MSA - Classification

•••

404-noname

19127614 - Nguyễn Anh Tuấn
19127517- Hồ Thiên Phước
19127165 - Võ Gia Huy

I. Giới thiệu

Học Máy gồm 4 loại phương thức học:

1. Học có giám sát gồm 2 bài toán lớn:
   a. **PHÂN LỚP**
   b. Hồi quy
2. Học không giám sát
3. Học nửa giám sát
4. Học tăng cường

PHÂN LỚP là quá trình tìm kiếm 1 mô hình phân lớp (model) giúp tách dữ liệu thành nhiều lớp theo nhiều nhãn (label) khác nhau.

Model được xây dựng dựa trên tập dữ liệu được xây dựng từ trước có gán nhãn (tập huấn luyện).

I. Động lực - Ý nghĩa khoa học

Định nghĩa:

Classification là một kĩ thuật nhằm phân tích các dữ liệu (observation), sau đó phân bố các đối tượng mới (observation) cho các class khác nhau đã được xác định trước đó.

Nhiệm vụ: Gồm hai việc chính:

- Sắp xếp các đối tượng (observation) vào 2 hay nhiều lớp đã được dán nhãn.
- Tìm ra một phương pháp tối ưu để dán nhãn đối tượng cho các lớp đã được dán nhãn trước. (Trọng tâm)

I. Động lực - Ứng dụng

Nhận dạng chữ số viết tay

Nhận dạng khuôn mặt

Khai thác văn bản (Text mining)

Truy vấn ảnh

Xếp loại học lực dựa trên ĐTB

Tìm và đánh giá tỉ lệ để dự đoán xác suất mắc bệnh tiểu đường của một người ở một độ tuổi nhất định dựa trên tuổi tác, tình trạng sức khỏe, truyền thống gia đình, thói quen ăn uống, thời kì mang thai, huyết áp, độ dày của da, insulyn, chỉ số khối cơ thể,...

II. Phát biểu bài toán - Đặt vấn đề

Ta có n lớp (population) R như sau:

Lớp R_1 có label π_1 , hàm mật độ xác suất $f_1(x)$, xác suất tiên nghiệm p_1

Lớp R_2 có label π_2, hàm mật độ xác suất $f_2(x)$, xác suất tiên nghiệm p_2

...

Lớp R_n có label π_n, hàm mật độ xác suất $f_n(x)$, xác suất tiên nghiệm p_n

Bài toán:

Ta có tập vector $X = [X_1, X_2, \dots X_p]$ và muốn phân bố tập vector này vào n lớp R

Để làm việc trên, ta phải tìm ra một Mô Hình Phân Lớp để biết nên phân bố được các vector trong tập X này vào lớp nào trong n lớp R

II. Phát biểu bài toán - Input, Output

Input:

Vector $X^T = [X_1, X_2, X_3 \ldots X_p]$

Output:

X được dán nhãn π_1 hoặc π_2 hoặc .... π_n

III. Bài toán phân lớp 2 populations:

Input:

Vector $X^T = [X_1, X_2, X_3 ... X_p]$

Output:

X được dán nhãn π_1 hoặc π_2

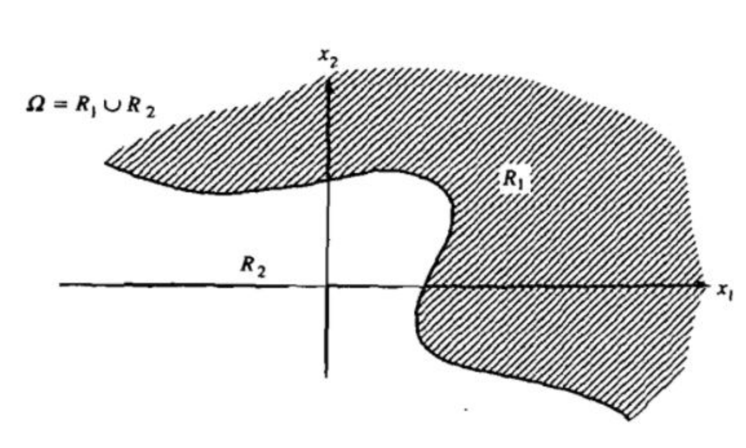
III.Bài toán phân lớp 2 populations

3.1) Phân tách và phân loại cho 2 quần thể:

-Ta định nghĩa Ω là tập không gian mẫu ⊂

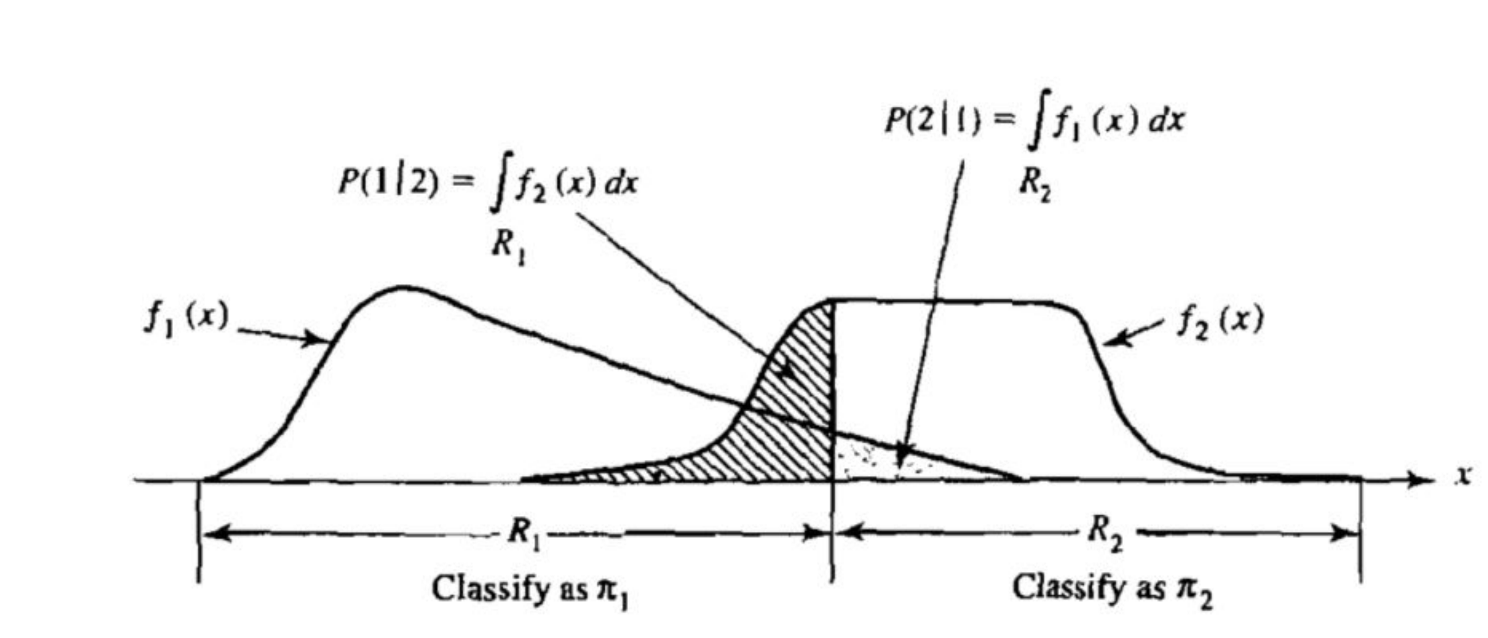
R1 ⊂ Ω là tập con của Ω trong đó X được dán nhãn π_1

R2 = Ω - R1 là tập con của Ω trong đó X được dán nhãn π_2



III.Bài toán phân lớp 2 populations

3.1) Phân tách và phân loại cho 2 quần thể:



III.Bài toán phân lớp 2 populations

3.1) Phân tách và phân loại cho 2 quần thể:

P(2|1): xác suất có điều kiện, phân lớp một đối tượng x vào π_2 trong khi nó thuộc về π_1

$$P(2|1) = P(\mathbf{X} \in R_2|\pi_1) = \int_{R_2} f_1(\mathbf{x})d\mathbf{x}$$

P(1|2): xác suất có điều kiện, phân lớp một đối tượng x vào π_1 trong khi nó thuộc về π_2

$$P(1|2) = P(\mathbf{X} \in R_1|\pi_2) = \int_{R_1} f_2(\mathbf{x})d\mathbf{x}$$

III.Bài toán phân lớp 2 populations

3.1) Phân tách và phân loại cho 2 quần thể:

Khi đó ta có các xác suất điều kiện:

P(Phân lớp sai X vào π_1)= P(X $\in R_1|\pi_2$) $P(\pi_2)$=$P(1|2)p_2$

P(Phân lớp sai X vào π_2)= P(X $\in R_2|\pi_1$) $P(\pi_1)$=$P(2|1)\,p_1$

Trong nhiều trường hợp thực tế chi phí phân lớp sai đối tượng không bằng nhau nên ta gọi c(i,j) là chi phí phân lớp sai đối tượng i vào j ; i,j=1,2.

III.Bài toán phân lớp 2 populations

3.1) Phân tách và phân loại cho 2 quần thể:

- Ta có ma trận chi phí sau:

		Classify as: π_1	π_2
Truth:	π_1	0	$c(2\|1)$
	π_2	$c(1\|2)$	0

- Để phân lớp hiệu quả ta cần một hàm độ lỗi sao cho khi phân lớp một đối tượng thì độ lỗi là nhỏ nhất. ECM-excepted cost of misclassification

$$ECM = c(2|1)P(2|1)p_1 + c(1|2)P(1|2)p_2$$

III.Bài toán phân lớp 2 populations

3.1) Phân tách và phân loại cho 2 quần thể:

Để cực tiểu hóa độ lỗi ECM thì ta có bất đẳng thức sau:

$$R_1 : \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)$$

$$R_2 : \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)$$

V. Bài toán phân nhiều lớp

5.2) Phân loại cho nhiều quần thể đa biến

Đặt:

$$\mathbf{d}_i^Q(\mathbf{x}) = -\frac{1}{2}ln(\boldsymbol{\Sigma}_i) - \left[\frac{-1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i{}^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right] + \boldsymbol{ln}(p_i)$$

Vậy để phân lớp x_0 vào lớp π_k thì:

$$d_k^Q(x_0) = lớn\ nhất\ trong\ d_1^Q(x_0), d_2^Q(x_0) \ldots, d_g^Q(x_0)$$

III.Bài toán phân lớp 2 populations

3.1) Phân tách và phân loại cho 2 quần thể:

-Với p1/p2=1 thì

$$R_1: \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{c(1|2)}{c(2|1)} \quad R_2: \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \frac{c(1|2)}{c(2|1)}$$

-Với c(1|2)/c(2|1) thì

$$R_1: \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{p_2}{p_1} \quad R_2: \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \frac{p_2}{p_1}$$

-Với p2/p1=c(1|2)/c(2|1)=1 hoặc p2/p1=1/(c(1|2)/c(2|1)) thì

$$R_1: \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq 1 \quad R_2: \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < 1$$

III.Bài toán phân lớp 2 populations

3.2) Phân loại cho 2 quần thể đa biến:

Giả định có hai hàm mật độ xác suất chuẩn là $f_1(x)$ và $f_2(x)$. Hàm đầu có mean vector μ_1 và covariance matrix $\sum_1$, hàm thứ hai có mean vector μ_2 và covariance matrix $\sum_2$

3.2.1) $\sum_1=\sum_2=\sum$

Hàm mật độ xác suất của biến $X^T=[X_1,...,X_p]$ của 2 quần thể π_1 và π_2:

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \, exp\left[\frac{-1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right]$$

với i =1,2

III.Bài toán phân lớp 2 populations

3.2) Phân loại cho 2 quần thể đa biến:

3.2.1) $\Sigma_1 = \Sigma_2 = \Sigma$

Σ , μ_1, μ_2 cho trước và $(2\pi)^{p/2}|\Sigma|^{½}$ là hằng số nên ta có:

$$\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} = \frac{exp\left[\frac{-1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)\right]}{exp\left[\frac{-1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2)\right]}$$
$$= exp\left[\frac{-1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) + \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2)\right]$$

III.Bài toán phân lớp 2 populations

3.2) Phân loại cho 2 quần thể đa biến:

3.2.1) $\Sigma_1 = \Sigma_2 = \Sigma$

Mặt khác

$$\frac{-1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(x-\boldsymbol{\mu}_2)+\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(x-\boldsymbol{\mu}_2)$$
$$=(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}x-\frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)$$

Và để cực tiểu ECM trong vùng R1, R2 thì

$$R_1: \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{c(1|2)}{c(2|1)}\frac{p_2}{p_1}$$

$$R_2: \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \frac{c(1|2)}{c(2|1)}\frac{p_2}{p_1}$$

III.Bài toán phân lớp 2 populations

3.2) Phân loại cho 2 quần thể đa biến:

3.2.1) $\Sigma_1 = \Sigma_2 = \Sigma$

Do đó ta có:

$$R_1: (\mu_1 - \mu_2)^T \Sigma^{-1} x - \frac{1}{2}(\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 + \mu_2) \geq ln\left[\frac{c(1|2)}{c(2|1)}\frac{p_2}{p_1}\right]$$

$$R_2: (\mu_1 - \mu_2)^T \Sigma^{-1} x - \frac{1}{2}(\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 + \mu_2) < ln\left[\frac{c(1|2)}{c(2|1)}\frac{p_2}{p_1}\right]$$

III.Bài toán phân lớp 2 populations

3.2) Phân loại cho 2 quần thể đa biến:

3.2.1) $\sum_1=\sum_2=\sum$

Tuy nhiên trong thực tế thì μ_1, μ_2, $\sum$ không xác định. Nhưng nếu có được n_1 biến X với $X^T=[X_1 ,...X_p]$ với p là số chiều dữ liệu có phân lớp π_1 và n2 biến X có phân lớp π_2 ($n_1+n_2-1 \geq p$). Ta có ma trận :

$$\mathbf{X_1} = \begin{bmatrix} \mathbf{x}_{11}^T \\ \mathbf{x}_{12}^T \\ \vdots \\ \mathbf{x}_{1n_1}^T \end{bmatrix} \text{ có thước } (n_1 \text{ x } p)$$

$$\mathbf{X_2} = \begin{bmatrix} \mathbf{x}_{21}^T \\ \mathbf{x}_{22}^T \\ \vdots \\ \mathbf{x}_{2n_2}^T \end{bmatrix} \text{ có thước } (n_2 \text{ x } p)$$

III.Bài toán phân lớp 2 populations

3.2) Phân loại cho 2 quần thể đa biến:

3.2.1) $\Sigma_1 = \Sigma_2 = \Sigma$

Ta tính được sample mean vector x_1 và x_2 và covariance matrix S_1, S_2 của X_1, X_2

Vì 2 covariance matrix của 2 quần thể bằng nhau nên:

$$S_{pooled} = \frac{n_1 - 1}{n_1 + n_2 - 2} S_1 + \frac{n_2 - 1}{n_1 + n_2 - 2} S_2$$

III.Bài toán phân lớp 2 populations

3.2) Phân loại cho 2 quần thể đa biến:

3.2.1) $\Sigma_1 = \Sigma_2 = \Sigma$

Khi đó:

$$R_1: (\overline{\mathbf{x}_1} - \overline{\mathbf{x}_2})^T S_{pooled}{}^{-1}\mathbf{x} - \frac{1}{2}(\overline{\mathbf{x}_1} - \overline{\mathbf{x}_2})^T S_{pooled}{}^{-1}(\overline{\mathbf{x}_1} + \overline{\mathbf{x}_2}) \geq ln\left[\frac{c(1|2)}{c(2|1)}\frac{p_2}{p_1}\right]$$

$$R_2: (\overline{\mathbf{x}_1} - \overline{\mathbf{x}_2})^T S_{pooled}{}^{-1}\mathbf{x} - \frac{1}{2}(\overline{\mathbf{x}_1} - \overline{\mathbf{x}_2})^T S_{pooled}{}^{-1}(\overline{\mathbf{x}_1} + \overline{\mathbf{x}_2}) < ln\left[\frac{c(1|2)}{c(2|1)}\frac{p_2}{p_1}\right]$$

III.Bài toán phân lớp 2 populations

3.2) Phân loại cho 2 quần thể đa biến:

3.2.1) $\Sigma_1 = \Sigma_2 = \Sigma$

Ta có nhận xét sau:

$$\hat{y} = (\overline{\mathbf{x}_1} - \overline{\mathbf{x}_2})^T S_{pooled}{}^{-1}\mathbf{x} = \hat{\mathbf{a}}^T\mathbf{x} \text{ (Tổ hợp tuyến tính)}$$

$$\hat{m} = \frac{1}{2}(\overline{\mathbf{x}_1} - \overline{\mathbf{x}_2})^T S_{pooled}{}^{-1}(\overline{\mathbf{x}_1} + \overline{\mathbf{x}_2}) = \frac{1}{2}(\overline{\mathbf{y}_1} + \overline{\mathbf{y}_2})$$

Trong đó:

$$\overline{\mathbf{y}_1} = (\overline{\mathbf{x}_1} - \overline{\mathbf{x}_2})^T S_{pooled}{}^{-1}\overline{\mathbf{x}_1} = \hat{\mathbf{a}}^T\overline{\mathbf{x}_1}$$

$$\overline{\mathbf{y}_2} = (\overline{\mathbf{x}_1} - \overline{\mathbf{x}_2})^T S_{pooled}{}^{-1}\overline{\mathbf{x}_1} = \hat{\mathbf{a}}^T\overline{\mathbf{x}_2}$$

III.Bài toán phân lớp 2 populations

3.2) Phân loại cho 2 quần thể đa biến:

3.2.1) $\sum_1=\sum_2=\sum$

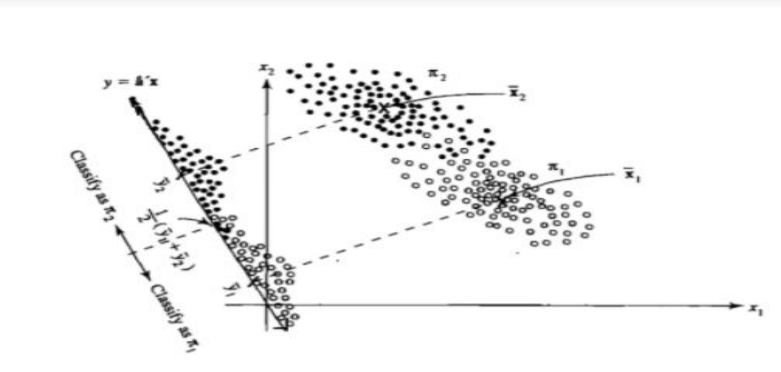
Trường hợp đặc biệt khi

$$ln = \left(\frac{c(1|2)}{c(2|1)}\frac{p_2}{p_1}\right) = 0$$

Thì x_0 được phân lớp vào π_1

$$y = (\overline{\mathbf{x}_1} - \overline{\mathbf{x}_2})^T S_{pooled}{}^{-1}\mathbf{x}_0 = \hat{\mathbf{a}}^T\mathbf{x}_0 \geq \hat{m}$$

Còn ngược lại thì x_0 được phân lớp vào π_2



III.Bài toán phân lớp 2 populations

3.2) Phân loại cho 2 quần thể đa biến:

3.2.2) Tiếp cận theo hướng của Fisher:

Chúng ta sẽ tập trung vào tính chất được ứng dụng để phân lớp đối tượng:

$$\hat{y} = (\overline{x_1} - \overline{x_2})^T S_{pooled}{}^{-1} x = \hat{a}^T x$$

Gỉa sử $y_{11},\dots,y_{1n}$ là tổ hợp tuyến tính của các giá trị x thuộc π_1 và $y_{11},\dots,y_{1n}$ là tổ hợp tuyến tính của các giá trị thuộc π_2

Sự khác biệt của hai tập chính là sự khác biệt giữa y_1 và y_2 được thể hiện bằng đơn vị độ lệch chuẩn:

$$seperation = \frac{|\overline{y_1} - \overline{y_2}|}{S_y}$$

III.Bài toán phân lớp 2 populations

3.2) Phân loại cho 2 quần thể đa biến:

3.2.2) Tiếp cận theo hướng của Fisher:

Trong đó:

$$S_y^2 = \frac{\sum_{j=1}^{n_1}(y_{1j} - \overline{y_1})^2 + \sum_{j=1}^{n_2}(y_{2j} - \overline{y_2})^2}{n_1 + n_2 - 2}$$

Vậy để seperation đạt cực đại thì:

$$seperation = \frac{|\overline{y_1} - \overline{y_2}|}{S_y} = \frac{(\overline{y_1} - \overline{y_2})^2}{{S_y}^2} = \frac{(\mathbf{\hat{a}}^T\overline{\mathbf{x}_1} - \mathbf{\hat{a}}^T\overline{\mathbf{x}_2})^2}{\mathbf{\hat{a}}^T S_{pooled}\mathbf{\hat{a}}} = \frac{(\mathbf{\hat{a}}^T\mathbf{d})^2}{\mathbf{\hat{a}}^T S_{pooled}\mathbf{\hat{a}}}$$

Trong đó : $d = \overline{x1} - \overline{x2}$

III.Bài toán phân lớp 2 populations

3.2) Phân loại cho 2 quần thể đa biến:

3.2.2) Tiếp cận theo hướng của Fisher:

Trong đó:

$$S_y^2 = \frac{\sum_{j=1}^{n_1}(y_{1j} - \overline{y_1})^2 + \sum_{j=1}^{n_2}(y_{2j} - \overline{y_2})^2}{n_1 + n_2 - 2}$$

Vậy để seperation đạt cực đại thì:

$$seperation = \frac{|\overline{y_1} - \overline{y_2}|}{S_y} = \frac{(\overline{y_1} - \overline{y_2})^2}{S_y{}^2} = \frac{(\hat{\mathbf{a}}^T\overline{\mathbf{x}_1} - \hat{\mathbf{a}}^T\overline{\mathbf{x}_2})^2}{\hat{\mathbf{a}}^T S_{pooled}\hat{\mathbf{a}}} = \frac{(\hat{\mathbf{a}}^T\mathbf{d})^2}{\hat{\mathbf{a}}^T S_{pooled}\hat{\mathbf{a}}}$$

Trong đó : $d = \overline{x1} - \overline{x2}$

III.Bài toán phân lớp 2 populations

3.2) Phân loại cho 2 quần thể đa biến:

3.2.2) Tiếp cận theo hướng của Fisher:

Tương đương với

$$D^2 = (\overline{x_1} - \overline{x_2})^T S_{pooled}{}^{-1} (\overline{x_1} - \overline{x_2})$$

III.Bài toán phân lớp 2 populations

3.2) Phân loại cho 2 quần thể đa biến:

3.2.3) $\sum_1 != \sum_2$

Ta có:

$$R_1: \frac{-1}{2} x^T(\Sigma_1^{-1} - \Sigma_2^{-1})\mathbf{x} + (\mu_1{}^T\Sigma_1^{-1} - \mu_2{}^T\Sigma_2^{-1})\,\mathbf{x} - k \geq \frac{c(1|2)}{c(2|1)}\frac{p_2}{p_1}$$

$$R_2: \frac{-1}{2} x^T(\Sigma_1^{-1} - \Sigma_2^{-1})\mathbf{x} + (\mu_1{}^T\Sigma_1^{-1} - \mu_2{}^T\Sigma_2^{-1})\,\mathbf{x} - k < \frac{c(1|2)}{c(2|1)}\frac{p_2}{p_1}$$

Trong đó :

$$k = \frac{1}{2}\, ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) + \frac{1}{2}(\mu_1{}^T\Sigma_1^{-1}\mu_1 - \mu_2{}^T\Sigma_2^{-1}\mu_2)$$

Vậy với giá trị x_0 nếu thỏa điều kiện của R_1 thì được phân vào π_1 và phân lớp vào π_2 trong trường hợp ngược lại

III.Bài toán phân lớp 2 populations

3.2) Phân loại cho 2 quần thể đa biến:

3.2.3) $\Sigma_1 != \Sigma_2$

Hàm mật độ xác suất f_i của π_i với i=1,2 :

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}_i|^{1/2}} \, exp\left[\frac{-1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i\right] \text{ với } i = 1,2$$

Áp dụng với kết quả sau:

$$R_1: \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{c(1|2)}{c(2|1)}\frac{p_2}{p_1}$$

$$R_2: \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \frac{c(1|2)}{c(2|1)}\frac{p_2}{p_1}$$

IV. Đánh giá quy tắc phân loại

Để đánh giá hiệu suất của quy trình phân loại mẫu là tốt hay không, chúng ta có thể kiểm tra "error rates", nói cách khác là tính xác suất phân loại sai (misclassification probabilities)

Trong thực tế, khó mà biết được các thông số của các population. Do đó, chúng ta sẽ ước tính "error rates" từ các dữ liệu được quan sát (observed data)

IV. Actual Error Rate (AER)

Một thước đo hiệu suất có thể được tính cho bất kỳ quy trình phân loại nào là Actual Error Rate (AER) được xác định bằng cách sử dụng các ước tính mẫu.

$$\mathrm{AER}(\hat{R}_1, \hat{R}_2) = p_1 \int_{\hat{R}_2} f_1(\mathbf{x})d\mathbf{x} + p_2 \int_{\hat{R}_1} f_2(\mathbf{x})d\mathbf{x}$$

IV. Apparent Error Rate(APER)

APER (Apparent Error Rate) được đề xuất như một cách tính AER đơn giản và trực quan.

IV. Apparent Error Rate (APER)

Ta có: n_1, n_2 lần lượt là tổng số phần tử được phân lớp vào π_1, π_2

		Predict	
		π_1	π_2
Actual	π_1	n_{1C}	$n_{1M} = n_1 - n_{1C}$
	π_2	$n_{2M} = n_2 - n_{2C}$	n_{2C}

n_{1C}: Số phần tử lớp π_1 được phân loại đúng vào lớp π_1

n_{1M}: Số phần tử lớp π_1 bị phân loại sai vào lớp π_2

n_{2C}: Số phần tử lớp π_2 được phân loại đúng

n_{2M}: Số phần tử lớp π_2 bị phân loại sai

IV. Apparent Error Rate (APER)

Khi đó thước đo apparent error rate (APER) được tính như sau:

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2}$$

APER có ưu điểm là khá dễ tính toán và trực quan.

Tuy nhiên, Do việc đánh giá lại dựa trên các dữ liệu dùng xây dựng model nên nó có xu hướng bị đánh giá thấp trong thực tế khi phân loại đối tượng mới.

IV. Phương pháp Lachenbruch's holdout

Vì vậy người ta đề xuất một phương pháp mới có tên: Lachenbruch's holdout .

Ý tưởng:

Xem xét một mẫu thử nghiệm độc lập và các đối tượng mới mà ta biết nhãn lớp thực sự. Điều này có nghĩa là ta chia mẫu ban đầu trong một mẫu đào tạo và mẫu thử nghiệm.

AER sau đó được ước tính bằng tỷ lệ của các đối tượng phân loại sai trong mẫu thử trong khi mẫu đào tạo được sử dụng để xây dựng quy tắc phân loại (model).

IV. Phương pháp Lachenbruch's holdout

Quy trình thực hiện như sau:

- Lấy một đối tượng ra khỏi mẫu và xây dựng quy tắc phân loại dựa trên n - 1 đối tượng còn lại.
- Phân loại đối tượng lấy ra dựa trên quy tắc phân loại được xây dựng ở bước 1.
- Lặp lại hai bước trước cho từng đối tượng trong mẫu.
- Gọi n_{1M}^{H} ,n_{2M}^{H} là số lượng các quan sát được phân loại sai trong lớp π_1 và π_2 tương ứng. Khi đó, AER sẽ được tính như sau:

$$\hat{E}(\mathrm{AER}) = \frac{n_{1M}^{(H)} + n_{2M}^{(H)}}{n_1 + n_2}$$

V. Bài toán phân nhiều lớp

5.1) Phân loại cho nhiều quần thể

Phần này trình bày về cách phân lớp cho g population với $g \geq 2$.

Gọi $f_i(x)$ là hàm mật độ xác suất của population π_i với i = 1, 2, 3, ... g, ta có:

- p_i : prior probability của population π_i với i = 1, 2, 3, ... g
- c(k,i): chi phi cho việc phân lớp 1 item vào π_k, Trong khi nó thuộc về π_i, với i,k=1,2,3,...,g. Với k=i => c(k|i)=0
- Với các giá trị x thuộc về R_k được phân lớp vào π_k và ta có:

P(k|i) = P(Phân lớp item vào lớp $\boldsymbol{\pi_k}|\boldsymbol{\pi_i}$)=$\int_{R_k} f_i(\mathbf{x})d\mathbf{x}$ với i,k=1,2,3,…,g và

$$P(i|i) = 1 - \sum_{i=1,k!=i}^{g} P(k|i)$$

V. Bài toán phân nhiều lớp

5.1) Phân loại cho nhiều quần thể

Như ở chương III, mục đích chúng ta vẫn là tìm một độ lỗi sao cho khi phân lớp đối tượng x thì độ lỗi này nhỏ nhất. Ta sẽ dùng lại ECM như chương III.

Xét chi phí phân lớp sai có điều kiện một giá trị x từ π_1 vào π_2, π_3, ...π_g là:

$$ECM(1)=c(2|1)P(2|1)+c(3|1)P(3|1)+\ldots+c(g|1)P(g|1)=\sum_{k=2}^{g} P(k|1)c(k|1)$$

V. Bài toán phân nhiều lớp

5.1) Phân loại cho nhiều quần thể

Tương tự có được ECM(2), ...,ECM(g). Ta có được ECM(i) khi phân lớp vào lớp π_i:

$$ECM(i) = \sum_{k!=i}^{g} c(k|i)P(k|i)$$

Với mỗi ECM(i) ta có được prior probability pi, nên ta có được ECM tổng thể:

$$ECM = \sum_{i=1}^{g} p_i ECM(i) = \sum_{i=1}^{g} p_i \left(\sum_{k!=i}^{g} c(k|i)\, P(k|i) \right)$$

V. Bài toán phân nhiều lớp

<u>5.1) Phân loại cho nhiều quần thể</u>

Để cực tiểu hóa ECM thì cần phải phân lớp x vào population π_k với k=1,2,3,...g sao cho

$$\sum_{i=1, i!=k}^{g} p_i \, f_i(x) \, c(k|i) \text{ nhỏ nhất}$$

Giả sử, c(k|i) như nhau. ECM được cực tiểu hóa khi

$$\sum_{i=1, i!=k}^{g} p_i \, f_i(x) \text{ nhỏ nhất}$$

V. Bài toán phân nhiều lớp

5.1) Phân loại cho nhiều quần thể

Do đó để phân lớp x_0 vào lớp π_k thì:

$$p_k f_k(\mathbf{x}_0) > p_i f_i(\mathbf{x}_0) \ \textbf{với} \ i != k$$

Tương đương:

$$\ln(p_k f_k(\mathbf{x}_0)) > \ln(p_i f_i(\mathbf{x}_0))\ \textbf{với} \ i != k$$

V. Bài toán phân nhiều lớp

5.2) Phân loại cho nhiều quần thể đa biến

Ta có:

$$f_i(x) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}_i|^{1/2}} exp\left[\frac{-1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right] \text{ với } i = 1,2,3\ldots,g$$

Vậy để phân lớp x_0 vào lớp π_k thì:

$$ln(p_k f_k(\mathbf{x})) = \boldsymbol{ln}(p_k) - \frac{p}{2}ln(2\pi) - \frac{1}{2}ln(\boldsymbol{\Sigma}_k) - \left[\frac{-1}{2}(\mathbf{x}-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}-\boldsymbol{\mu}_k)\right] \text{ lớn nhất}$$

V. Bài toán phân nhiều lớp

5.2) Phân loại cho nhiều quần thể đa biến

Do p/2 ln(2π) là hằng số nên:

$$ln(p_k f_k(\mathrm{x})) = -\frac{1}{2} ln(\boldsymbol{\Sigma}_k) - \left[\frac{-1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k{}^{-1}(\mathbf{x} - \boldsymbol{\mu}_k)\right] + \boldsymbol{ln}(p_k) \text{ lớn nhất}$$

$$max\left(-\frac{1}{2} ln(\boldsymbol{\Sigma}_i) - \left[\frac{-1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i{}^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right] + \boldsymbol{ln}(p_i)\right) \text{ với } i = 1,2,3\ldots, g$$

V. Bài toán phân nhiều lớp

5.2) Phân loại cho nhiều quần thể đa biến

Trường hợp: μ_i và $\sum_i$ chưa biết, ta vẫn có tập training set X_i tương ứng với populations π_i với i=1,2,3,..g. Khi đó ta có được các thông tin sau:

$$\bar{x}_i = \textit{sample mean vector}$$

$$S_i = \textit{sample covariance matrix}$$

$$n_i = \textit{kích thước sample}$$

V. Bài toán phân nhiều lớp

5.2) Phân loại cho nhiều quần thể đa biến

Xét trường hợp $\sum_i = \sum$

$$d_i^Q(\mathbf{x}) = -\frac{1}{2}ln(\boldsymbol{\Sigma}) - \frac{1}{2}\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x} + \boldsymbol{\mu}_i{}^T\boldsymbol{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i{}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i + \boldsymbol{ln}(p_i) \text{ với } i = 1,2,3,\ldots,g$$

Nhận thấy tất cả $-\frac{1}{2}ln(\boldsymbol{\Sigma}) - \frac{1}{2}\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x}$ đều giống nhau trong mọi d

=> linear discriminant score có dạng $d_i(x) = \boldsymbol{\mu}_i{}^T\boldsymbol{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i{}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i + \boldsymbol{ln}(p_i)$

V. Bài toán phân nhiều lớp

5.2) Phân loại cho nhiều quần thể đa biến

Do μ_i và $\sum_i$ ở linear discriminant score đều chưa biết nên chưa biết nên sẽ được thay thế.

Biến đổi thành=>

$$\hat{d}_i(x) = \overline{\mathbf{x}_i}^T \boldsymbol{S}_{\boldsymbol{pooled}}{}^{-1}\mathbf{x} - \frac{1}{2}\overline{\mathbf{x}_i}^T \boldsymbol{S}_{\boldsymbol{pooled}}{}^{-1}\overline{\mathbf{x}_i} + \boldsymbol{ln}(p_i)$$

Với i =1, 2, ... g

Trong đó:

$$\boldsymbol{S}_{\boldsymbol{pooled}} = \frac{1}{n_1 + n_2 + \ldots + n_g - g}\left((n_1 - 1)S_1 + (n_2 - 1)S_2 + \ldots + (n_g - 1)S_g\right)$$

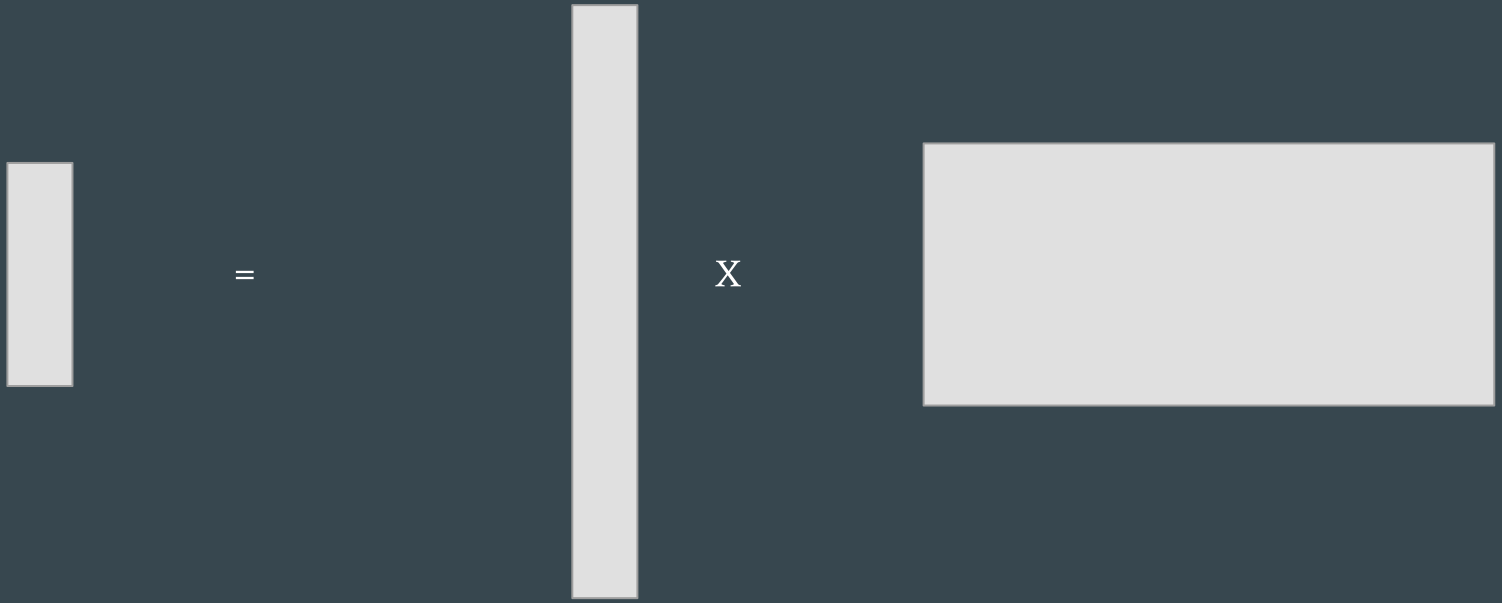
V. Bài toán phân nhiều lớp

5.2) Phân loại cho nhiều quần thể đa biến

Vậy để phân lớp x_0 vào lớp π_k thì:

$$\widehat{d_k}(x_0) = lớn\ nhất\ trong\ \widehat{d_1}(x_0), \widehat{d_2}(x_0) \dots, \widehat{d_k}(x_0)$$

LDA/Fisher - Giảm chiều



Ma trận ít dòng ít cột (Y)

Ma trận ít cột (w^T cần tìm)

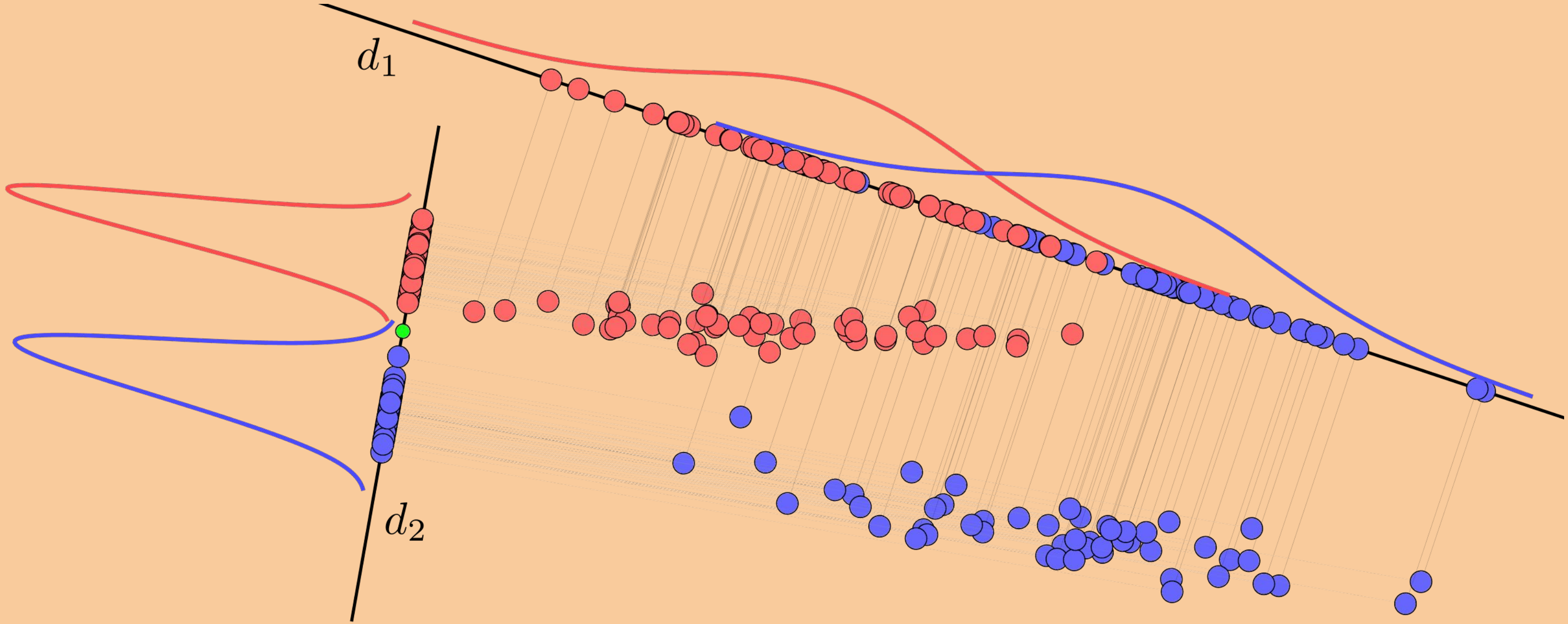
Ma trận ít dòng (X đang có/đề cho)

So sánh PCA và LDA/Fisher

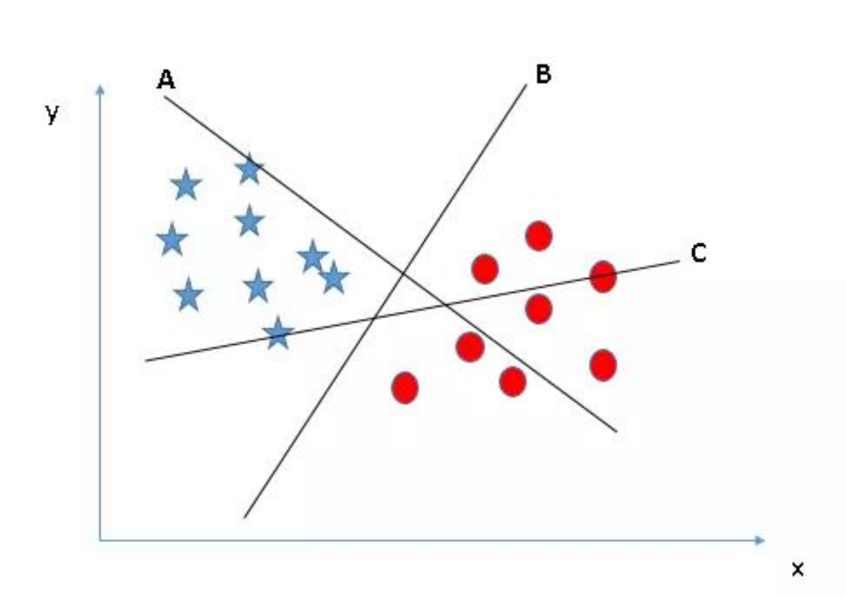
	LDA	PCA
Mô hình học	Có giám sát - là thuật toán tiên đoán nhãn cho dữ liệu mới dựa trên tập huấn luyện (các mẫu trong tập này đều đã được gán nhãn)	Học không giám sát - là thuật toán tiên đoán nhãn cho dữ liệu mới dựa trên tập huấn luyện (các mẫu trong tập này đều chưa được gán nhãn)
Xét phương sai	Có	Không
Giảm chiều	Có	Có
Số lượng thông tin	Cần thiết	Nhiều nhất
Kết quả phân loại	Tốt hơn	..

So sánh PCA và LDA/Fisher

d1 PCA ,d2 LDA



Chọn đường A. Hãy cùng tìm hiểu vì sao LDA lại chọn đường A thay vì B, C



LDA/Fisher cho 2 quần thể

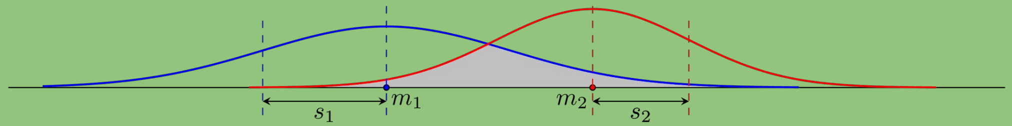
Độ rộng của mỗi đường hình chuông thể hiện độ lệch chuẩn của dữ liệu.

Dữ liệu “bu lại” thì độ lệch chuẩn nhỏ, dữ liệu “lệch ra xa” thì độ lệch chuẩn lớn.

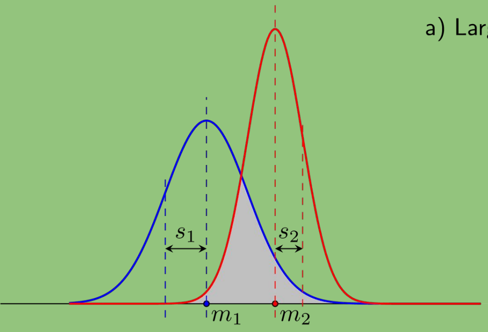
Mục tiêu: Thuật toán có giám sát đi tìm phép chiếu từ không gian đặc trưng này sang không gian đặc trưng khác có số chiều nhỏ hơn mà vẫn giữ được tính khả tách của dữ liệu.

LDA/Fisher cho 2 quần thể - tính khả tách

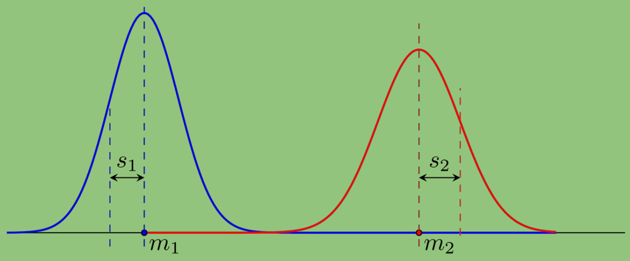
Nhưng độ lệch chuẩn nhỏ trong mỗi class chưa thể hiện độ khả tách của dữ liệu (hình b), mà đòi hỏi thêm khoảng cách 2 kì vọng phải lớn (hình c).



a) Large $(m_1 - m_2)^2$, large $s_1^2 + s_2^2$



b) Small $(m_1 - m_2)^2$, small $s_1^2 + s_2^2$



c) Large $(m_1 - m_2)^2$, small $s_1^2 + s_2^2$

Mong muốn:

- Between-class: khoảng cách 2 kì vọng đủ lớn (2 cái chuông xa nhau)
- Within-class: phương sai s đủ nhỏ (Dữ liệu “bu lại” centroid)

LDA/Fisher cho 2 quần thể- Nguyên lí

LDA(w)= $\dfrac{(m_1 - m_2)^2}{s_1^2 + s_2^2}$ maximum

LDA/Fisher

B1: Tính ma trận within s_w

B2: Tính ma trận between s_B

B3 Tìm vector phép chiếu tốt nhất (tính trị riêng, tính vector riêng, tính vector cơ sở theo vector riêng) $s_w^{-1}.s_B.w = \lambda.w$

B4: Giảm số chiều $y = w^T x$

LDA/Fisher cho 2 quần thể - Công thức

chain rule cho đạo hàm nhiều biến nếu ma trận A đối xứng: $\nabla_{\mathbf{w}} \mathbf{w}\mathbf{A}\mathbf{w} = 2\mathbf{A}\mathbf{w}$

N: số điểm dữ liệu 2 class

w^T : phép chiếu / phép biếnđổi tuyến tính (vector 1 dòng)

đẳng thức: $(\mathbf{a}^T\mathbf{b})^2 = (\mathbf{a}^T\mathbf{b})(\mathbf{a}^T\mathbf{b}) = \mathbf{a}^T\mathbf{b}\mathbf{b}^T\mathbf{a}$

N_1 Điểm đầu tiên của class 1

$N_2 = N - N_1$ Điểm cuối cùng class 2

C_1 Các điểm thuộc class 1

C_2 Các điểm thuộc class 2

$y_n = \mathbf{w}^T\mathbf{x}_n, 1 \leq n \leq N$ Mỗi điểm sau khi chiếu

Vector kỳ vọng của mỗi class:

$$\mathbf{m}_k = \frac{1}{N_k}\sum_{n\in\mathcal{C}_k}\mathbf{x}_n, \quad k = 1, 2$$

Khoảng cách 2 chuông:

$$m_1 - m_2 = \frac{1}{N_1}\sum_{i\in\mathcal{C}_1} y_i - \frac{1}{N_2}\sum_{j\in\mathcal{C}_2} y_j = \mathbf{w}^T(\mathbf{m}_1 - \mathbf{m}_2)$$

Mức “bu lại” của mỗi class:

$$s_k^2 = \sum_{n\in\mathcal{C}_k}(y_n - m_k)^2, \quad k = 1, 2$$

LDA/Fisher cho 2 quần thể

Between

$$[w^T(m_1 - m_2)]^2$$

$$= \mathbf{w}^T \underbrace{(\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T}_{\mathbf{S}_B} \mathbf{w} = \mathbf{w}^T \mathbf{S}_B \mathbf{w}$$

Within

$$s_1^2 + s_2^2 = \sum_{k=1}^{2} \sum_{n \in \mathcal{C}_k} \left(\mathbf{w}^T(\mathbf{x}_n - \mathbf{m}_k)\right)^2 = \mathbf{w}^T \underbrace{\sum_{k=1}^{2} \sum_{n \in \mathcal{C}_k} (\mathbf{x}_n - \mathbf{m}_k)(\mathbf{x}_n - \mathbf{m}_k)^T}_{\mathbf{S}_W} \mathbf{w} = \mathbf{w}^T \mathbf{S}_W \mathbf{w}$$

=> w*

$$= \arg\max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}$$

LDA/Fisher cho 2 quần thể - Tìm phép chiếu w đạt mong muốn

Cho (w*)' = 0:

$$\frac{1}{(\mathbf{w}^T\mathbf{S}_W\mathbf{w})^2}\left(2\mathbf{S}_B\mathbf{w}(\mathbf{w}^T\mathbf{S}_W\mathbf{w}) - 2\mathbf{w}^T\mathbf{S}_B\mathbf{w}^T\mathbf{S}_W\mathbf{w}\right) = \mathbf{0}$$

$$\Leftrightarrow \mathbf{S}_B\mathbf{w} = \frac{\mathbf{w}^T\mathbf{S}_B\mathbf{w}}{\mathbf{w}^T\mathbf{S}_W\mathbf{w}}\mathbf{S}_W\mathbf{w}$$

$$\Leftrightarrow \mathbf{S}_W^{-1}\mathbf{S}_B\mathbf{w} = Lw$$

LDA(w) là số vô hướng => w là vector riêng của $\mathbf{S}_W^{-1}\mathbf{S}_B$ ứng với trị riêng L mà L = w*. Mong muốn max => w* max => L là trị riêng max của $\mathbf{S}_W^{-1}\mathbf{S}_B$

LDA/Fisher - Tiêu chí để xét cá thể mới thuộc lớp nào

Chiếu x_{new}, y_{new} xuống w^T :

$$new = w^T . \begin{bmatrix} x_{new} \\ y_{new} \end{bmatrix}$$

Tính d1 khoảng cách giữa 'cá thể mới' sau khi chiếu với m1:

$$d_1 = |new - m_1|$$

Tính d2 khoảng cách giữa 'cá thể mới' sau khi chiếu với m2:

$$d_2 = |new - m_2|$$

Khoảng cách nào bé hơn thì thuộc vào lớp đó:

d1 < d2 : R1

d2 < d1 : R2

Ví dụ

$$X_1 = \{(4,1),(2,4),(2,3),(3,6),(4,8)\}$$
$$X_2 = \{(9,10),(6,8),(9,5),(8,7),(10,8)\}$$

Ví dụ -B1: $S_W = S_1 + S_2$

$$S_1 = \sum_{x \in C_1} (x - \mu_1)(x - \mu_1)^\top$$

$$\mu_1 = \left\{ \frac{4+2+2+3+4}{5}, \frac{1+4+3+6+4}{5} \right\}$$
$$\mu_1 = \begin{bmatrix} 3.00 & 8.60 \end{bmatrix}$$

$$(\chi_1 - \mu_1) = \begin{bmatrix} 1. & -1 & -1 & 0 & 1 \\ -2.6 & 0.4 \cdot -0.6 & 2.4 & 0.4 & \end{bmatrix}$$

$$x_{1,1}.x_{1,1}^T = \begin{bmatrix} 1 \\ -2.6 \end{bmatrix} \cdot \begin{bmatrix} 1 & -2.6 \end{bmatrix} = \begin{bmatrix} 1 & -2.6 \\ -2.6 & 6.76 \end{bmatrix}$$

$$x_{1,2}.x_{1,2}^T = \begin{bmatrix} -1 \\ 0.4 \end{bmatrix} \begin{bmatrix} -1 & 0.4 \end{bmatrix} = \begin{bmatrix} 1 & -0.4 \\ -0.4 & 0.16 \end{bmatrix}$$

$$x_{1,3}.x_{1,3}^T = \begin{bmatrix} -1 \\ -0.6 \end{bmatrix} \begin{bmatrix} -1 - 0.6 \end{bmatrix} = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 0.36 \end{bmatrix}$$

$$x_{1,4}.x_{1,4}^T = \begin{bmatrix} 0 \\ 2.4 \end{bmatrix} \begin{bmatrix} 0 & 2.4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 5.76 \end{bmatrix}$$

$$x_{1,5}.x_{1,5}^T = \begin{bmatrix} 1 \\ 0.4 \end{bmatrix} \begin{bmatrix} 1 & 0.4 \end{bmatrix} = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 0.16 \end{bmatrix}$$

$$S_1 = \begin{bmatrix} 0.8 & -0.4 \\ -0.4 & 2.6 \end{bmatrix}$$

$$\mu_2 = \begin{bmatrix} 8.4 & 7.60 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 1.84 & -0.04 \\ -0.04 & 2.64 \end{bmatrix}$$

$$S_W = S_1 + S_2.$$

$$S_W = \begin{bmatrix} 2.64 & -0.44 \\ -0.44 & 6.28 \end{bmatrix}$$

Ví dụ - B2: Tìm S_B

$$\begin{aligned} S_B =& (\mu_1 - \mu_2)(\mu_1 - \mu_2)^\top \\ =& \begin{pmatrix} -5.4 \\ -4 \end{pmatrix} \begin{pmatrix} -5.4 & -4 \end{pmatrix} = \begin{pmatrix} 29.16 & 21.6 \\ 21.6 & 16.00 \end{pmatrix} \end{aligned}$$

Ví dụ - B3: Tìm vector chiếu thoả 2 mong muốn

$$S_W^{-1} S_B W = \lambda W \quad (*)$$

$$\left| S_W^{-1} S_B - \lambda I \right| = 0$$

I=[1 0 0
0 1 0
0 0 1]

$$\begin{vmatrix} 11.89 - \lambda & 8.81 \\ 5.08 & 3.76 - \lambda \end{vmatrix} = 0$$

$$\lambda = 15.65$$

CHỌN LAMDA LỚN NHẤT

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$S_W = \begin{bmatrix} 2.64 & -0.44 \\ -0.44 & 5.28 \end{bmatrix}$$

$$S_W^{-1} = \frac{1}{13.74} \begin{bmatrix} 5.28 & 0.44 \\ 0.44 & 2.64 \end{bmatrix} = \begin{bmatrix} 0.384 & 0.032 \\ 0.032 & 0.122 \end{bmatrix}$$

(*) =>

$$\begin{bmatrix} 11.89 & 8.81 \\ 5.08 & 3.76 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 15.65 \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 0.91 \\ 0.39 \end{bmatrix}$$

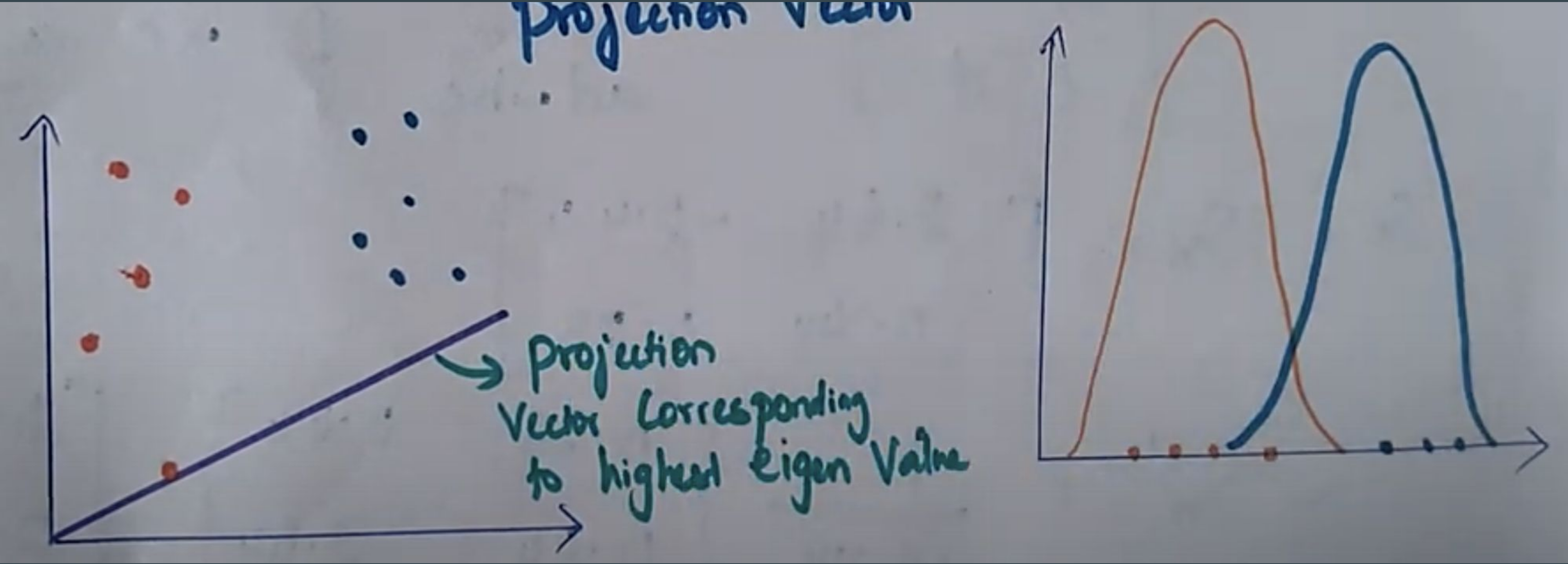
Ví dụ - B4: Giảm chiều

$$y = W^\top x$$

Y1 = $\begin{pmatrix} 0.91 & 0.39 \end{pmatrix} \cdot \begin{pmatrix} 4 & 2 & 2 & 3 & 4 \\ 1 & 4 & 3 & 6 & 8 \end{pmatrix} = \begin{pmatrix} \frac{403}{100} & \frac{169}{50} & \frac{299}{100} & \frac{507}{100} & \frac{169}{25} \end{pmatrix}$

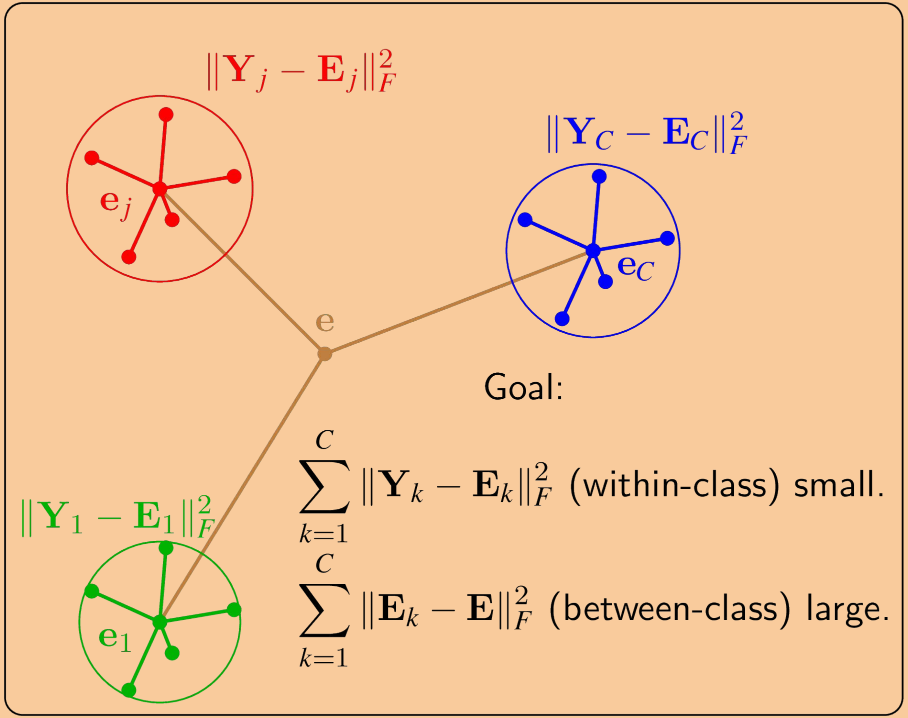
Y2 = $\begin{pmatrix} 0.91 & 0.39 \end{pmatrix} \cdot \begin{pmatrix} 9 & 6 & 9 & 8 & 10 \\ 10 & 8 & 5 & 7 & 8 \end{pmatrix} = \begin{pmatrix} \frac{1209}{100} & \frac{429}{50} & \frac{507}{50} & \frac{1001}{100} & \frac{611}{50} \end{pmatrix}$

Ví dụ - Vẽ hình


Projection Vector
Projection Vector Corresponding to highest Eigen Value

LDA/Fisher cho nhiều quần thể

Xem thêm tại báo cáo đính kèm



XIN CÁM ƠN MỌI NGƯỜI ĐÃ DÀNH THỜI GIAN THEO DÕI