CS6690: Pattern Recognition Assignment #2

Group 3: Akshay & Suthirth

September 25, 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:

2.1 Bayes Classifier with Covariance same for all classes

2.2 Bayes Classifier with Covariance different for all classes

2.3 Naive Bayes Classifier with C = $\Sigma^2 * I$

2.4 Naive Bayes Classifier with C same for all classes

2.5 Naive Bayes Classifier with C different for all classes

3 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:

$TruePositiveRate = \frac{\Sigma TruePositive}{\Sigma ConditionPositive}$

$FalsePositiveRate = \frac{\Sigma FalsePositive}{\Sigma ConditionNegative}$

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

4 Experiments

4.1 Data

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

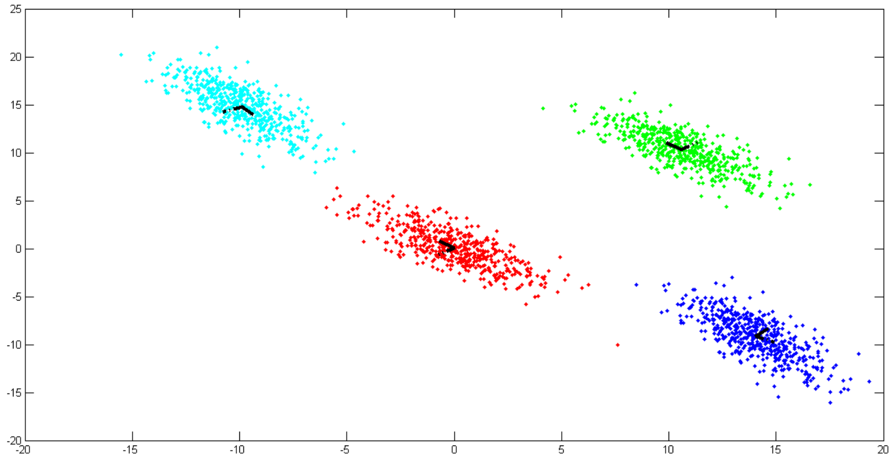

Figure 1: Linearly Separable Data

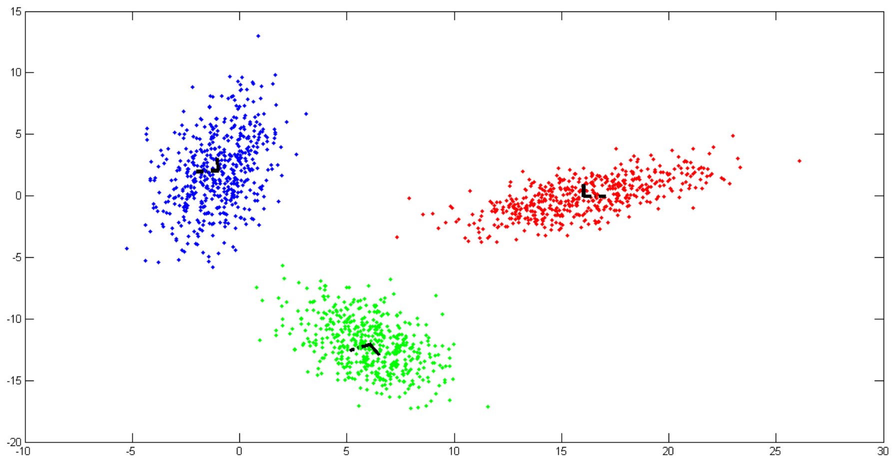

Figure 2: Non-linearly Separable Data

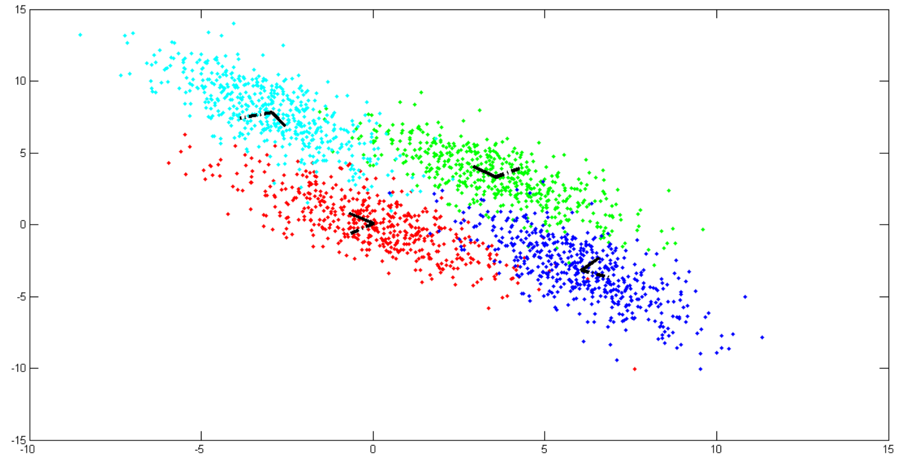

Figure 3: Overlapping Data

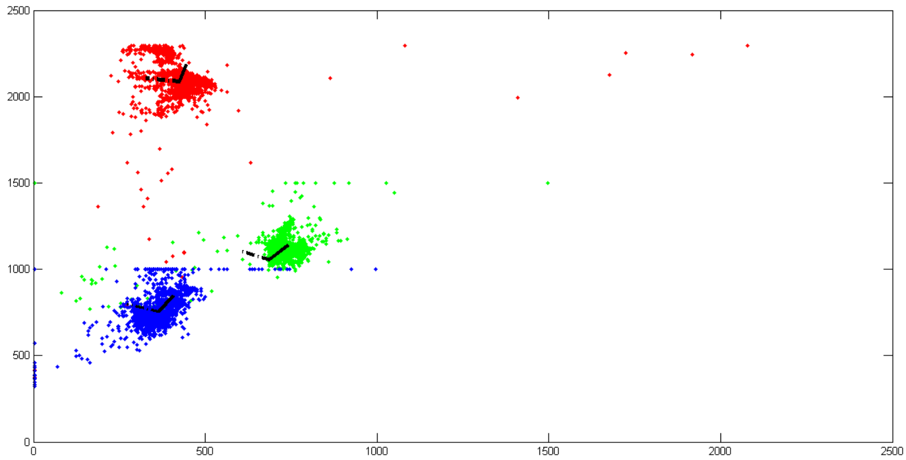

Figure 4: Real World Data: Vowel utterance formant frequencies F1 and F2

4.2 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

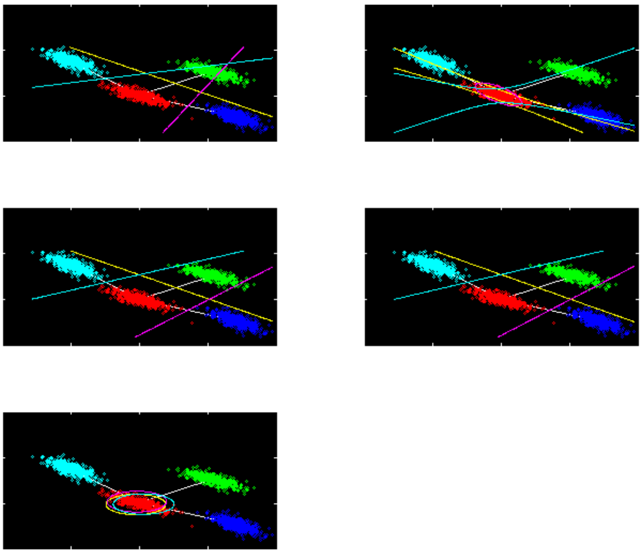

Figure 5: Decision Boundary for Dataset 1

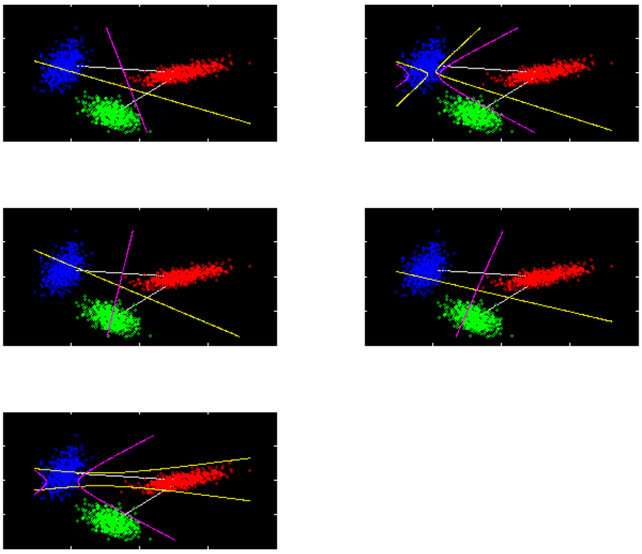

Figure 6: Decision Boundary for Dataset 2

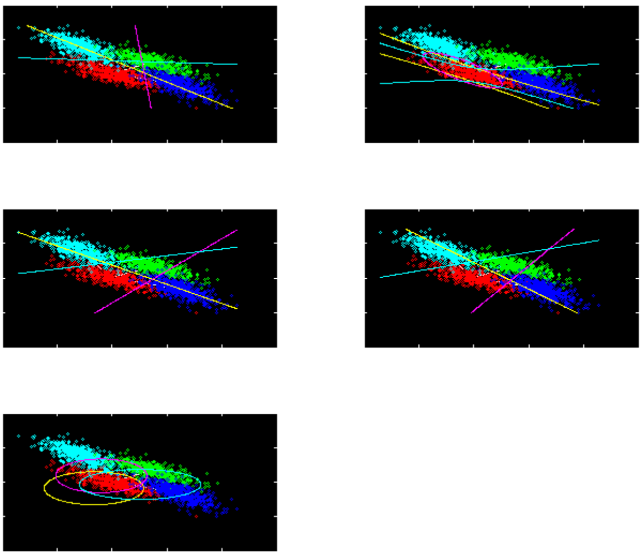

Figure 7: Decision Boundary for Dataset 3

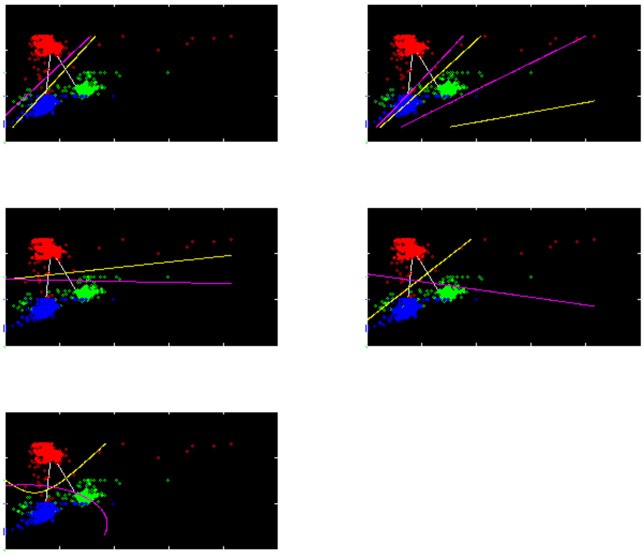

Figure 8: Decision Boundary for Dataset 4

4.3 DET Curves

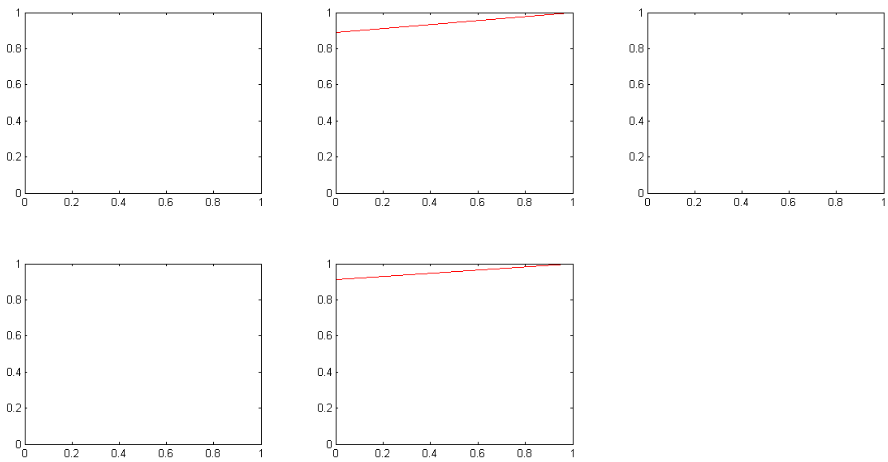

Figure 9: DET Curves for Linearly Separable Data

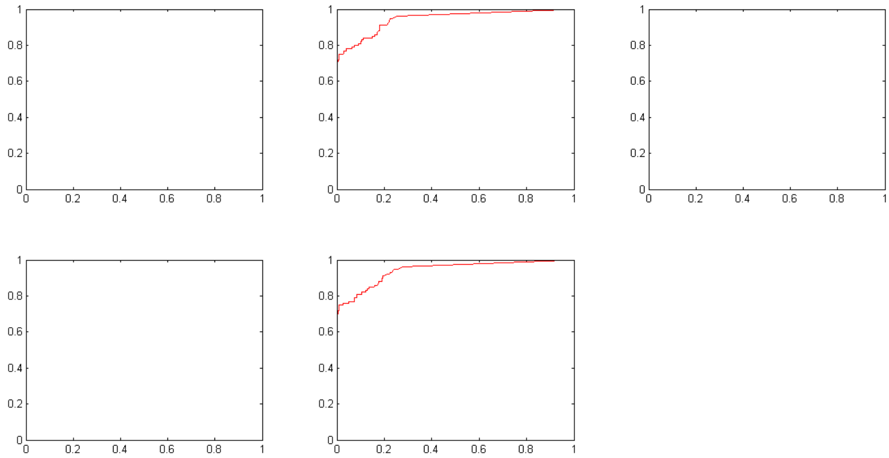

Figure 10: DET Curves for Non-linearly Separable Data

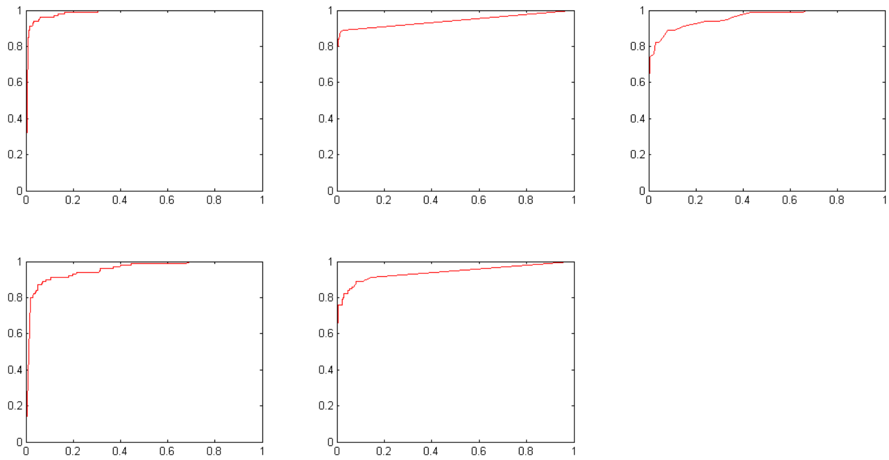

Figure 11: DET Curves for Overlapping Data