Memorial University of Newfoundland

Pattern Recognition

Lecture 5
May 16, 2006

http://www.engr.mun.ca/~charlesr

Office Hours: Tuesdays & Thursdays 8:30 - 9:30 PM
EN-3026

### Housekeeping

- Assignment 1 now due on May 23
- Project abstracts due Thursday
  - Submit via email
  - Expand on your topics to include a problem statement, background, and potential pattern recognition technique
- Word document format or PDF
- Include references (no Wikipedia)
Recap

Feature Weighting

$$d_w^2(x, z) = \sum_{j=1}^{n} \left( \frac{x_j - z_j}{s_j} \right)^2$$

Good for uncorrelated features

$$x' = Wx = \begin{bmatrix} w_1 & w_2 & 0 \\ 0 & \ddots & w_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Recap

Orthonormal Whitening

$$y = Wx$$

$$W = W_2W_1 = \Lambda^{-1/2}\Phi^T$$

Full transformation:
1. Rotate feature axes ($\Phi^T$)
2. Scale to account for variances ($\Lambda^{-1/2}$)

![Diagram showing the transformation process from $x$ to $y$](image)
Recap

MICD Metric

Euclidean distance in the new space is a good measure of pattern similarity.

\[
d_{E}^{2}(y, z_{Y}) = (x - z)^{T}S^{-1}(x - z) = d_{M}^{2}(x, z)
\]

d_{M}^{2}(x, z) is the multivariate version of measuring distance in standard deviation units.

Compare \((x - z)^{T}S^{-1}(x - z)\) with \((\frac{x - z}{s})^{2}\)

This metric is called the **Minimum Intra-Class Distance** (MICD) metric, because it corresponds to the transformation that makes all members of a class as similar as possible. It is also known as the **Mahalanobis distance**.

---

Properties of Pattern-to-Class Distance

Consider equidistance contours:

\[
(x - m)^{T}S^{-1}(x - m) = c^{2}, \text{ where } S = \Phi \Lambda \Phi^{T}
\]

What shape are the distributions?
Case 1: \( S = I \) (identity)

\[
\therefore (x - m)^T S^{-1} (x - m) = (x - m)^T (x - m)
\]

The MICD is the same as the MED.

Case 2: \( S \) is diagonal

\[
d_M^2(x, m) = (x - m)^T \begin{bmatrix}
\frac{1}{s_1^2} & 0 \\
0 & \ddots \\
0 & \frac{1}{s_n^2}
\end{bmatrix} (x - m)
\]

\[
= \sum_{i=1}^{n} \left( \frac{x_i - m_i}{s_i} \right)^2 = 1 \text{ for } c = 1
\]
Case 2: $S$ is arbitrary

Contour is an oriented ellipse. $\Phi$ maps $S$ to $\Lambda$ with pure rotation, so equidistance contour does not change shape.

If $y = \Phi^T x$, in $y$-space we have:

$$\sum_{i=1}^{n} \left( \frac{y_i - m_{yi}}{\sqrt{\lambda_i}} \right)^2 = 1 \text{ for } c^2 = 1$$

So before rotation we have:

The eigenvectors are principle axes directions.
The (eigenvalues)$^{1/2}$ are the lengths of the principle axes.

Distance to the unit contour along any eigenvector is the square root of the corresponding eigenvalue, which is the standard deviation of the distribution along that eigenvector (from class covariance matrix).
### MICD Classifier

\[ x \in C_1 \text{ iff } d^2_M(x, C_1) < d^2_M(x, C_2) \]

iff \((x - m_1)^T S_1^{-1} (x - m_1) < (x - m_2)^T S_2^{-1} (x - m_2)\)

Measure the distance to each of the classes using its own metric (i.e. its own standard deviation units).

![Diagram showing decision boundary](image)

Classifier defines a decision boundary. This partitions the feature space into regions for each class.

### Decision Boundary (2-classes)

The decision boundary follows the path where the MICD is equal to both classes.

\[
(x - m_1)^T S_1^{-1} (x - m_1) - (x - m_2)^T S_2^{-1} (x - m_2) = 0
\]

\[
\vdots
\]

\[x^T (S_1^{-1} - S_2^{-1}) x + 2(m_2^T S_2^{-1} - m_1^T S_1^{-1}) x + m_1^T S_1^{-1} m_1 - m_2^T S_2^{-1} m_2 = 0\]

We can write this as...

\[x^T Q x + w^T x + w_0 = 0\]

The form of the decision surface depends on \(Q, w,\) and \(w_0\) which in turn depend on the class means and covariance matrices.
Decision Boundary

In general the decision boundary is a hyperquadratic. Consider some special cases.

1. Equal Covariance Matrices ($S_1 = S_2$)
   (Typical of problems where all class measurements are corrupted by similar amounts of noise.)

   Within-class variability is the same for all classes. So $S^{-1}_1 - S^{-1}_2 = 0$.
   Thus $Q = 0$ and the boundary equation is:
   
   $$ w^T x + w_0 = 0 $$
   
   $$ 2(m_2 - m_1)^T S^{-1} x + m_1^T S^{-1} m_1 + m_2^T S^{-1} m_2 = 0 $$
   
   $$ (m_2 - m_1)^T S^{-1} \left( x - \frac{m_1 + m_2}{2} \right) = 0 $$

   This is an equation of a hyperplane through $\frac{m_1 + m_2}{2}$
   with normal vector $S^{-1}(m_2 - m_1)$
2. $S_1 = c_1I, S_2 = c_2I$
Both distributions are aligned to the feature axes (features are uncorrelated).

3. $S_1 = cS_2$
Both distributions have same shape, but different amounts of scatter.
4. General case

Hyperquadratic

Hyperparabola

Hyperhyperbola

From Figure 2.14 from Pattern Classification by Duda, Hart, and Stork.

5. $m_1 = m_2, S_1 \neq S_2$

Both classes have same mean but different variances

MICD: $d^2_M(x, C_1) > d^2_M(x, C_2) \forall x$

because we measure in standard deviation units.

$\Rightarrow x \in C_2 \forall x$
However, if $C_1$ and $C_2$ are both equally probable, we need a better classifier.

Consider the 1-dimensional case:

![Graph of probability density functions](image)

This suggests that we want to modify the MICD to favour dense classes when near to the mean.... thus we need to look at probabilistic classification!

Modification of Figure 2.13 From Pattern Classification by Duda, Hart, and Stork

Distance-based Classification Example

Suppose a class is known to include the following samples:

\[
\begin{bmatrix}
5 & 6 & 6 & 7 & 7 & 8 & 8 & 9 \\
2 & 3 & 4 & 2 & 3 & 2 & 3 & 4
\end{bmatrix}
\]

a. Plot the points.

b. Calculate the mean vector and the covariance matrix.

c. Find the eigenvectors and eigenvalues of the class covariance matrix.

d. Sketch the unit standard deviation contour, labelling the class mean and the principal axes.

e. Suppose a second class with mean vector $\mathbf{m}_2=(3,7)$ and the same covariance matrix is present. Find the MICD decision boundary. Plot the USD contour and decision boundary.

f. Find the transformation that maps the USD contours of both classes to unit circles. What are the class means after the transformation?