Ahora si los valores están en la frontera:

$$Fr(\text{decisión}) = \{ x \in \mathbb{R}^n \mid f_1(x)\Pi_1 = f_2(x)\Pi_2 \}$$ Si esto pasa, se echa un volado para ver si va a Π_1 ó Π_2

LDA (caso homogéneo) $\quad (\Sigma_1 = \Sigma_2)$

$$f_1(x) \sim N_n(\mu_1, \Sigma_1) \quad \wedge \quad f_2(x) \sim N_n(\mu_2, \Sigma_2)$$

Tarea: 27 de marzo

Expo: Ari ⟶ 23 marzo

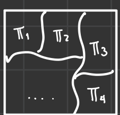
Naïve Bayes y su implementación en Python

jueves 16 marzo

Análisis lineal por discriminantes (LDA)

Método de reducción en dimensión y es un clasificador de variables

Habíamos definido las probabilidades:



(i) $C = \bigcup_{i=1}^{n} \Pi_i$ ajena $\qquad \Pi_i$ es un conjunto (el i-ésimo grupo)

$$\Pi_i \cap \Pi_j = \Phi \quad \forall i \neq j$$

es decir, $\{\Pi_i\}_{i=1}^{n}$ forma una partición de C

(ii) Sea $X = (X_1, \dots, X_n)$ un vector aleatorio

$$\{ \pi_i = \mathbb{P}(X \in \Pi_i) \}_{i=1}^{n}$$

Probabilidades a posteriori

$$\mathbb{P}(X = x \mid X \in \Pi_i) = f_i(x)$$

a priori (Regla de Bayes)

$$\mathbb{P}(\Pi_i \mid x) = \mathbb{P}(X \in \Pi_i = X = x)$$

Por Bayes (Proba I)

$$\boxed{P(\Pi_i \mid x) = \frac{f_i(x)\pi_i}{\sum_{i=1}^{n} f_i(x)\pi_i}}$$

como tener...

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}$$

Regla de decisión: para $i \neq j$

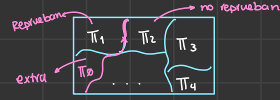
$$\underbrace{\frac{P(\Pi_i \mid x)}{P(\Pi_j \mid x)}}_{\text{estadístico "odd"}} = \frac{f_i(x)\pi_i}{f_j(x)\pi_j} > 1 \qquad x \rightsquigarrow \Pi_i$$

tienen el mismo denominador

en otro caso $\quad x \rightsquigarrow \Pi_j$

Si estamos $R = \{ x \in \mathbb{R}^n \mid f_i(x)\pi_i = f_j(x)\pi_j \}$, hacemos un volado para decidir → **ejemplo:**

ejemplo: Leo está en la intersección

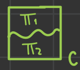


se hecha un volado para ver a donde va → R ó NR

LDA (Fisher)

Hipótesis: $f_i(\cdot)$ es la densidad de una v.a. $N(\mu_i, \Sigma_i)$

Caso $i = 1,2$ $\quad \pi_1, \pi_2$



$$f_i(x) = \frac{1}{\sqrt{(2\pi)^n |\det \Sigma_i|}} \exp\left\{ -\frac{1}{2}(x-\mu_i)^t \Sigma_i^{-1} (x-\mu_i) \right\}$$

Hipótesis: Caso homogénea $\Sigma_1 = \Sigma_2$

$$\frac{P(\pi_1|x)}{P(\pi_2|x)} = \frac{e^{-\frac{1}{2}(x-\mu_1)^t \Sigma^{-1}(x-\mu_1)}}{e^{-\frac{1}{2}(x-\mu_2)^t\Sigma^{-1}(x-\mu_2)}} \frac{\pi_1}{\pi_2}$$

log inyectiva

$$\log\left(\frac{P(\pi_1|x)}{P(\pi_2|x)}\right) = -\frac{1}{2}(x-\mu_1)^t\Sigma^{-1}(x-\mu_1) + \frac{1}{2}(x-\mu_2)^t\Sigma^{-1}(x-\mu_2) + \log\left(\frac{\pi_1}{\pi_2}\right)$$

"$=$" $b_\emptyset + bx$, b y x son vectores

$$(1) = \underbrace{(\mu_1-\mu_2)}_{\text{vector de medias}}{}^t\Sigma^{-1}x - \frac{1}{2}\Sigma^{-1}(\mu_1+\mu_2)(\mu_1-\mu_2)^t$$

$$x_1 = (x_{11}, \dots, x_{1n})$$

$$\mu_1 = (\mathbb{E}(x_{11}), \dots, \mathbb{E}(x_{1n}))$$

$(1) = b_\emptyset + bx$ donde: $b_\emptyset = -\frac{1}{2}\left(\mu_1^t\Sigma^{-1}\mu_1 - \mu_2^t\Sigma^{-1}\mu_2\right) + \log\left(\frac{\pi_1}{\pi_2}\right)$

$$b = \Sigma^{-1}(\mu_1-\mu_2)$$

$$\boxed{\log\left(\frac{P(\pi_1|x)}{P(\pi_2|x)}\right) = b_\emptyset + b^t_x}$$

... **LDA lineal**

¿Que programar LDA?

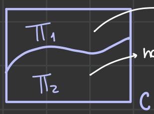
Esquema:

2 clase
- clase 1
- clase 2

Matriz de <u>dispersión</u>

S_W, S_B se tienen qué mostrar las proyecciones

<u>2 grupos</u>



hay n_1 elementos

hay n_2 elementos

$X = (X_1, \ldots, X_n)$

cardinalidad

Suponer $n_1 = \# \Pi_1$

$n_2 = \# \Pi_2$

$n = n_1 + n_2$

Recordemos que una proyección de una variable x a una variable y : es un valor w tal que

$$y = w^t x$$

$w = (w_1, \ldots, w_n)$ $\qquad$ $x = (x_1, \ldots, x_n)$

$$\left.\begin{aligned} \Pi_1 &= \{ x_1, \ldots, x_{n_1} \} \\ \Pi_2 &= \{ x_1^2, \ldots, x_{n_2}^2 \} \end{aligned}\right\} \text{Calcular la media de c/uno de los puntos}$$

Calcular la media

$i = 1,2 \qquad \mu_i = \frac{1}{n_i} \sum_{x \in \Pi_i} x$

$$\tilde{\mu}_i = \frac{1}{n_i} \sum_{y \in \Pi_i} y$$

NOTA :

$$\tilde{\mu}_i = \frac{1}{n_i} \sum_{y \in \Pi_i} w^t x = \frac{1}{n_i} w^t \sum_{x \in \Pi_i} x = w^t \frac{1}{n_i} \sum_{x \in \Pi_i} x = w^t \mu_i$$

$$\boxed{\therefore \tilde{\mu}_i = w^t \mu_i}$$

Se puede tener: la distancia ... ?

$$J(w) = |\tilde{\mu}_1 - \tilde{\mu}_2| = |w^t \mu_1 - w^t \mu_2| = |w^t (\mu_1 - \mu_2)|$$

↓

distancia entre las medias proyectadas

Definición: (Scatter)

Con las notaciones anteriores

$$\tilde{S}_i^2 = \sum_{y \in \Pi_i} (y - \tilde{\mu}_i)^2 \quad \Big\} \text{ Variabilidad de los datos despuès de ser proyectados}$$

($\boxed{\text{NOTA}}$ Varianza para los valores proyectados)

Vamos a definir la i-ésima matriz scatter como:

$$S_i = \sum_{x \in \Pi_i} (x - \mu_i)^t (x - \mu_i) \qquad i = 1,2$$

$\boxed{S_W = S_1 + S_2}$ matriz scatter within y la distancia entre las medias proyectadas normalizadas por la variabilidad de las medias proyectadas

$$J(w) = \frac{|w^t(\mu_1-\mu_2)|}{\tilde{S}_1^2 + \tilde{S}_2^2}$$ .... suma de varianzas

De aquí;

$$\tilde{S}_i^2 = \sum_{x \in \Pi_i} (w^t x - w^t \mu_i)^2$$

$$= \sum_{x \in \Pi_i} (w^t(x-\mu_i))^2 = \sum_{x \in \Pi_i} w^t \underbrace{(x-\mu_i)(x-\mu_i)^t}_{} w = w^t S_i w$$

$$\therefore \tilde{S}_i^2 = w^t S_i w$$

La matriz scatter es:

$$S_B = (\mu_1-\mu_2)(\mu_1-\mu_2)^t$$ ... llamada: matriz scatter between

Así:

$$J(w) = \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|}{\tilde{S}_1^2 + \tilde{S}_2^2} = \frac{|w^t(\mu_1-\mu_2)|}{w^t S_w w}$$

elevando al cuadrado

$$|\tilde{\mu}_1 - \tilde{\mu}_2|^2 = w^t(\mu_1-\mu_2)(\mu_1-\mu_2)^t w$$

Finalmente,

$$\underbrace{J(w) = \frac{w^t S_B w}{w^t S_w w}}_{\text{parte empírica}} \longleftrightarrow \underbrace{\log\left(\frac{P(\Pi_1|x)}{P(\Pi_2|x)}\right)}_{\text{parte teórica}}$$

¿Como obtener a w?

$$\frac{dJ(w)}{dw} = \emptyset$$ equivalentemente $$\emptyset = (w^t S_w w)\frac{dw^t S_B w}{dw} - (w^t S_B w)\frac{dw^t S_w w}{dw}$$

$$\emptyset = (w^t S_w w) 2 S_B w - (w^t S_B w) 2 S_w w$$

$$\emptyset = S_B w - J(w) \cdot S_w w \Longrightarrow J(w) S_w w = S_B w$$

multiplicando por S_w^{-1}

$$\longrightarrow \quad J(w) w = S_w^{-1} S_B w$$ es una medida (es un número)

Teorema (Fisher)

Con las notaciones anteriores

$$\boxed{(S_w^{-1} S_B) w = J(w) \cdot w}$$ satisface a los eigenvalores

$J(w)$ es eigenvalor de $S_w^{-1} S_B$ ■

De hecho,

$S_x = S_w^{-1} S_B$ ... llamada la matriz de covarianzas