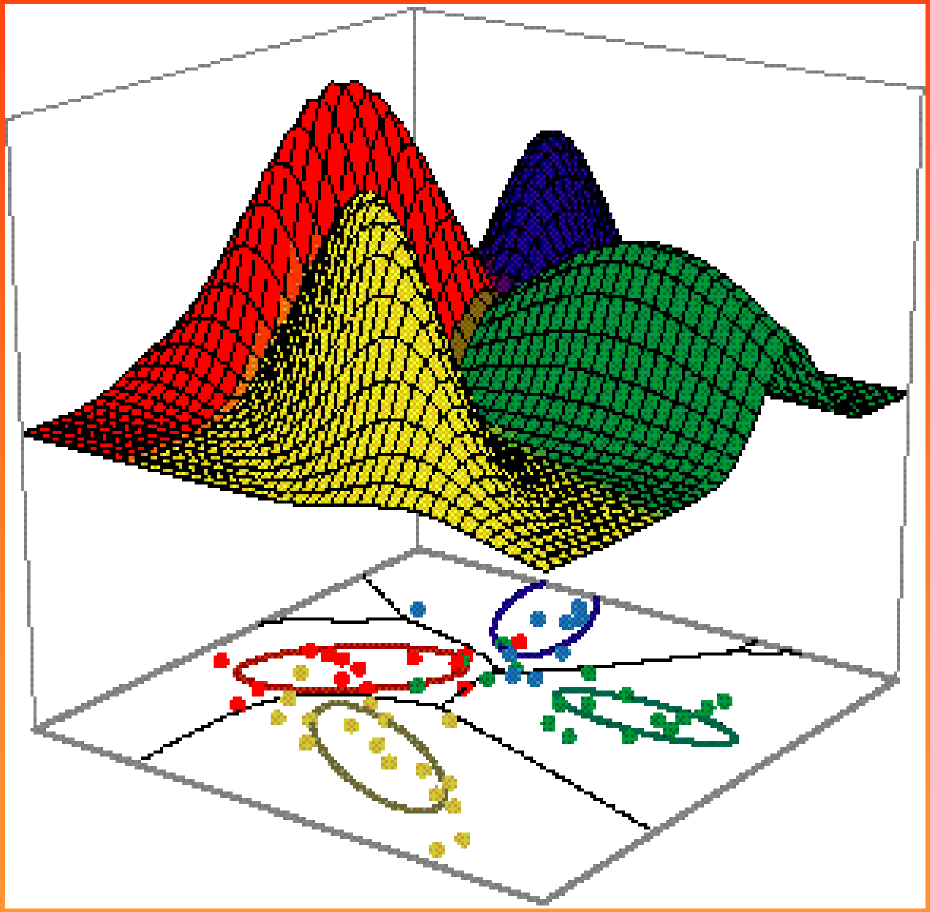


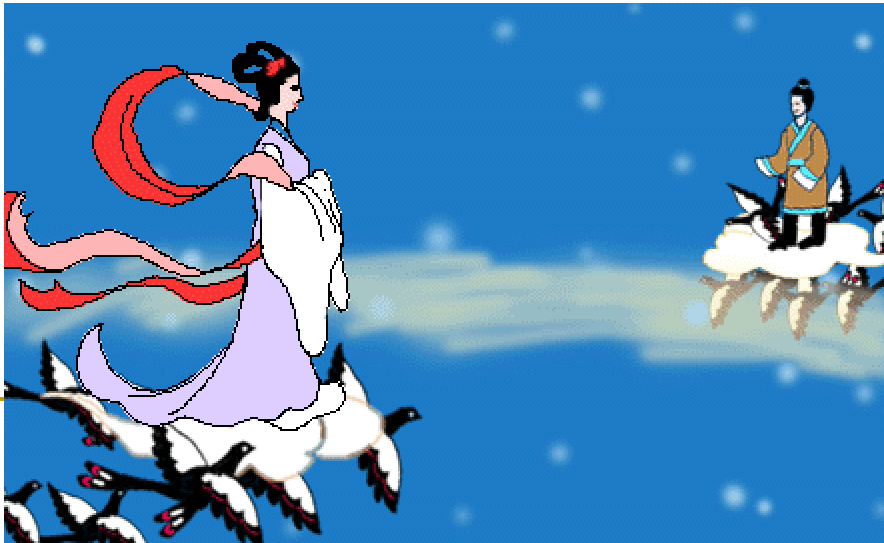
Pattern Classification

All materials in these slides were taken from
Pattern Classification (2nd ed) by R. O. *Duda, P. E. Hart and D. G. Stork, John Wiley & Sons, 2000*
with the permission of the authors and the publisher

5.8.2 Relation to Fisher's Linear Discriminant

- With the proper choice of the vector b, the MSE discriminant function $a^t y$ is directly related to Fisher's linear discriminant (Chapter 3 Section 3.8.2).

- To do so, we use the linear rather than generalized linear discriminant function.



Assumption

- A set of n d-dimensional samples $x_1,\ldots x_n$, n_1 of which are in the subset D_1 labeled w_1, and n_2 of which are in the subset D_2 labeled w_2.
- A sample y_i if formed from x_i by adding a threshold component $x_0=1$ to make an augmented pattern vector.
- If the sample is labeled w_2, then the entire pattern vector is multiplied by -1
- With no loss in generality, assume the first n_1 samples are labeled w_1 and the second n_2 are labeled w_2.

- MSE method based on Ya=b can be equivalent to Fisher's Linear Discriminant

- Condition: the number of the samples approaches to infinity.

Then the matrix Y can be partitioned as follows:

$$Y = \begin{bmatrix} 1_1 & X_1 \\ -1_2 & -X_2 \end{bmatrix},$$

Where 1_i is a column vector of n_i ones, and X_i is an n_i-by-d matrix whose rows are the samples labeled w_i.

Correspondingly,

$$a = \begin{bmatrix} \omega_0 \\ w \end{bmatrix} \qquad b = \begin{bmatrix} \frac{n}{n_1} 1_1 \\ \frac{n}{n_2} 1_2 \end{bmatrix}$$

By Writing Eq.45 for a in terms of the partitioned matrices:

$$\begin{bmatrix} 1_1^t & -1_2^t \\ X_1^t & -X_2^t \end{bmatrix} \begin{bmatrix} 1_1 & X_1 \\ -1_2 & -X_2 \end{bmatrix} \begin{bmatrix} \omega_0 \\ w \end{bmatrix} = \begin{bmatrix} 1_1^t & -1_2^t \\ X_1^t & -X_2^t \end{bmatrix} \begin{bmatrix} \frac{n}{n_1} 1_1 \\ \frac{n}{n_2} 1_2 \end{bmatrix}. \quad (49)$$

Defining the sample means

$$m_i = \frac{1}{n_i} \sum_{x \in D_i} x \quad i = 1, 2 \qquad (50)$$

And the pooled sample scatter matrix

$$Sw = \sum_{i=1}^{2} \sum_{x \in D_i} (x - m_i)(x - m_i)^t \qquad (51)$$

Multiply the matrices of Eq.49 and obtain

$$\begin{bmatrix} n & (n_1 m_1 + n_2 m_2)^t \\ (n_1 m_1 + n_2 m_2) & S_w + n_1 m_1 m_1{}^t + n_2 m_2 m_2{}^t \end{bmatrix} \begin{bmatrix} \omega_0 \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ n(m_1 - m_2) \end{bmatrix}$$

This can be viewed as a pair of equations, the first of which can be solved for w_0 in terms of w:

$$\omega_0 = -m^t w, \qquad (52)$$

Where m is the mean of all of the samples.

Substituting this in the second equation, we obtain

$$\left[\frac{1}{n} S_w + \frac{n_1 n_2}{n^2} (m_1 - m_2)(m_1 - m_2)^t \right] w = m_1 - m_2 . \qquad (53)$$

Because the vector $(m_1 - m_2)(m_1 - m_2)^t w$ is in the direction of $m_1 - m_2$ for any value of w, we can write

$$\frac{n_1 n_2}{n^2}(m_1 - m_2)(m_1 - m_2)^t w = (1\text{-}a)(m_1 - m_2),$$

where a is some scalar.

Then Eq.53 yields

$$w = S_w^{-1}(m_1 - m_2), \qquad (54)$$

Which, except for an unimportant scale factor, is identical to the solution for Fisher's linear discriminant.

In addition, we obtain the threshold weight w_0 and the following decision rule:

$$\text{Decide } \omega_1 \text{ if } w^t(x - m) > 0; \text{ otherwise decide } \omega_2 .$$

References of LDA

- David Zhang, FengXi Song, Yong Xu, ZhiZhen Liang
- "Advanced Pattern Recognition Technologies with Applications
- to Biometrics", Medical Information Science Reference, 2009

- Yong Xu, David Zhang, Represent and fuse bimodal biometric images at the feature level: complex-matrix-based fusion scheme, Opt. Eng. 49(3), 037002, 2010

- Yong Xu, Jing-Yu Yang, Zhong Jin, A novel method for Fisher discriminant Analysis. Pattern Recognition, 37 (2), 381-384, 2004