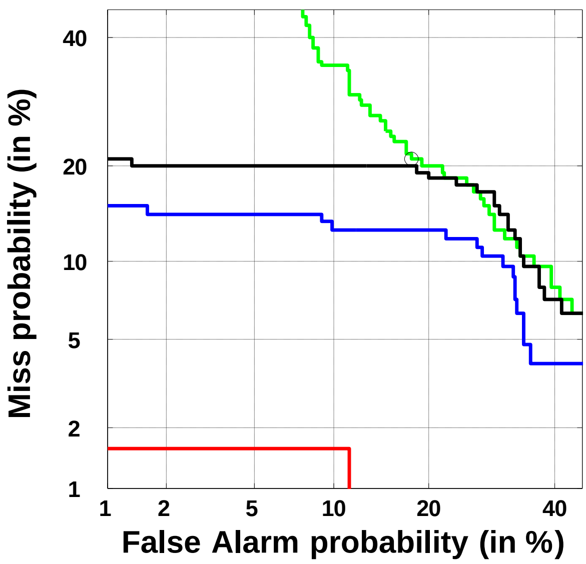
Covariance Matrix	$\sum = \sigma^2 I$	$\sum_i = \sum$	$\sum$ = Different
Accuracy	100	100	100
Class-wise Accuracy	100 100 100 100	100 100 100 100	100 100 100 100
Confusion Matrix	125 0 0 0 0 125 0 0 0 0 125 0 0 0 0 125	125 0 0 0 0 125 0 0 0 0 125 0 0 0 0 125	125 0 0 0 0 125 0 0 0 0 125 0 0 0 0 125

As seen before, as the data is well separated, the performance in all three cases is the same and an accuracy of 100% is obtained.

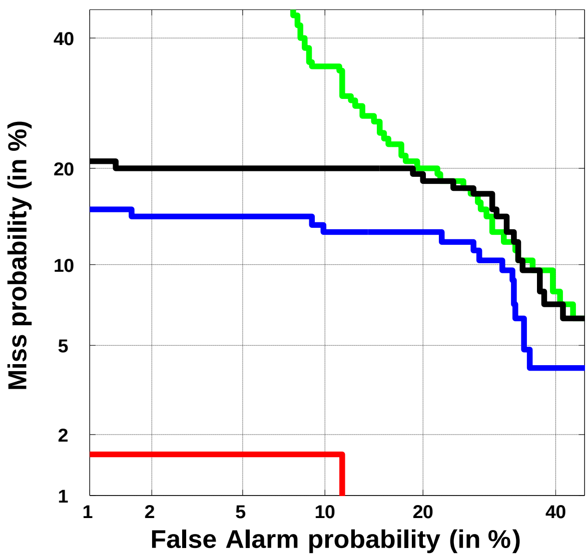
DET Plot - 2a



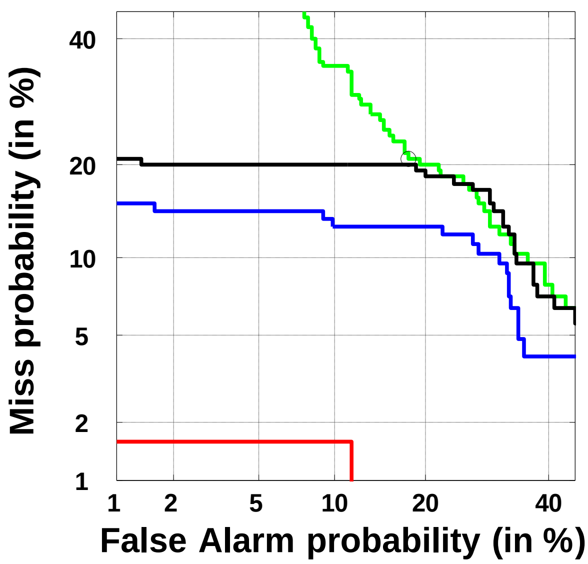
DET Curve - 2a- Same C



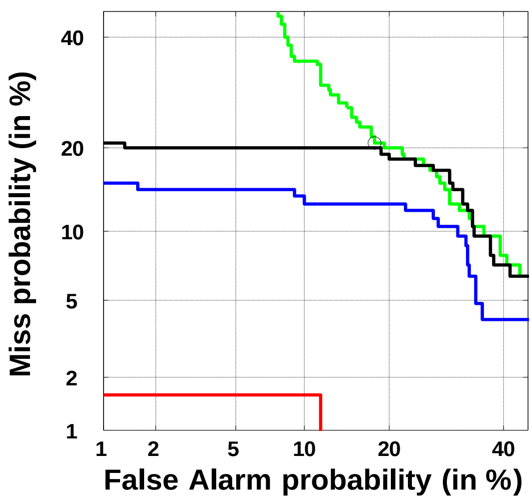
DET Curves - 2a - NB



DET Curves - 2a - NB - Same C



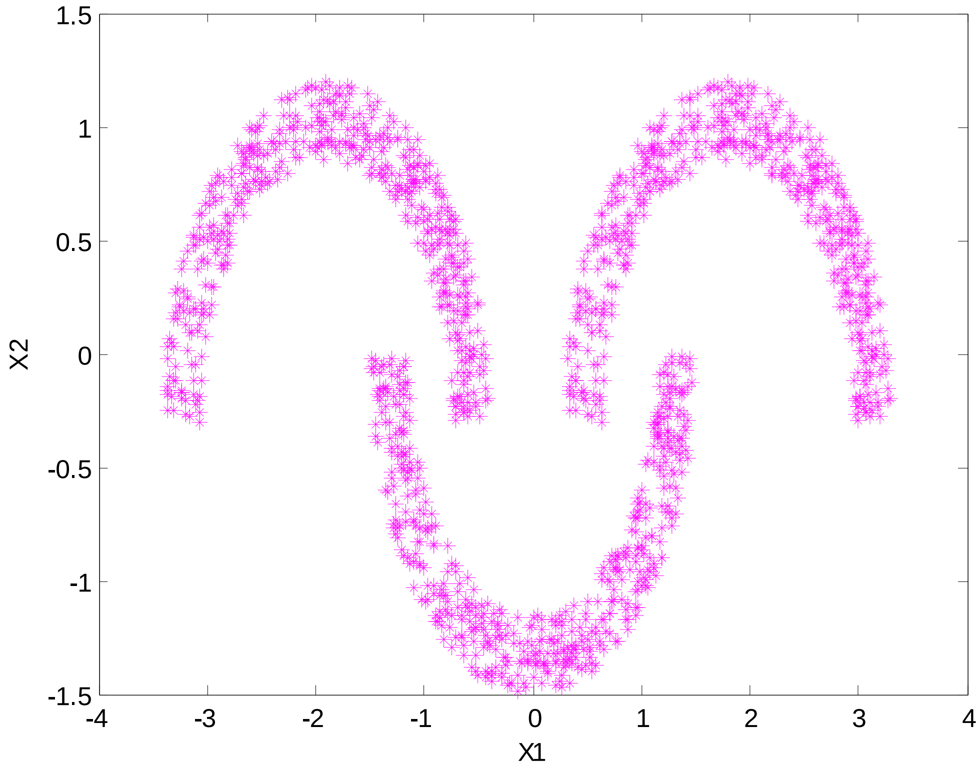
DET Curve - 2a- NB - Ci



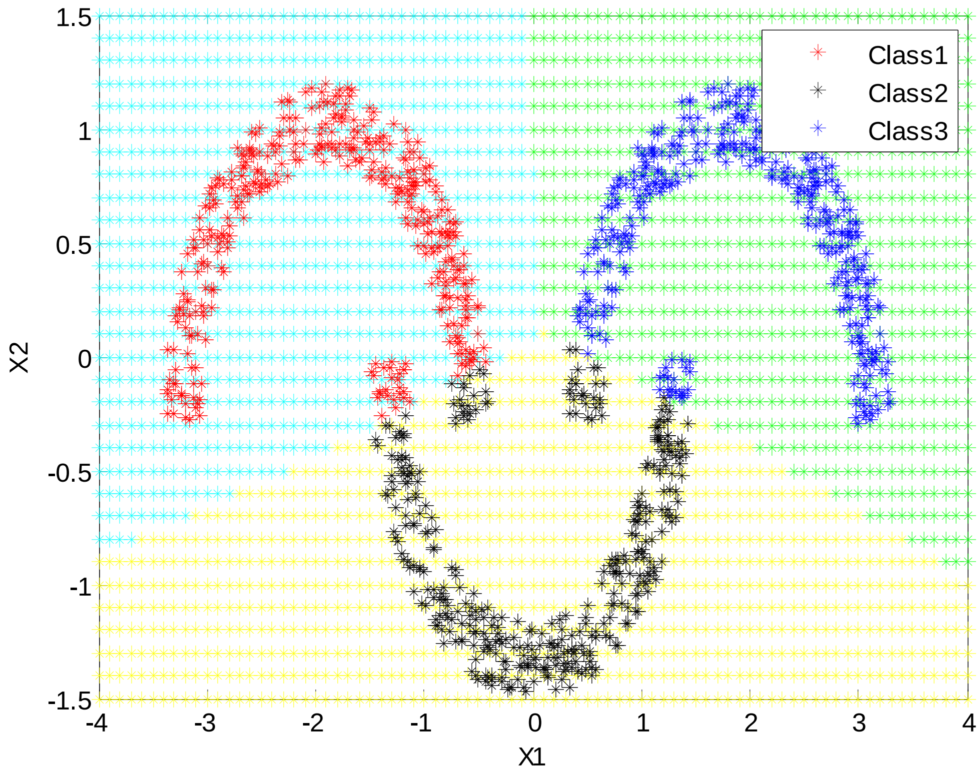
The performance is same in all the cases except in Bayes Classifier using different covariance matrices. This is because the classes are well separated.

B) Non-Linearly separable classes

Training data before classification :

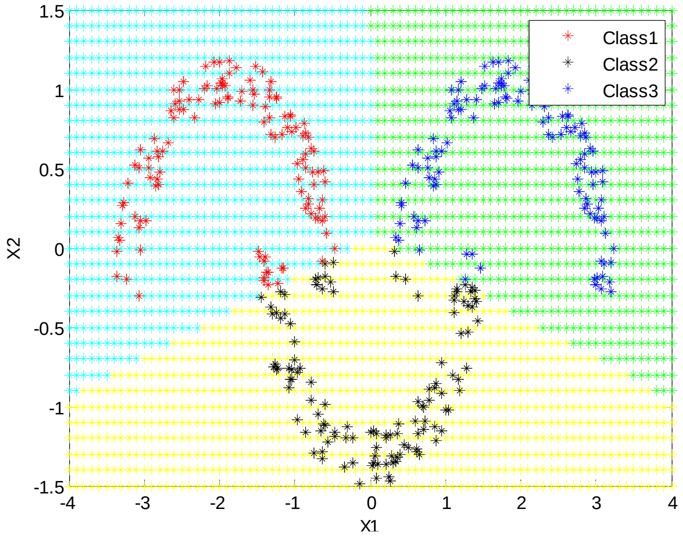


Training data after classification :

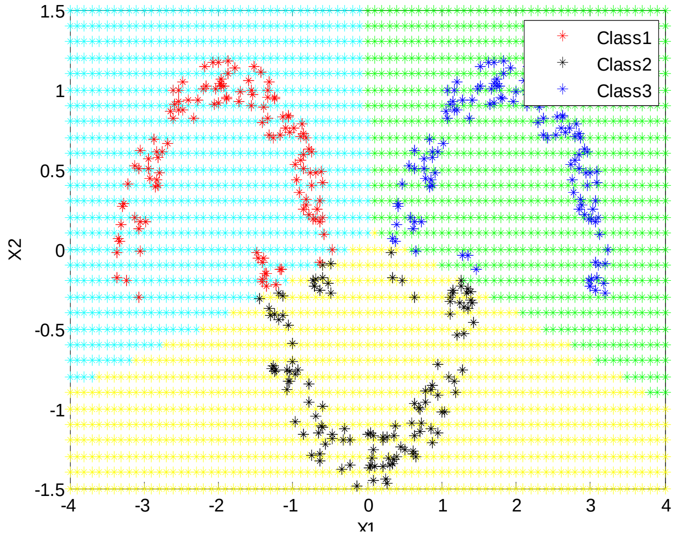


Baye's classifier:

$\sum_i = \sum$



$\sum$ = Different

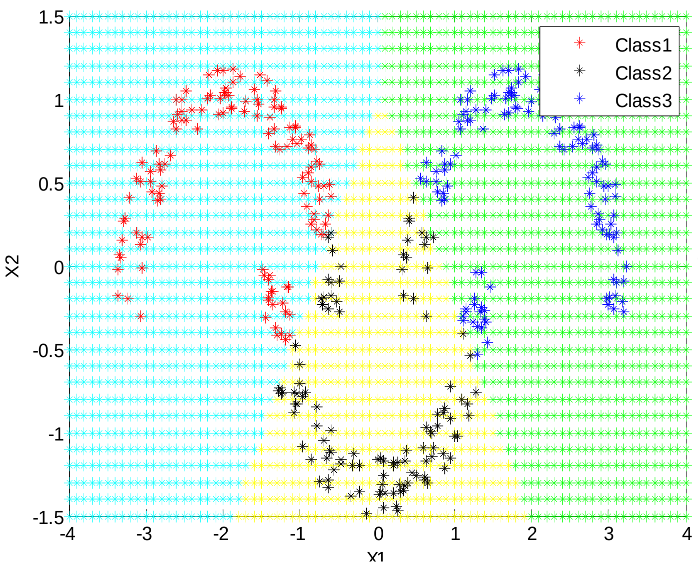


Performance of Bayes classifier:

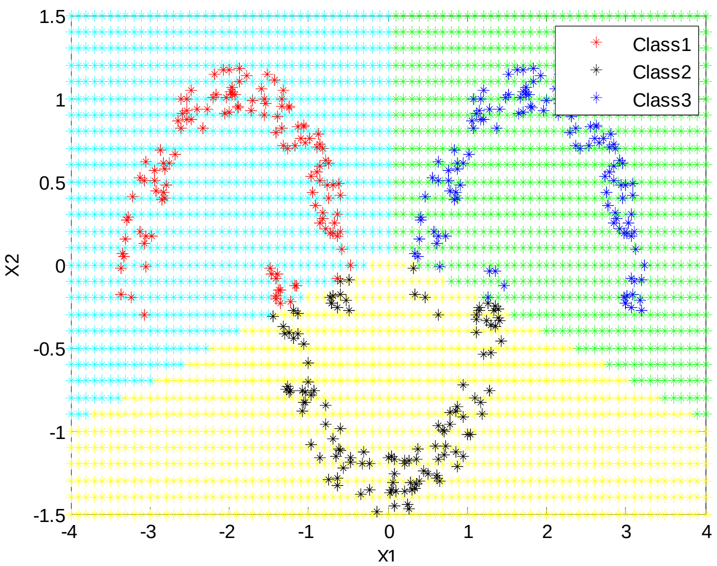
Covariance Matrix	$\sum_i = \sum$	$\sum$ = Different
Accuracy	92	92.2667
Class-wise Accuracy	92 87.2 96.8	92 88 96.8
Confusion Matrix	115 10 0 12 109 4 0 4 121	115 10 0 12 110 3 0 4 121

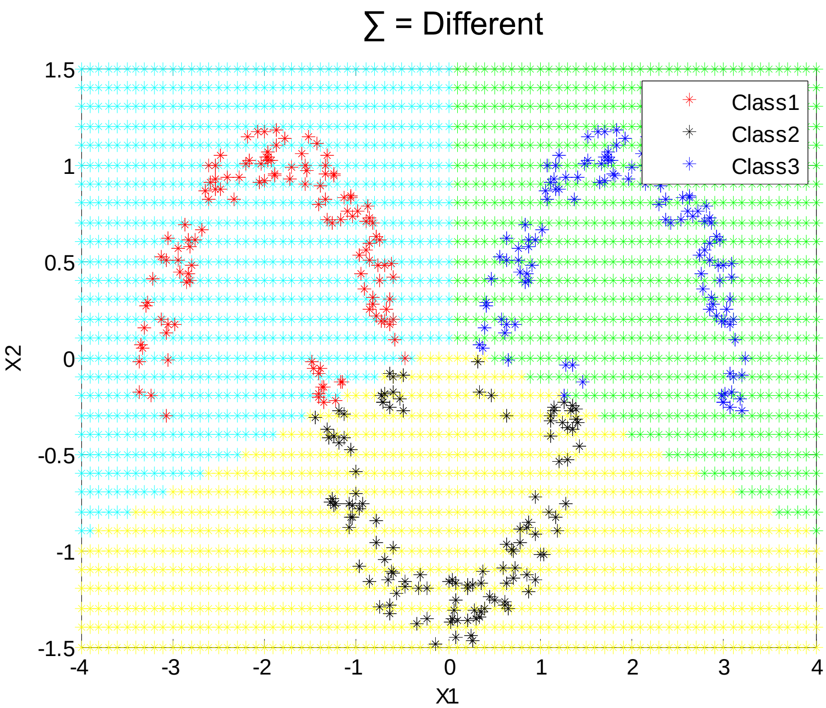
Naïve Baye's classifier

$\sum = \sigma^2 I$



$\sum_i = \sum$



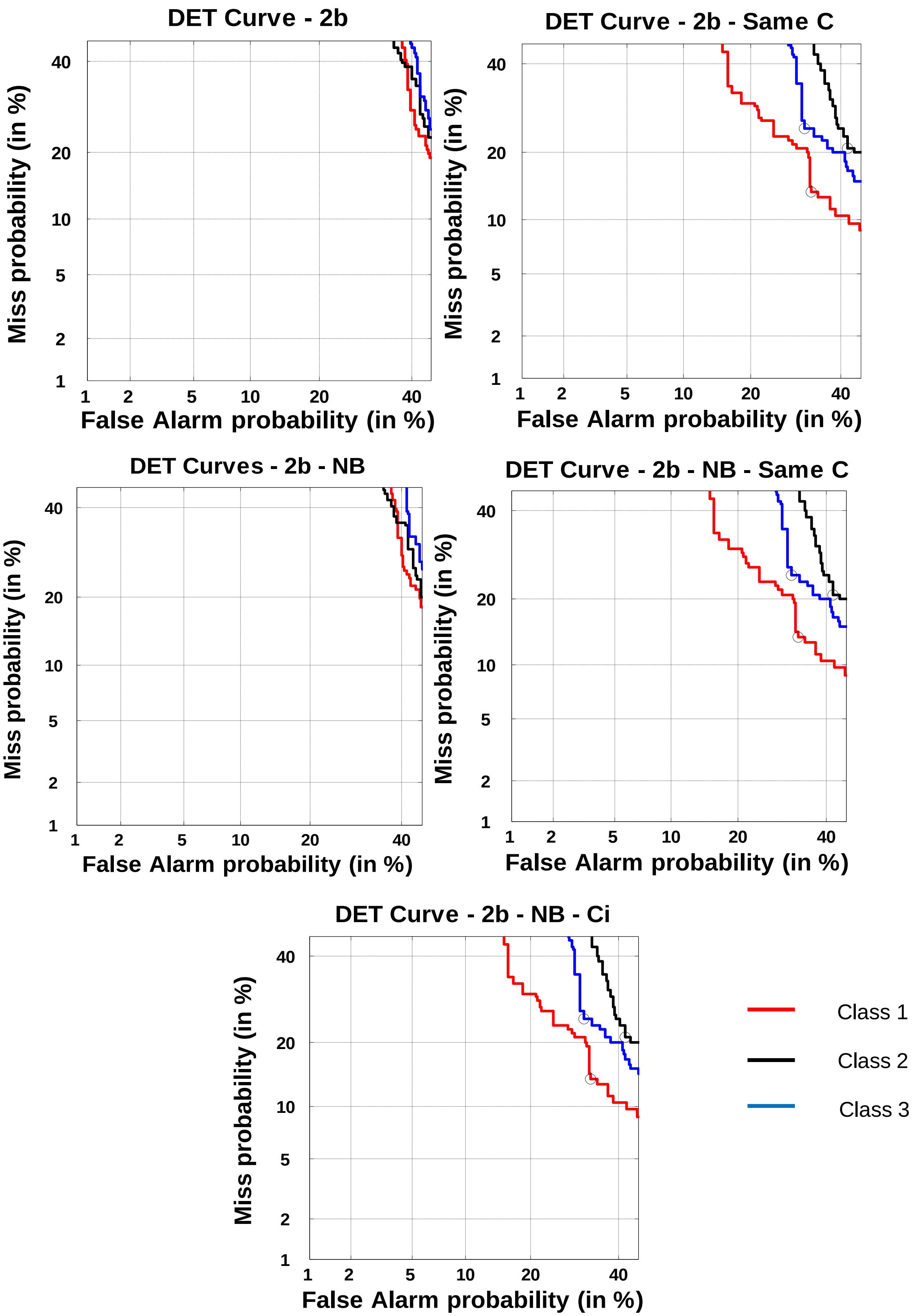


Performance of Naive Bayes classifier:

Covariance Matrix	$\sum = \sigma^2 I$	$\sum_i = \sum$	$\sum$ = Different
Accuracy	81.6	92	91.733
Class-wise Accuracy	88 68.8 88	92 87.2 96.8	91.2 87.2 96.8
Confusion Matrix	110 15 0 20 86 19 0 15 110	115 10 0 12 109 4 0 4 121	114 11 0 12 109 4 0 4 121

Observations:

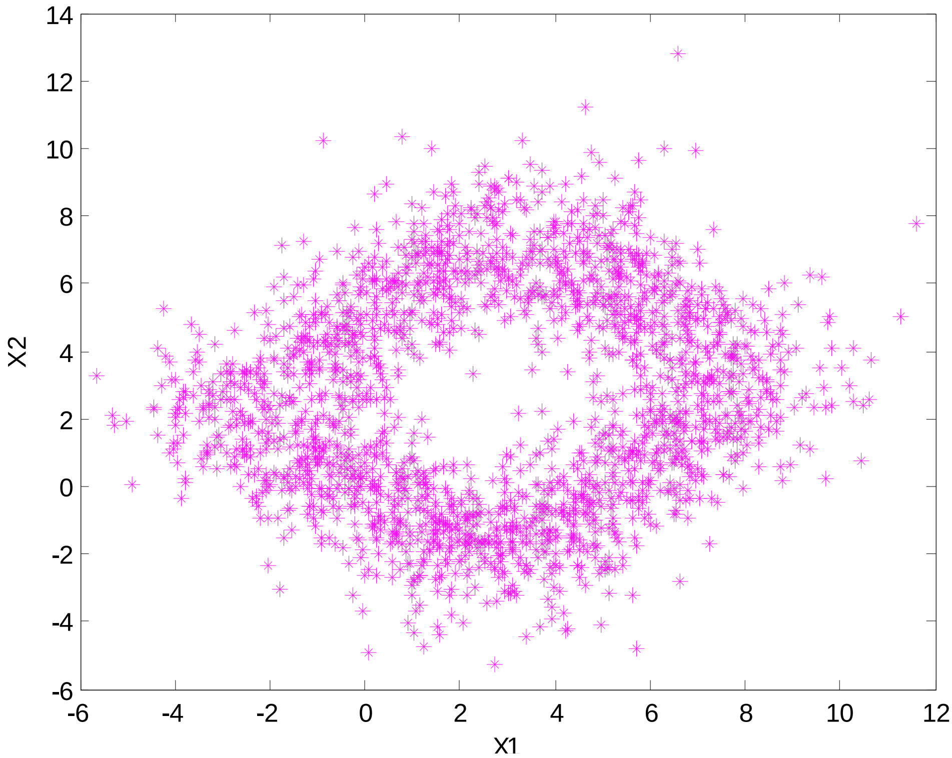
- In Bayes classifier the performance for same covariance and different covariance is almost the same as the covariance matrices for the 3 classes were almost the same
- In Naive Bayes classifier the accuracy when covariance matrix is considered to be $\sigma^2 I$ is reduced to a great extent because the difference between elements along the diagonal of the covariance matrix is large. When the root mean square value of these elements is considered as σ^2, the accuracy decreases.
- As the data is not overlapping, they are pretty well separated.



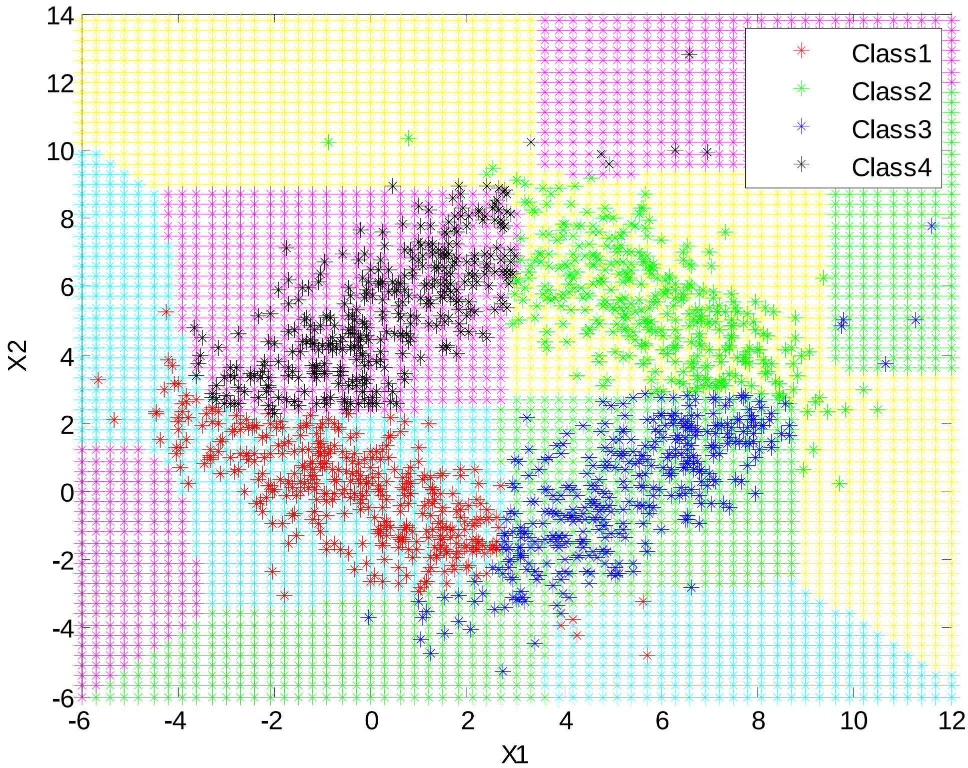
It can be seen that the performance increases when same covariance matrix is considered for all the classes as seen in the accuracies.

C) Overlapping classes

Train data before classification

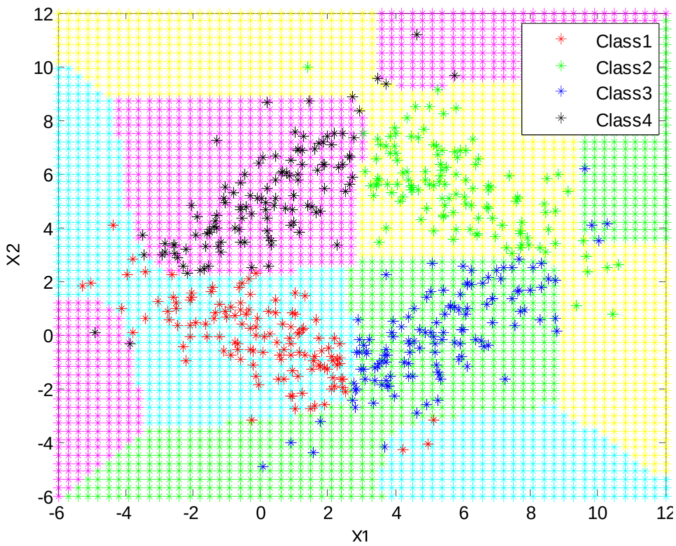


Train data after classification

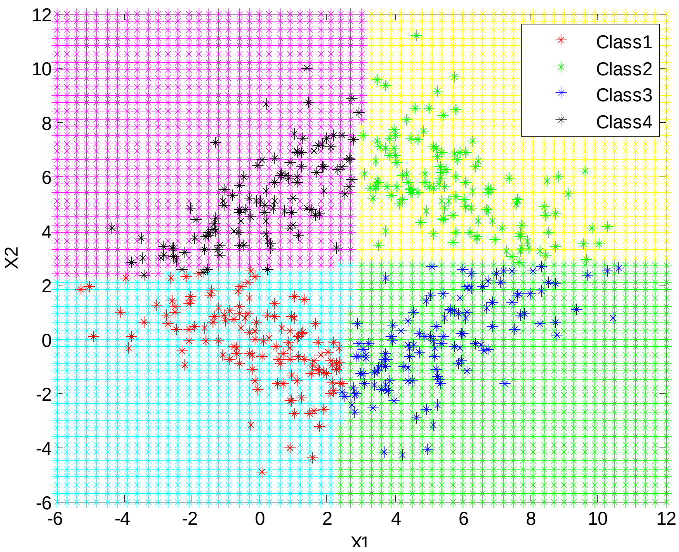


Bayes Classification

$\sum$ = Different



$\sum_i = \sum$

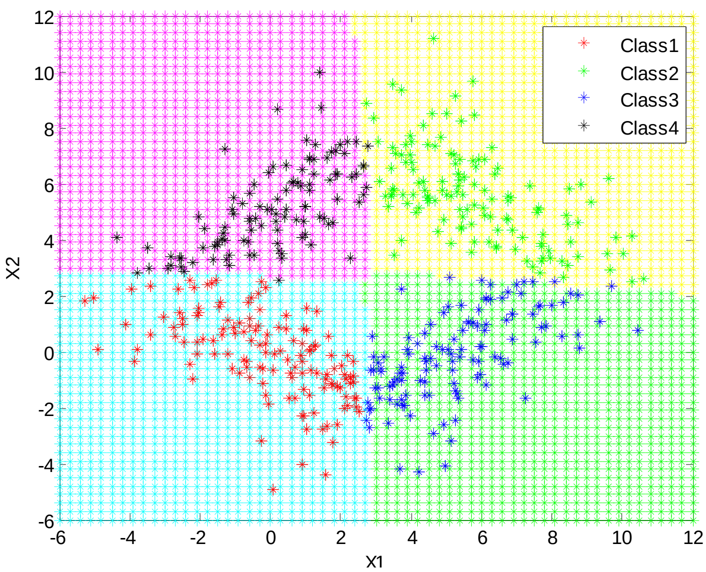


Performance of Bayes classifier:

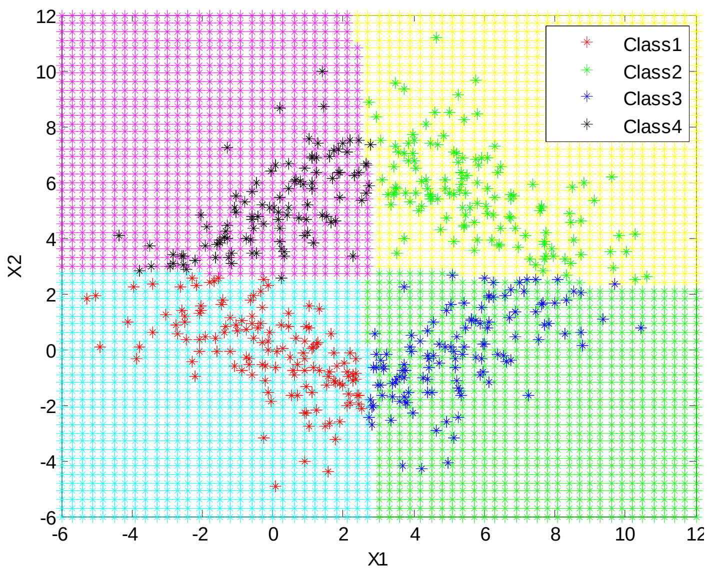
Covariance Matrix	$\sum_i = \sum$	$\sum$ = Different
Accuracy	84.4	84.4
Class-wise Accuracy	84.8 87.2 82.4 83.2	84.2 87.2 83.2 84
Confusion Matrix	106 0 17 2 0 109 9 7 14 8 103 0 9 12 0 104	104 0 16 5 0 109 8 8 14 7 104 0 10 10 0 105

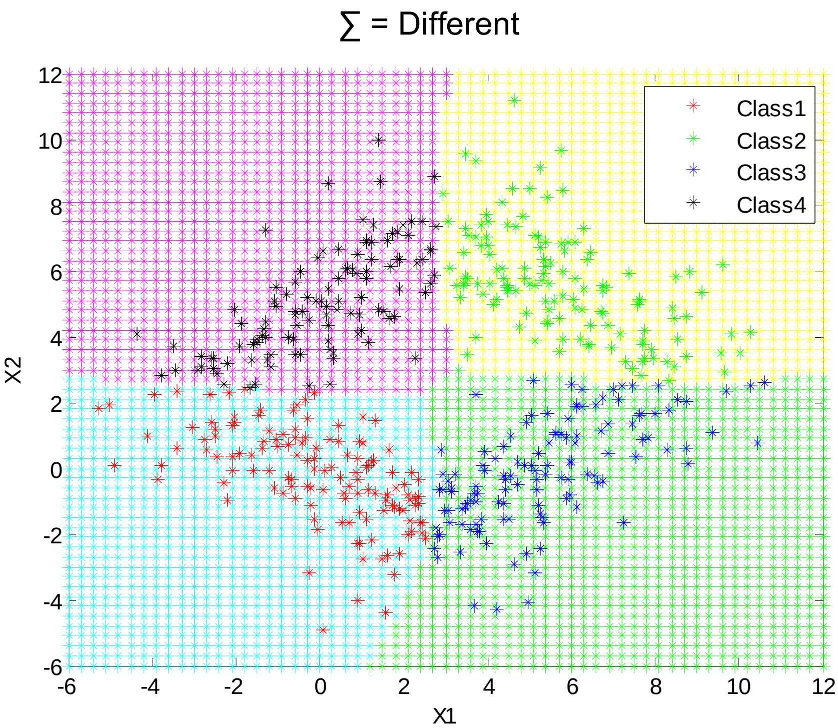
Naïve Baye's classifier

$\sum = \sigma^2 I$



$\sum_i = \sum$



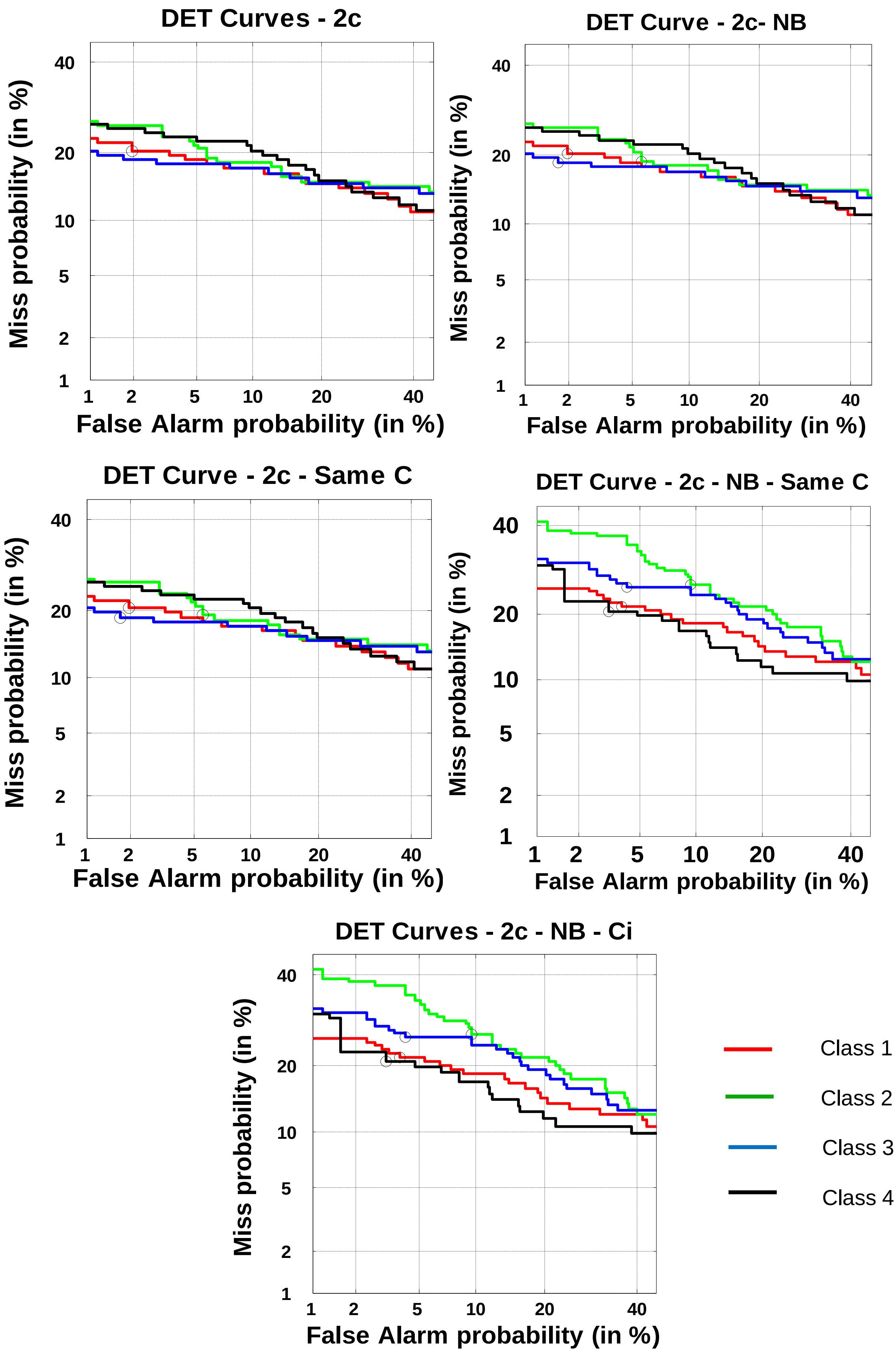


Performance of Naive Bayes classifier:

Covariance Matrix	$\Sigma = \sigma^2 I$	$\Sigma_i = \Sigma$	Σ = Different
Accuracy	83.3	83.8	83.6
Class-wise Accuracy	85.6 89.6 80.8 79.2	85.6 89.6 80.8 79.2	84 88 80.8 81.6
Confusion Matrix	107 0 17 1 0 112 6 7 16 8 101 0 12 14 0 99	107 0 17 1 0 112 6 7 16 8 101 0 12 14 0 99	105 0 17 3 0 110 8 7 16 8 101 0 10 13 0 102

Observations:

- As the data is overlapping, the accuracy has reduced compared to the previous two cases.
- In Bayes Classifier, though the class wise accuracies varied, the overall accuracy remained the same in both the cases of same covariance and different covariance.
- But the decision region is more complex when different covariance is considered.
- In Naive Bayes classifier, σ^2 was well approximated, therefore tha performance in all the three cases has been more or less the same.



Due to overlapping nature of the data, as seen in the accuracies, in all the cases the performance in similar.

Experimentation:

1. Changing the amount of training data

For Non-linear separable data 2(b) and for overlapping data 2(c) the amount of training data was varied and the corresponding accuracies were seen.

Percentage of Training Data (%)	Accuracy-2(b) (%)	Accuracy-2(c) (%)
50	92.26	83.4
60	92.33	83.5
75	92.27	84.4
80	93	85.25

It was observed that as the amount of training data increased, the performance of the system was getting better. This is because, more the training data the better the classifier.

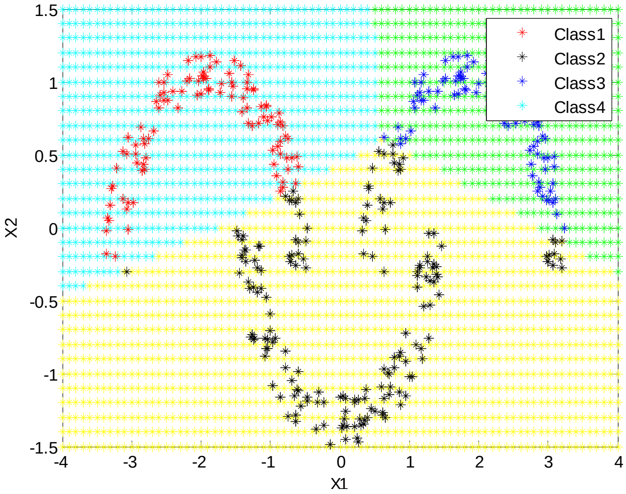
2. Varying the mean

A constant value was added to the mean of the classes during training. Because of this shift in mean, the classifier performance decreases.

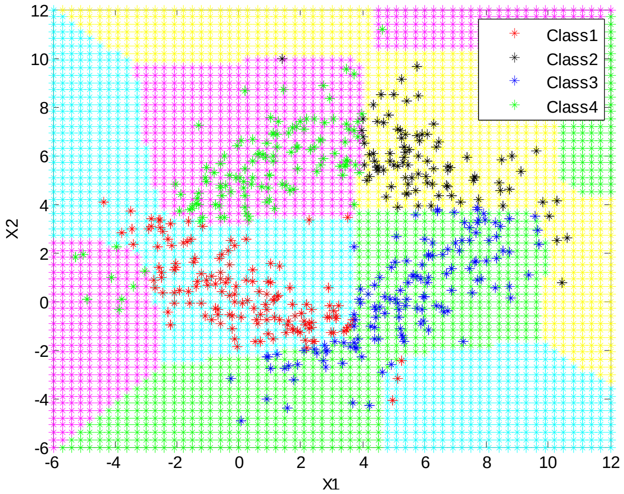
Value added to mean	Accuracy-2(b) (%)	Accuracy-2(c) (%)
0 0	92.27	84.4
0.5 0.5	85.3	82.8
-0.5 -0.5	88.3	82.6
1 1	66.4	64

Due to the shift in mean the decision boundary changes, which includes data points of other nearby classes.

2(b) with [0.5 0.5] added to mean



2(c) with [0.5 0.5] added to mean

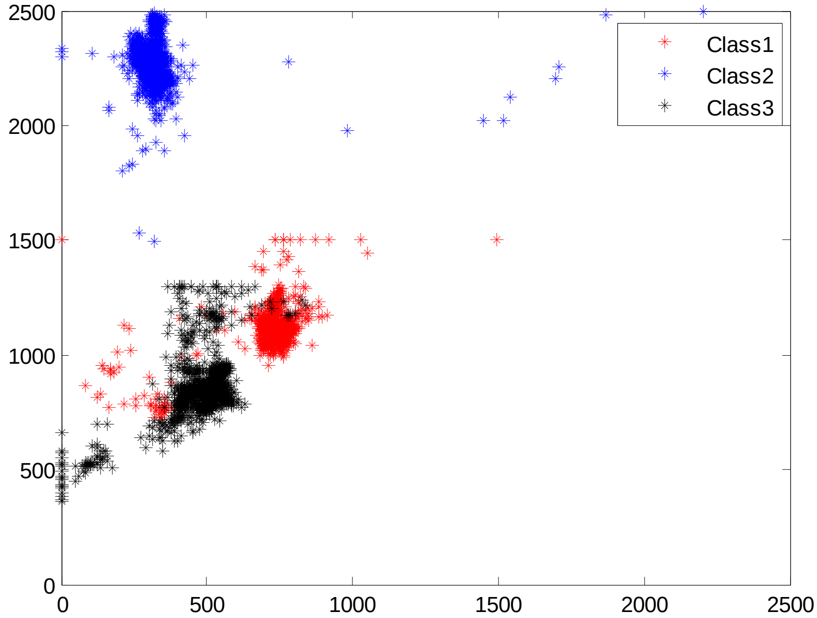


Dataset 3: Speech Data

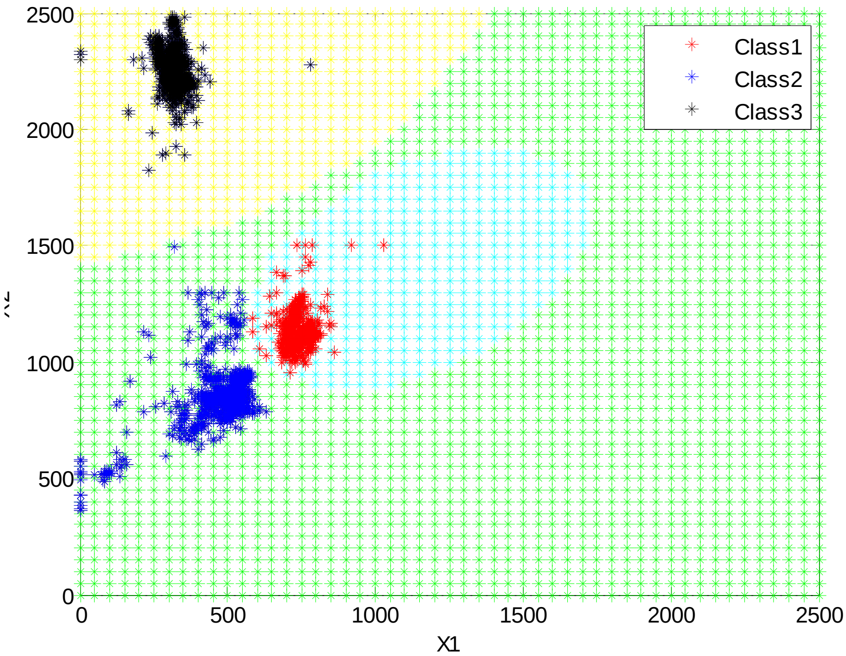
1. Data is read from the file and 75% of data is used as training set and remaining is used as testing set.
2. Mean vectors are calculated for corresponding classes.
3. For the case $\sum = \sigma^2 I$, σ^2 is the average of variance of all features
4. For $\sum_i = \sum$, $\sum$ is the covariance matrix is taken as the average of all covariance matrices
5. For arbitrary $\sum$, $\sum$ for each class is different and specific to the class.
6. In Naïve Baye's classifier, $\sum$ is a diagonal matrix and the off diagonal elements of $\sum$ matrix are made 0.

A) Speech Data of isolated utterances

Training data before classification:

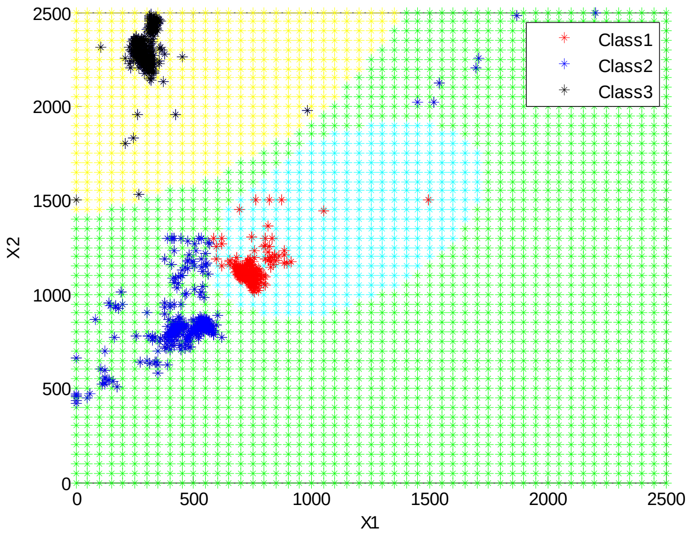


Training data after classification

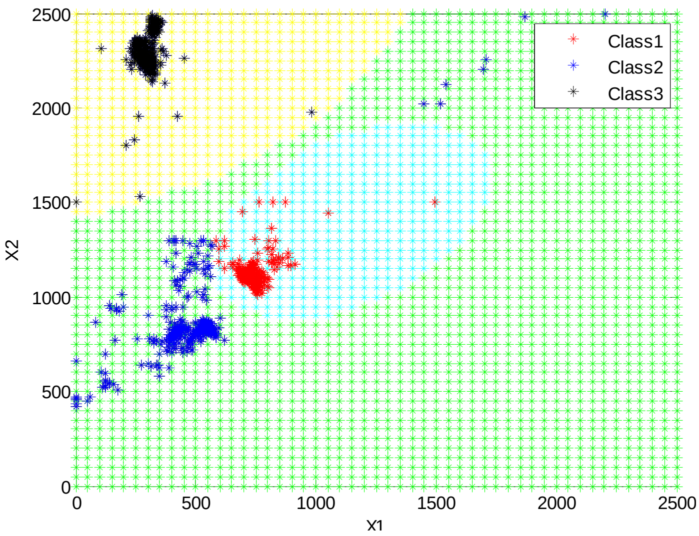


Bayes Classification

∑ = Different



$\sum_i = \sum$

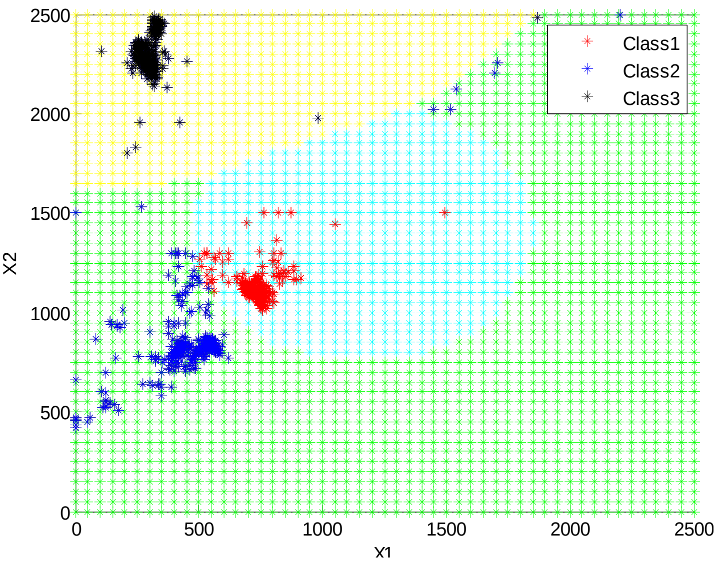


Performance of Bayes classifier:

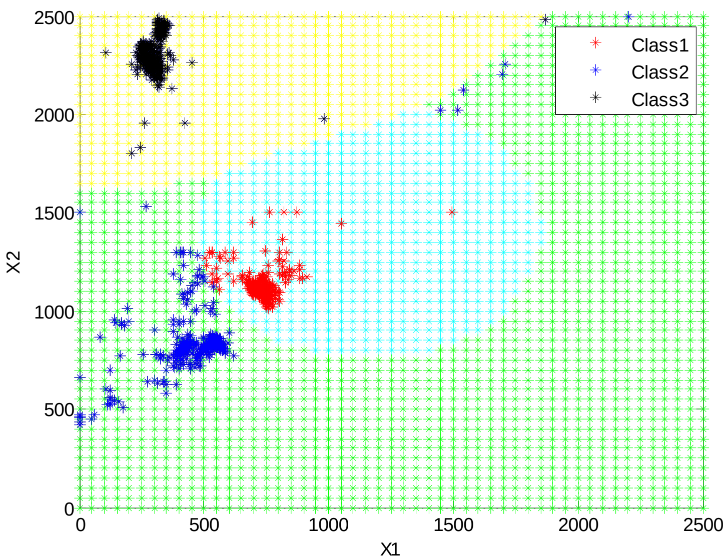
Covariance Matrix	$\sum_i = \sum$	$\sum$ = Different
Accuracy	95.8	97.2
Class-wise Accuracy	94.6 98.5 94.1	94.1 98.7 98.8
Confusion Matrix	512 1 28 8 533 0 32 0 509	509 1 31 0 534 7 6 0 535

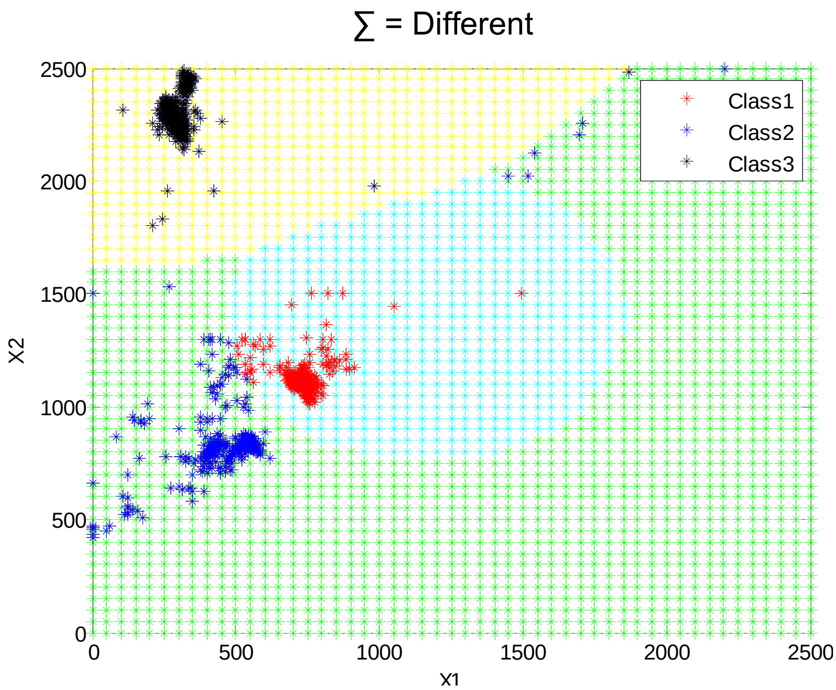
Naïve Baye's classifier

$\sum = \sigma^2 I$



$\sum_i = \sum$



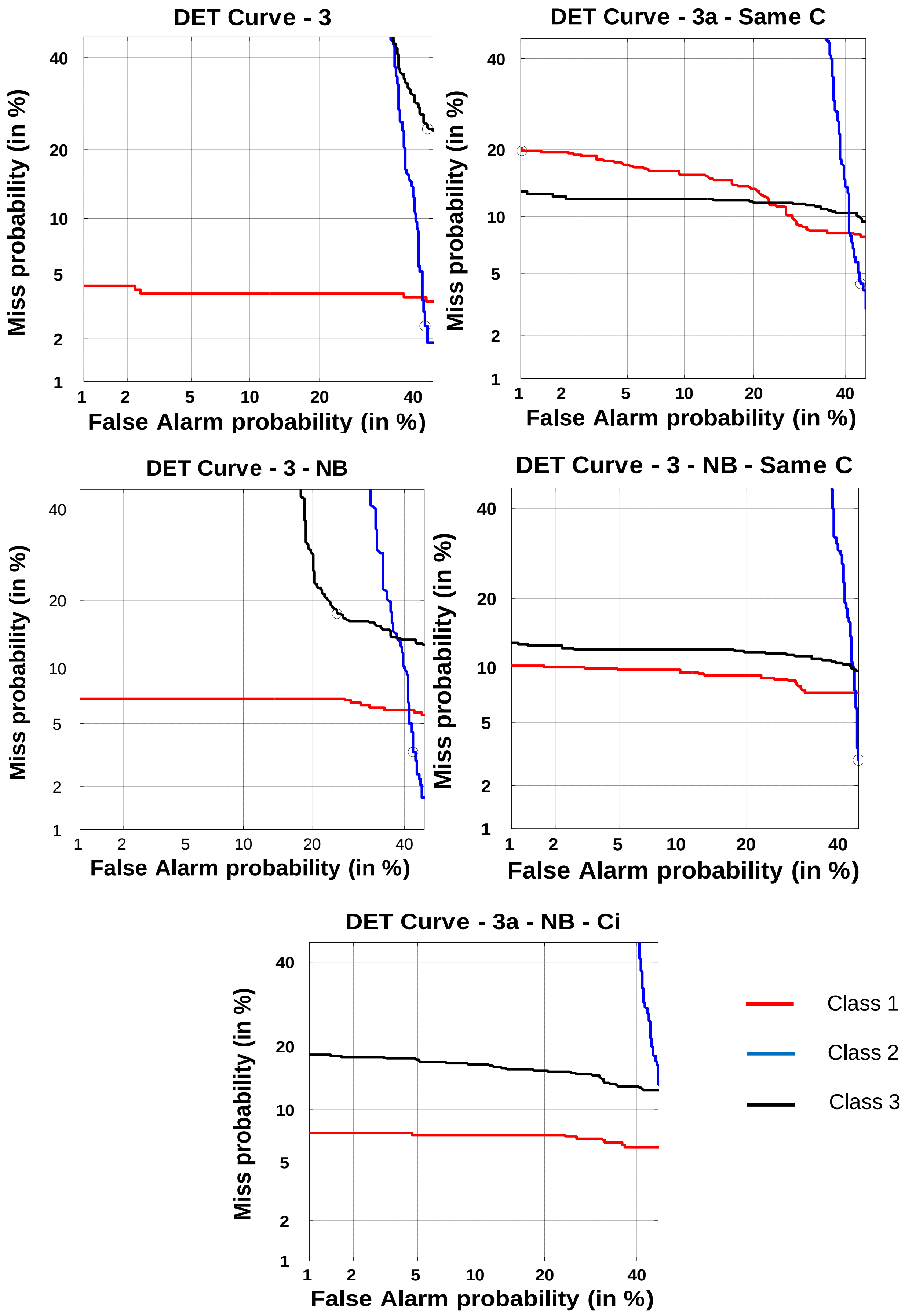


Performance of Naive Bayes classifier:

Covariance Matrix	$\Sigma = \sigma^2 I$	$\Sigma_i = \Sigma$	Σ = Different
Accuracy	95.87	95.74	96.4
Class-wise Accuracy	94.63 99.44 93.33	94.6 98.7 93.9	94.3 98.7 96.1
Confusion Matrix	512 0 29 3 538 0 35 0 506	512 1 28 7 534 7 33 0 509	510 0 31 0 534 7 21 0 520

Observations:

- In this real time data set, there are 3 classes with 2 dimensions.
- As seen from the plot of training data before classification class 1 and class 3 are overlapping, therefore they cannot be 100% separated.
- But class 2 is well separated from class 1 and class 3, therefore is all cases it is noticed that the accuracy for class 2 is higher.
- Different Covariance matrix for each class gave a better performance in both Bayes and Naive Bayes Classifier.



In Bayes classifier, the increase in class 3 accuracy is reflected in the DET plot.

Dataset 3b : 39-dimensional speech data corresponding to the speech emotions

Each file had a 39 dimensional feature vector. Files for training and testing were given separately. Using Naive Bayes Classifier the accuracy and confusion matrix for different cases were:

Covariance Matrix	$\Sigma = \sigma^2 I$	$\Sigma_i = \Sigma$	Σ = Different
Accuracy	38.1	91.8	96.4
Class-wise Accuracy	14.8 14.6 8.45	29.5 20.8 41.5	54.5 67.7 72.6
Confusion Matrix	353 701 1324 451 349 1578 919 1258 201	702 502 1174 696 495 118 849 541 988	1296 731 351 626 1612 140 386 264 728

Experimentation:

1. For the case $\Sigma = \sigma^4 I$

Confusion Matrix:

874	220	1284
680	603	1095
46	62	2270

Accuracy: 36.7536 25.3574 95.4584

2. Adding very small number (random number multiplied by 0.1) to the full covariance matrix

(i) Confusion Matrix:

1381	638	359
790	1458	130
323	308	1747

Accuracy: 58.0740 61.3120 73.4651

(ii) Confusion Matrix:

1921	0	457
2048	0	330
437	0	1941

Accuracy: 80.7822 0 81.6232

(iii) Confusion Matrix:

1123	977	278
479	1853	46
537	342	1499

Accuracy: 47.2246 77.9226 63.0362

(iv) Confusion Matrix:

0	1767	611
0	2149	229
0	252	2126

Accuracy: 0 90.3701 89.4029

(v) Confusion Matrix:

1461	570	347
970	1246	162
322	436	1620

Accuracy: 61.4382 52.3970 68.1245

3. Performing k-means clustering and then doing Bayesian estimation

Confusion Matrix:

1415	603	360
612	1686	80
400	182	1796

Accuracy: 59.5038 70.8999 75.5257

Observations:

- It is observed that the overall accuracy of real world data is less compared to the other data.
- Better accuracy is observed when covariance matrix of each class is different.
- From the experiments it is observed that, high accuracy is obtained when nothing gets classified into a class. But this is not desirable.
- Performing k-means algorithm and then doing the Bayesian Estimation improves the accuracy of classification.