MULTIVARIATE STATISTICAL ANALYSIS

Lecture 7
Multivariate Linear Classification

Associate Professor Lý Quốc Ngọc

KHOA CÔNG NGHỆ THÔNG TIN
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

fit@hcmus

Contents

7.1. Purpose of Classification

Discrimination (separation) to describe the differential features of objects from several known populations.

Classification (allocation) to sort objects into two or more labeled classes.

7.2. Problem statement

LDA for two classes

Select the projection that maximizes the ratio of dissimilarity between classes and dissimilarity in class.

Suppose that we have 2 classes lớp C_1, C_2,with expect. μ_1, μ_2

C_1, C_2 are presented by $X^{(1)}(N_1 \times p),\ X^{(2)}(N_2 \times p)$

Projecting data onto a straight line can be described using a coefficient vector w :

$$y_n = w^T x_n,\ 1 \le \mathrm{n} \le \mathrm{N}$$

The expectation of each class after projection :

$$m_i = \frac{1}{N_i} \sum_{n=1}^{N_i} y_n = w^T \mu_i,\ \mathrm{i} = 1{,}2$$

7.2. Problem statement

Dissimilarity between classes:

$$(m_1 - m_2)^2 = w^T(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T w = w^T S_B w$$

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$$

Dissimilarity in class:

$$s_1^2 + s_2^2 = \sum_{k=1}^{2} \sum_{n \in C_k} (w^T(x_n - \mu_k))^2 =$$

$$= w^T \Big(\sum_{k=1}^{2} \sum_{n \in C_k} (x_n - \mu_k)(x_n - \mu_k)^T\Big) w$$

$$= w^T S_w w$$

7.2. Problem statement

Find w to maximize $F_{fisher}(w)$

$$F_{fisher}(w) = \frac{w^T S_B w}{w^T S_w w}$$

7.3. Method

fit@hcmus

The necessary condition for $F_{fisher}(w)$ to achieve a maximum value:

$$\nabla_w F_{fisher}(w) = 0$$

$$\nabla_w \left(\frac{w^T S_B w}{w^T S_w w} \right) = 0$$

$$(w^T S_w w)(2 S_B w) - (w^T S_B w)(2 S_w w) = 0$$

$$S_B w = \frac{w^T S_B w}{w^T S_w w} (S_w w)$$

$$S_w^{-1} S_B w = F_{fisher}(w) w$$

7.3. Method

Select w such that $(\mu_1 - \mu_2)^T w = F_{fisher}(w) = L$
is the largest eigen value of $S_w^{-1} S_B$

$$Lw = S_w^{-1}(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T w = L S_w^{-1}(\mu_1 - \mu_2)$$

$$w = \alpha S_w^{-1}(\mu_1 - \mu_2)$$

7.2. Problem statement

LDA for multiclass

Discussion in class

7.3. Method

LDA for multiclass

Discussion in class

7.4. Geometrical Explanation

Bàn luận trên lớp

7.5. Model Checking

Discussion in class

7.6. Case Study

Discussion in class