University of Waterloo

SD 372 Pattern Recognition

Winter 2003

Lab 1: Orthonormal Transformation and Decision Boundaries
(Due Monday February 10, 2003.)

Lab is to be done in groups of 2 or 3 students. One report per group.

1 Purpose

This lab investigates two related areas: calculating an orthonormal transformation and creating pairwise decision boundaries. You are invited to use whatever software at your disposal in order to assist completion of these labs, however the use of Matlab is recommended (implement your work using M-files).

2 Generating Clusters

2.1 Directions

Use the Matlab function **randn** to assist generation of a 2-d cluster with the following bivariate Gaussian distribution parameters:

$$N = 50 \qquad \mu = [15 \;\; 20]^T \qquad \Sigma = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

where N is the number of samples. Set $a = 9$, $b = 3$ and $c = 4$.

The **randn** function will produce normally distributed data with mean 0 and variance 1.0. To create the correlated, unequal feature variance data as required, you will need to apply a transformation to the uncorrelated, univariance data. Plot the original data, the transformed data and the unit standard deviation contour (for both the original and transformed data) on the same plot. Note: Matlab will probably scale the plot so that isocontours are skewed. You can use the "**axis equal**" Matlab command to alleviate this problem.

2.2 Questions

1. What are the eigenvectors and eigenvalues of the covariance matrix? Do these match the observed shape?
2. What happens to the shape of the cluster as $\{a, b, c\}$ are modified? What happens when b alone is modified? What happens when $b = 0$?

3 Generating Pairwise Discriminants

3.1 Directions

For each pair of classes below, determine and plot each pairwise class discriminant using the Minimum Euclidean Distance (MED), Minimum Intra-Class Distance (MICD), and Maximum A Priori (MAP) measures. For each case:

- Create N_A samples of class A and N_B samples of class B.
- Create a plot of the class pair, showing the class samples, the unit standard deviation contour, and the discriminant boundary.
- For MAP, use the number of samples as an indication of the *a priori* class probability.
- Use the given means and covariances, not the sampled means and covariances.
- Estimate the error rate for each classifier for the given cases. Explain how you determined the error rate and provide rationale for your methodology.

CASE 1:

Class A and Class B

$N_A = 100 \qquad \mu_A = [15 \;\; 10]^T \qquad \Sigma_A = \begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix}$

$N_B = 100 \qquad \mu_B = [20 \;\; 16]^T \qquad \Sigma_B = \begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix}$

CASE 2:

Repeat CASE 1 using $N_A = 50$ and $N_B = 250$.

CASE 3:

Class C and Class D

$N_C = 100 \qquad \mu_C = [15 \;\; 10]^T \qquad \Sigma_C = \begin{bmatrix} 40 & 4 \\ 4 & 16 \end{bmatrix}$

$N_D = 50 \qquad \mu_D = [20 \;\; 25]^T \qquad \Sigma_D = \begin{bmatrix} 8 & 3 \\ 3 & 8 \end{bmatrix}$

CASE 4:

Instead of using MED, MICD, or MAP, suggest a 'reasonable' linear decision boundary (that is, pick any linear boundary you find suitable). Apply this method to Case 3.

3.2 Questions

1. Summarize your error results in a table. Are the results as expected? Explain.
2. CASE 1: What type of decision boundary (ie. line, ellipse, etc.) is produced in this case? Why?
3. CASE 2: What happens to the decision boundary when the *a priori* probabilities are modified? Why?
4. CASE 3: What happens to the decision boundary as the number of samples in class C increases? Sketch the direction of this change in the decision boundary on your plot.
5. CASE 4: Is finding a linear boundary a reasonable approach? How does the error rate compare to the MED/MICD/MAP approaches in Case 3?

4 Report

Include in your report:

- A brief introduction.
- Discussion of your implementations and results (Include brief derivations for equations implemented in M-files. Don't bother generating equations using a word processor. Handwritten equations are ok as long as they are readily legible.)
- Printouts of pertinent graphs.
- M-files for each section.
- Include responses to all questions.
- A brief summary of your results with conclusions.