CS6690: Pattern Recognition Assignment #2

Group 3: Akshay & Suthirth

September 26, 2014

1 Bayesian Classifiers

According to Bayes Theorem, for a dataset x with classes ω_i,
Probability of a datapoint belonging to class ω_i is defined as:

$$P(\omega_i|x) = \frac{(P(x|\omega_i)P((\omega_i))}{P(x)} \tag{1}$$

- Here, $P(x|\omega_i)$ is known as the class likelihood.
  To estimate this value, we require the distribution of ω_i. Based on the central limit theorem, we can assume that this would be Gaussian distribution for large datasets.

- The value $P(\omega_i)$ is the class prior and is calculated using:

$$P(\omega_i) = N_i/N \tag{2}$$

  This term becomes irrelevant if the classes have equal probabilities.

- P(x) is termed as 'evidence' and can be calculated as:

$$P(x) = \sum_i P(x|\omega_i)P(\omega_i) \tag{3}$$

2 Gaussian Likelihood Distribution

For multi-dimensional data, the Gaussian Distribution is:

$$P(x;\mu,\Sigma) = \frac{1}{2\pi^{k/2}|\Sigma|^{1/2}} e^{-(x-\mu)^T \Sigma^{-1}(x-\mu)} \tag{4}$$

where

- μ is the mean

- Σ is the covariance matrix

The above parameters are calculated for the following cases:

3 Data

Red - Class 1, Green - Class 2, Blue - Class 3, Cyan - Class 4
Black (solid) - 1st Eigenvector
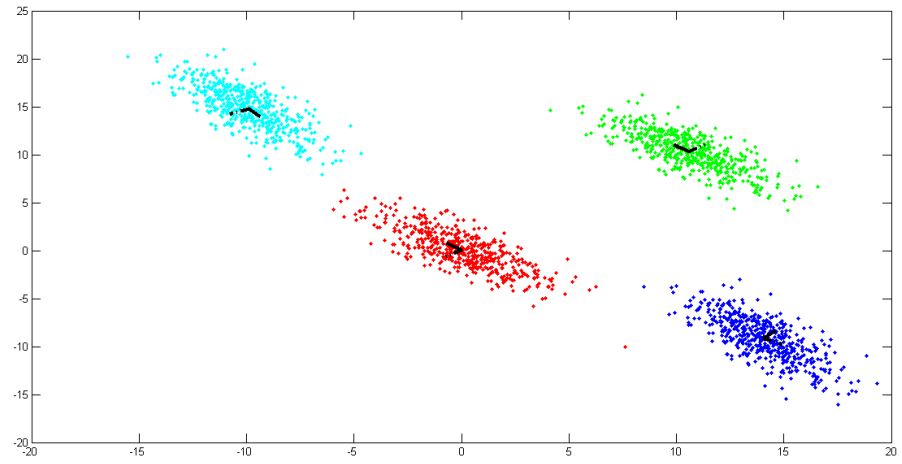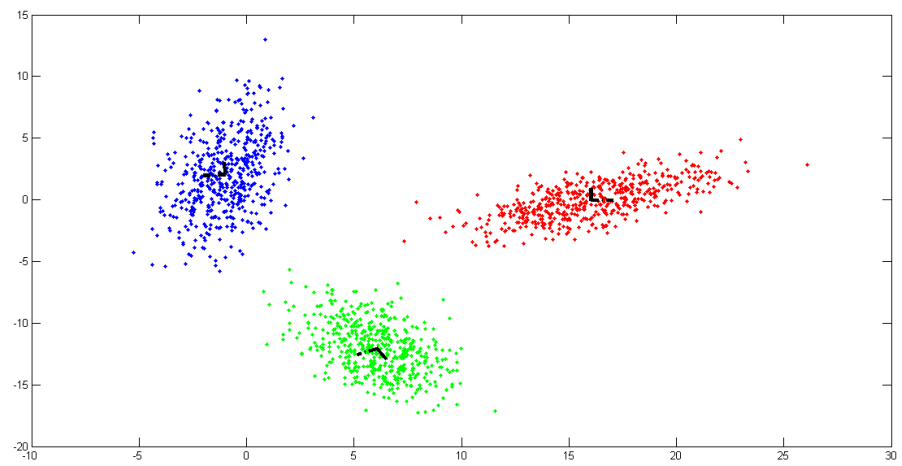Black (dashed) - 2nd Eigenvector

Figure 1: Linearly Separable Data
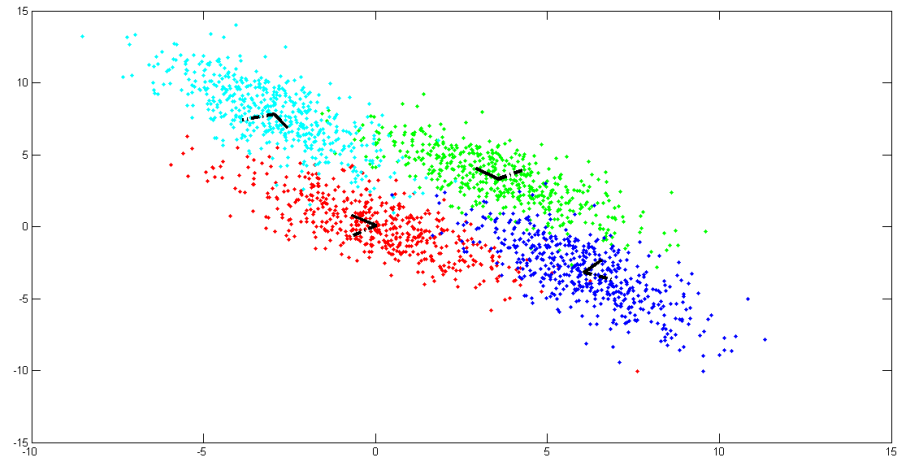

Figure 2: Non-linearly Separable Data
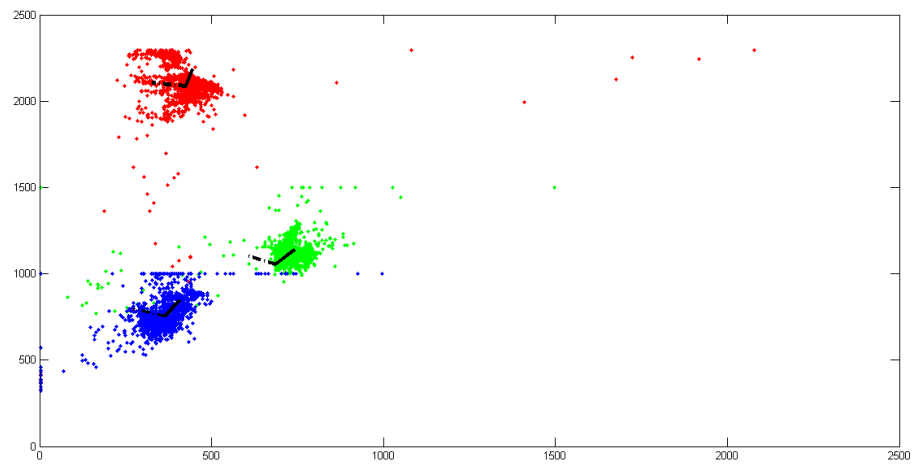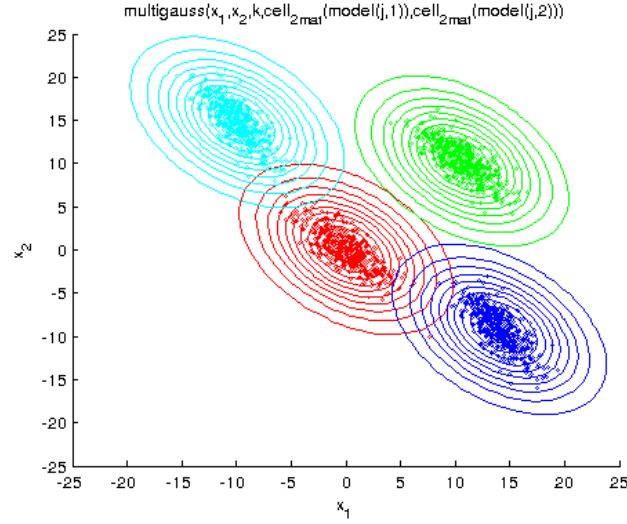
Figure 3: Overlapping Data

Figure 4: Real World Data: Vowel utterance formant frequencies F1 and F2

3.1 Bayes Classifier with Covariance same for all classes

multigauss(x$_1$,x$_2$,k,cell$_{2\text{ma1}}$(model(j,1)),cell$_{2\text{ma1}}$(model(j,2)))

For a bayesian classifier with a common covariance matrix, we combine the matrices by taking the mean of all the class covariance matrices. Geometrically this will result in a linear rotation of the contours, while at the same time, scaling the axes of the contours.

3.2 Bayes Classifier with Covariance different for all classes

multigauss(x$_1$,x$_2$,k,cell$_{2\text{ma1}}$(model(j,1)),cell$_{2\text{ma1}}$(model(j,2)))
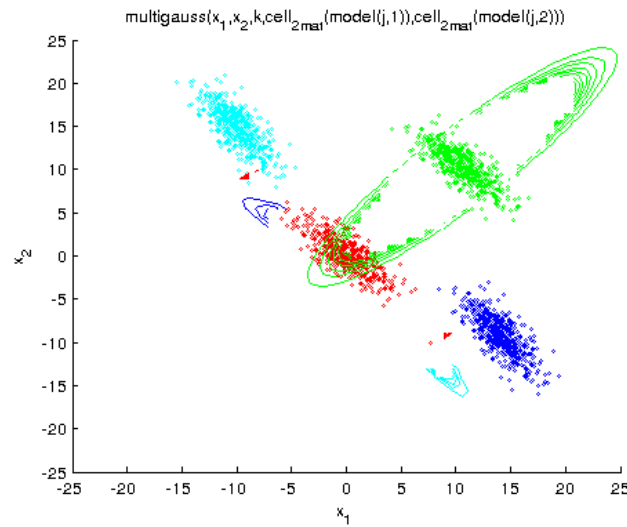
In this case, we take the class matrices and find out standard covariance to get the covariance matrix of each class. This acts as the C in our formula

3.3 Naive Bayes Classifier with $C = \sigma^2 * I$

multigauss(x$_1$,x$_2$,k,cell$_{2mat}$(model(j,1)),cell$_{2mat}$(model(j,2)))
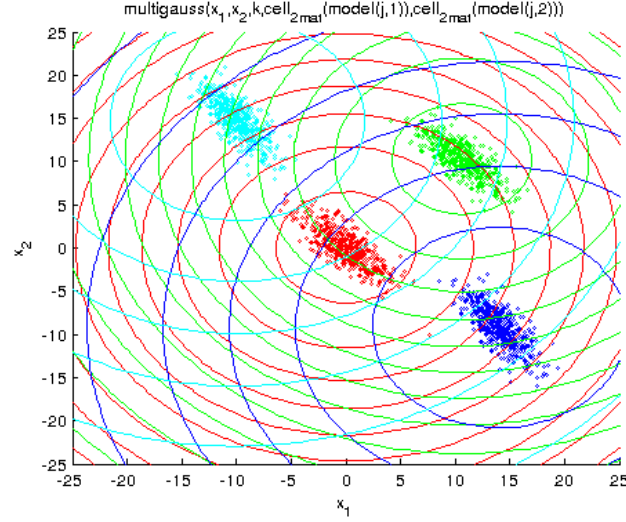
To get one single value of σ, we use a weighted mean for all the features of each class and come up with a common $\sigma_i I$ matrix; We then take a weighted mean of these matrices to come up with one single $\sigma^2 I$ matrix

3.4 Naive Bayes Classifier with C same for all classes
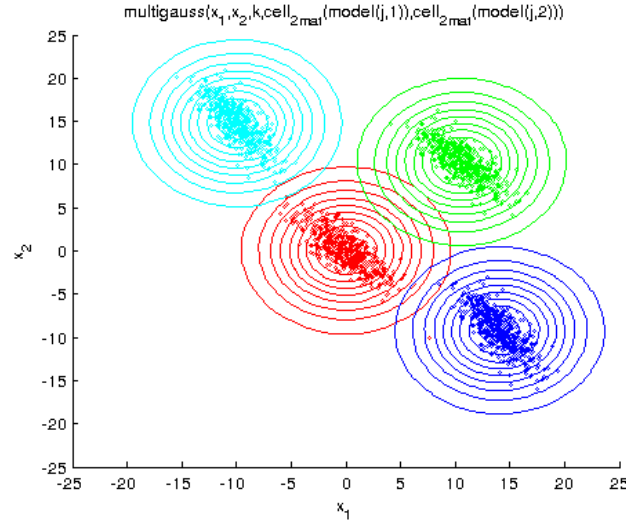
multigauss(x$_1$,x$_2$,k,cell$_{2mat}$(model(j,1)),cell$_{2mat}$(model(j,2)))

We follow the same principle as the non-naive case, but with features being independent of each other, thus setting off diagonal elements to 0

3.5 Naive Bayes Classifier with C different for all classes
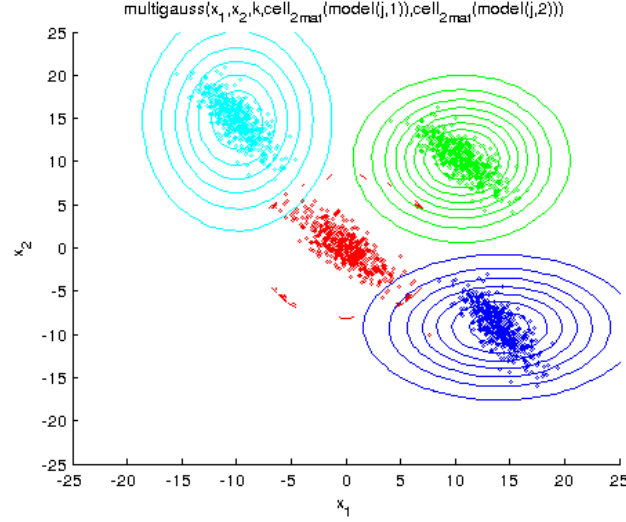
multigauss(x_1,x_2,k,cell$_{2ma1}$(model(j,1)),cell$_{2ma1}$(model(j,2)))

We follow the same principle as the non-naive case, but with features being independent of each other, thus setting off diagonal elements to 0

4 Bayes Classification

If $P(\omega_1|x) > P(\omega_2|x)$ then x belongs to class ω_1
If $P(\omega_1|x) < P(\omega_2|x)$ then x belongs to class ω_2

Using equation (1), this can be written as:

$$P(x|\omega_1)P(\omega_1) \gtrless P(x|\omega_2)P(\omega_2) \tag{5}$$

The value x_0 at which the RHS and LHS is called the threshold value.

This classification rules minimizes number of misclassifications.

Based on this classification, the following quantities can be defined:

$$TruePositiveRate = \frac{\sum TruePositive}{\sum ConditionPositive} \tag{6}$$

$$FalsePositiveRate = \frac{\sum FalsePositive}{\sum ConditionNegative} \tag{7}$$

A plot of these two quantities is known as a DET Curve.

5 Experiments

5.1 Decision Boundaries

Following plots describe the decision boundaries for various datasets with different Bayesian classifiers.

For every figure,
Row 1: Bayes with: (L) Same covariance for all classes, (R) Different covariance for all classes
Row 2: Naive Bayes with: (L) $C = \sigma^2 * I$, (R) Same C for all classes
Row 3: Naive Bayes with different C for all classes

Legends:

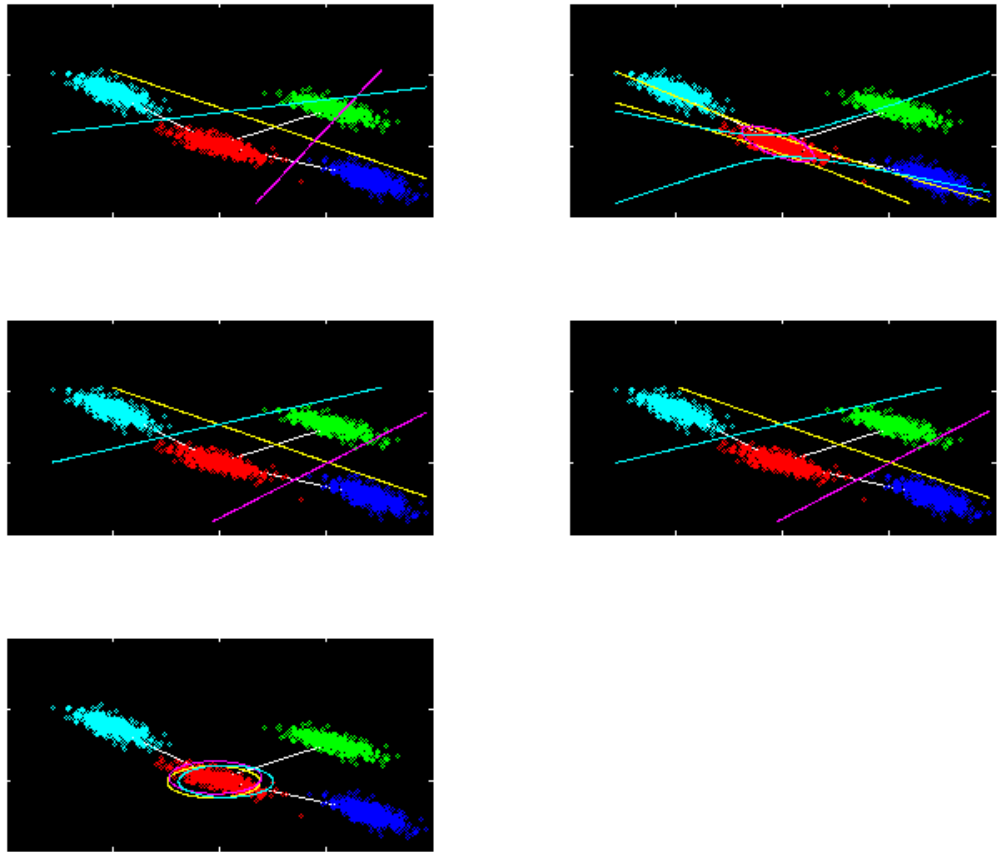|	
Red	Class 1
Green	Class 2
Blue	Class 3
Cyan	Class 4
White	$\mu_1 - \mu_2, \mu_3, \mu_4$
Yellow	Decision Boundary b/w Class 1 and 2
Magenta	Decision Boundary b/w Class 1 and 3
Cyan (line)	Decision Boundary b/2 Class 1 and 4

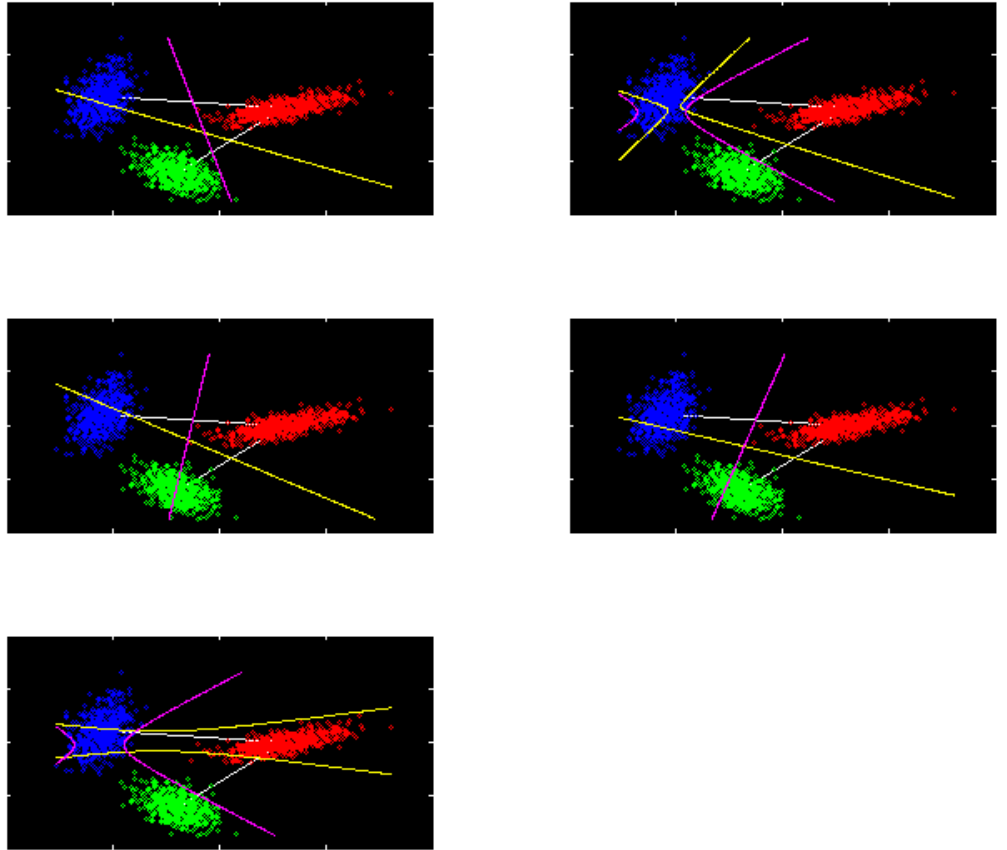Figure 5: Decision Boundary for Dataset 1
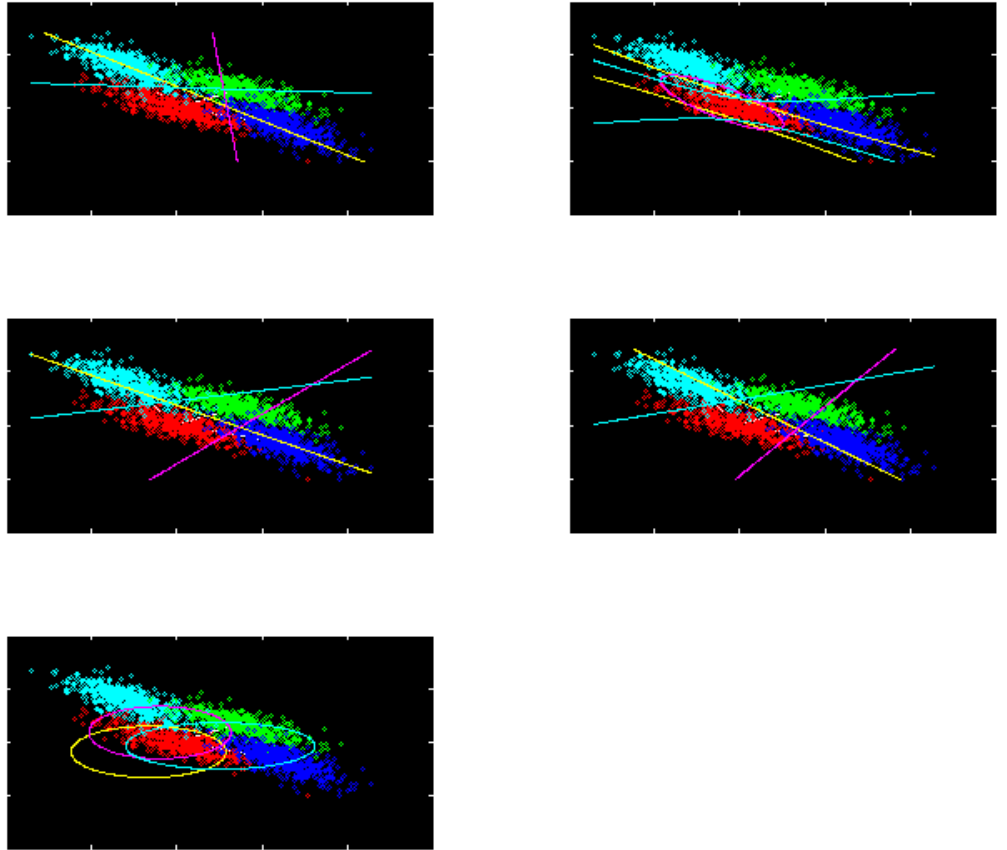
Figure 6: Decision Boundary for Dataset 2

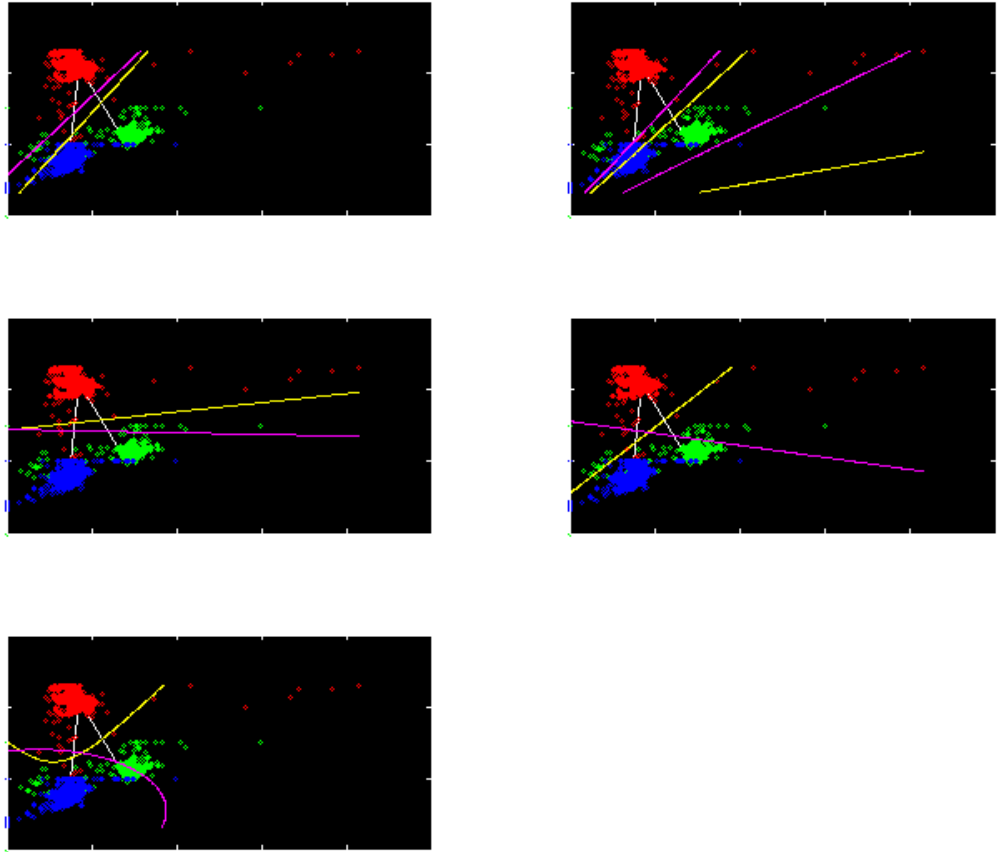Figure 7: Decision Boundary for Dataset 3

Figure 8: Decision Boundary for Dataset 4

6 Confusion Matrix

6.1 Bayes with same covariance for all classes

$$LS = \begin{bmatrix} 100 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 100 \end{bmatrix}$$

$$NLS = \begin{bmatrix} 100 & 0 & 0 \\ 0 & 100 & 2 \\ 0 & 0 & 98 \end{bmatrix}$$

$$OD = \begin{bmatrix} 91 & 0 & 1 & 2 \\ 0 & 92 & 13 & 1 \\ 3 & 2 & 86 & 0 \\ 6 & 6 & 0 & 97 \end{bmatrix}$$

6.2 Bayes with different covariance for all classes

$$LS = \begin{bmatrix} 71 & 0 & 0 & 0 \\ 10 & 100 & 0 & 0 \\ 8 & 0 & 100 & 0 \\ 11 & 0 & 0 & 100 \end{bmatrix}$$

$$NLS = \begin{bmatrix} 100 & 0 & 1 \\ 0 & 100 & 1 \\ 0 & 0 & 98 \end{bmatrix}$$

$$OD = \begin{bmatrix} 64 & 0 & 0 & 0 \\ 12 & 95 & 13 & 4 \\ 14 & 2 & 87 & 0 \\ 10 & 3 & 0 & 96 \end{bmatrix}$$

6.3 Naive Bayes with $C = \sigma^2 * I$

$$LS = \begin{bmatrix} 100 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 100 \end{bmatrix}$$

$$NLS = \begin{bmatrix} 100 & 0 & 0 \\ 0 & 100 & 0 \\ 0 & 0 & 100 \end{bmatrix}$$

$$OD = \begin{bmatrix} 86 & 0 & 1 & 2 \\ 0 & 86 & 14 & 4 \\ 6 & 7 & 85 & 0 \\ 8 & 7 & 0 & 94 \end{bmatrix}$$

6.4 Naive Bayes with C same for all

$$LS = \begin{bmatrix} 100 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 100 \end{bmatrix}$$

$$NLS = \begin{bmatrix} 100 & 0 & 0 \\ 0 & 100 & 0 \\ 0 & 0 & 100 \end{bmatrix}$$

$$OD = \begin{bmatrix} 86 & 0 & 1 & 4 \\ 0 & 87 & 14 & 2 \\ 6 & 7 & 85 & 0 \\ 8 & 6 & 0 & 94 \end{bmatrix}$$

6.5 Naive Bayes with C different for all

$$LS = \begin{bmatrix} 84 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 6 & 0 & 100 & 0 \\ 10 & 0 & 0 & 100 \end{bmatrix}$$

$$NLS = \begin{bmatrix} 95 & 0 & 1 \\ 5 & 100 & 3 \\ 0 & 0 & 96 \end{bmatrix}$$

$$OD = \begin{bmatrix} 66 & 0 & 0 & 0 \\ 2 & 89 & 14 & 4 \\ 14 & 5 & 86 & 0 \\ 18 & 6 & 0 & 96 \end{bmatrix}$$
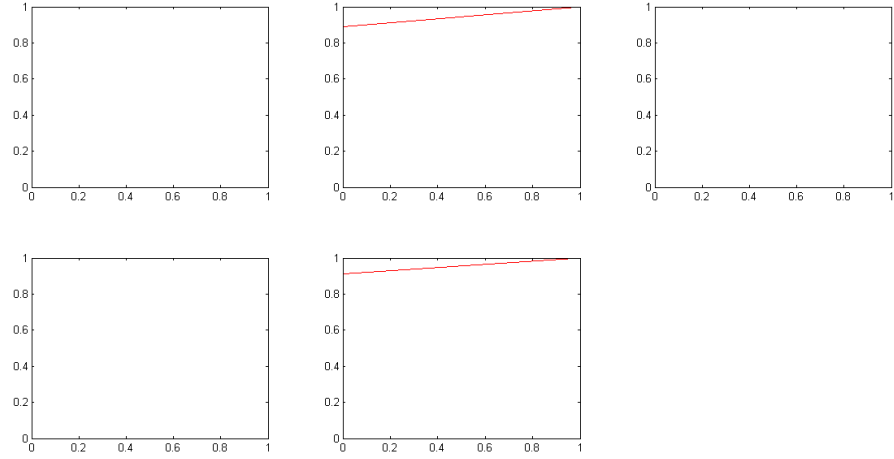
7 DET Curves

DET Curves for Class 1.

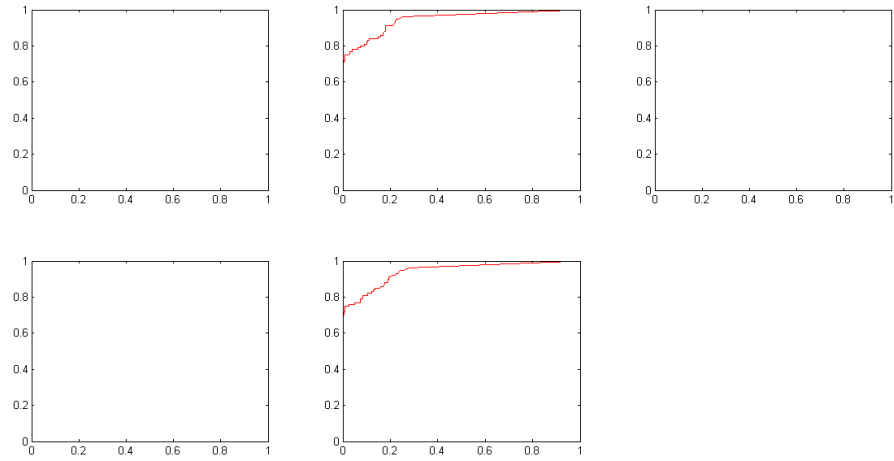Figure 9: DET Curves for Linearly Separable Data
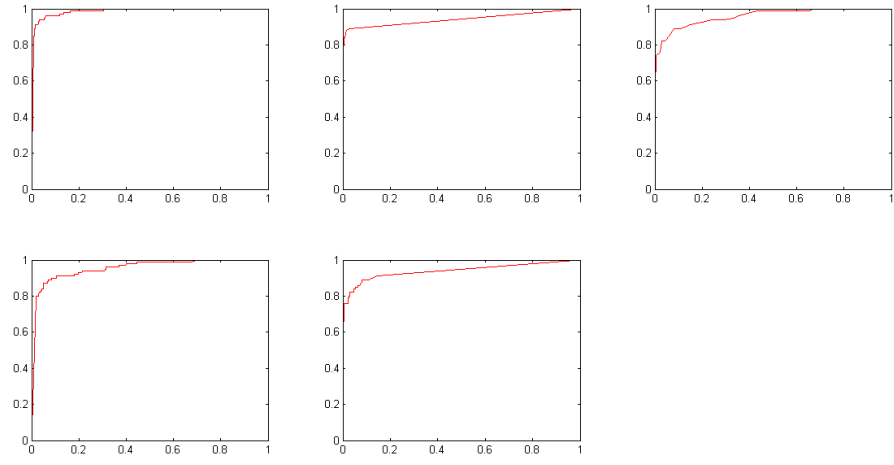

Figure 10: DET Curves for Non-linearly Separable Data

15

Figure 11: DET Curves for Overlapping Data

16