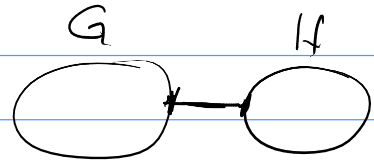
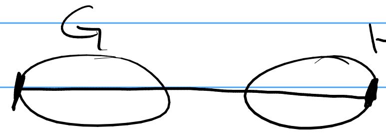
dendograma

Single linkage



$$d_{SL}(G,H) = \min_{\substack{i\in G\\ i'\in H}} d_{i,i'}$$

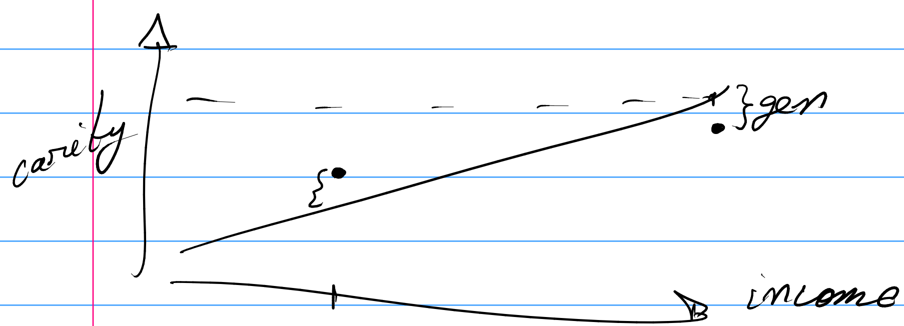
Complete Linkage



$$d_{CL}(G,H) = \max_{\substack{i\in G\\ i'\in H}} d_{i,i'}$$

Group Average

$$d_{Ga}(G,H) = \frac{1}{N_G N_H} \sum_{i\in G} \sum_{i'\in H} d_{i,i'}$$



Clasificación

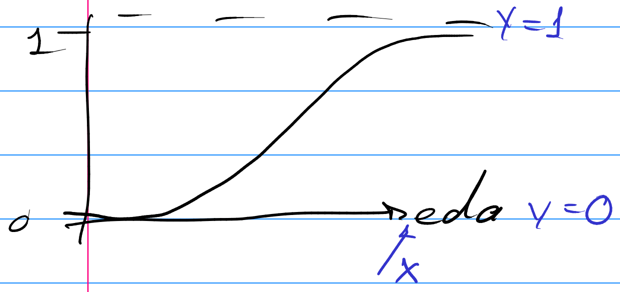
LDa (Linear Discriminant Analysis)

Análisis Discriminante Lineal

Clasificador Óptimo de Bayes:

$$g^* \in \underset{g: \mathcal{X} \to \mathcal{Y}}{\operatorname{argmin}} \ \mathbb{P}(g(X) \neq Y \mid X)$$

i.e., que minimiza la probabilidad posterior de equivocarse



Edad $\longrightarrow$ Coronarios

$\longleftarrow$ $\mathbb{P}$

Supongamos que $Y = k \Rightarrow$

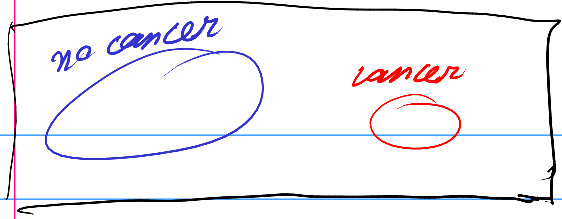
$$g^* \in \underset{g: \mathcal{X} \to \mathcal{Y}}{\operatorname{argmax}} \ \mathbb{P}(g(x) = k \mid X = x)$$

teorema de Bayes

$$\mathbb{P}(g(x) = k \mid X = x) = \frac{\mathbb{P}(X = x \mid Y = k)\, \mathbb{P}(Y = k)}{\mathbb{P}(X = x)}$$

i.e. $\underbrace{\mathbb{P}(g(x) = k \mid X = x)}_{\text{posterior}} \propto \underbrace{\mathbb{P}(X = x \mid Y = k)}_{\text{verosimilitud}} \underbrace{\pi_k}_{\text{previa}} \quad \cdots \ (*)$

De aquí que $\mathbb{P}(Y = k \mid X = x)$ se le llama la proba. posterior y a g^* por eso se le llama clasificador bayesiano óptimo



Regresión logística modela directamente la proba. posterior, en particular si usamos la liga logit:

$$\log \frac{\mathbb{P}(Y=k \mid X=x)}{\mathbb{P}(Y=K \mid X=x)} = x^T \vec{\beta}_k$$

$$\log \frac{\mathbb{P}(Y=1 \mid X=x)}{1-\mathbb{P}(Y=1 \mid X=x)} = x^T \vec{\beta}$$

Pero podríamos modelar el lado derecho

π_k: Proba. previa de que $Y \in$ Pob.$_k$

$\mathbb{P}(X=x \mid Y=k)$: Verosimilitud
$= f_k(x)$

$\phi_k(x; \mu_k, \Sigma_k)$ Mezcla gaussiana

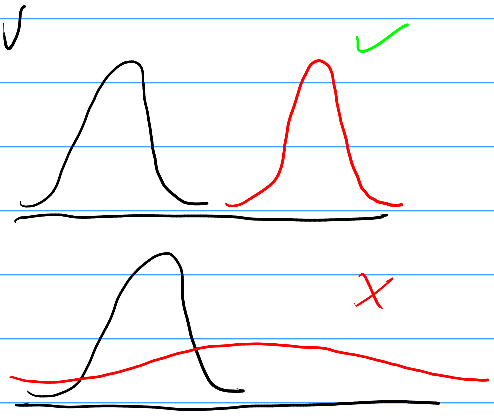
$$\mathbb{P}(Y=k \mid X=x) = \frac{\mathbb{P}(X=x \mid Y=k)\,\mathbb{P}(Y=k)}{\mathbb{P}(X=x)}$$

$$= \frac{\mathbb{P}(X=x \mid Y=k)\,\mathbb{P}(Y=k)}{\sum_{l=1}^{K} \mathbb{P}(X=x \mid Y=l)\,\mathbb{P}(Y=l)}$$

$$= \frac{f_k(x)\pi_k}{\sum_{l=1}^{K} f_l(x)\pi_l}$$

Así, $$\frac{\mathbb{P}(Y=k \mid X=x)}{\mathbb{P}(Y=j \mid X=x)} = \frac{f_k(x)}{f_j(x)} \, \frac{\pi_k}{\pi_j} \quad \cdots (**)$$

$$f_i(x) = \mathcal{N}_p(\mu_i, \Sigma) \quad \forall i$$



$$x \subseteq \mathbb{R}^p$$

De $(**)$, vamos a asignar $Y \in$ Pob. k si

$$\frac{\mathbb{P}(Y=k \mid X=x)}{\mathbb{P}(Y=j \mid X=x)} \geq 1 \qquad \forall j \neq k$$

i.e. $$f_k(x)\pi_k \geq f_j(x)\pi_j$$

$$\log f_k(x) + \underline{\log \pi_k} \geq \log f_j(x) + \underline{\log \pi_j}$$

$\Rightarrow$

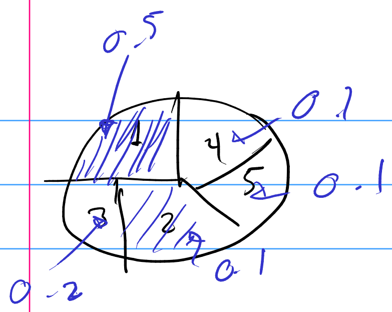
$$\cancel{-\tfrac{1}{2}\log|\Sigma|} - \tfrac{1}{2}(x-\mu_k)^T \Sigma^{-1} (x-\mu_k) + \underline{\log \pi_k}$$

$$\geq \cancel{-\tfrac{1}{2}\log|\Sigma|} - \tfrac{1}{2}(x-\mu_j)^T \Sigma^{-1} (x-\mu_j) + \underline{\log \pi_j}$$

$\Rightarrow$

$$\log \pi_k \cancel{- \tfrac{1}{2} x^T \Sigma^{-1} x} - \tfrac{1}{2}\mu_k^T \Sigma^{-1} \mu_k + x^T \Sigma^{-1} \mu_k$$

$$\geq \log \pi_j \cancel{- \tfrac{1}{2} x^T \Sigma^{-1} x} - \tfrac{1}{2}\mu_j^T \Sigma^{-1} \mu_j + x^T \Sigma^{-1} \mu_j$$



Supuesto $\quad X=x \mid Y=j \sim \mathcal{N}_p(\mu_j, \Sigma)$

$$f_j(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu_j)^T\Sigma^{-1}(x-\mu_j)\right\} \mathbb{1}_{\mathbb{R}^p}(x)$$

$$\propto |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu_j)^T\Sigma^{-1}(x-\mu_j)\right\}$$

$$\log f_j(x) \propto -\frac{1}{2}\log|\Sigma| - \frac{1}{2}(x-\mu_j)^T\Sigma^{-1}(x-\mu_j)$$

i.e.

Asignamos a la población k si:

$$\left(\log\pi_k - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k\right) + \underbrace{x^T\left(\Sigma^{-1}\mu_k\right)}_{(\mu_k^T\Sigma^{-1})x} \geq \left(\log\pi_j - \frac{1}{2}\mu_j^T\Sigma^{-1}\mu_j\right) + x^T\left(\Sigma^{-1}\mu_j\right)$$

$\forall j \neq k$

Sea $\quad w_{0i} = \log\pi_i - \frac{1}{2}\mu_i^T\Sigma^{-1}\mu_i \quad \forall i=1,\ldots,K$

y $\quad w_i^T = \mu_i^T\Sigma^{-1}$

Asignamos a la Pob. k si

$$\boxed{w_{0k} + w_k^T x \geq w_{0j} + w_j^T x} \quad \forall j \neq k$$

LDA

QDA (Quadratic Discriminant Analysis)

$$X = x \mid Y = i \sim N_p(\mu_i, \Sigma_i)$$

Asignamos a la población k si

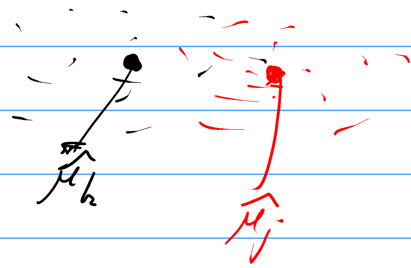
$$-\frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k) + \log \pi_k$$

$$\geq -\frac{1}{2}\log|\Sigma_j| - \frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j) + \log \pi_j$$

Reglas basadas en datos

$$D_n = ((X_1, Y_1), \dots, (X_n, Y_n))$$

$$X_i \in \mathbb{R}^p, \quad i = 1, \dots, n$$



$$n_k = \sum_{i=1}^{n} \mathbb{1}\{Y_i = k\}$$

$$\Rightarrow \hat{\mu}_k = \frac{1}{n_k}\sum_{i=1}^{n} X_i \mathbb{1}\{Y_i = k\}$$

QDA:

$$\hat{\Sigma}_k \equiv S_k = \frac{1}{n_k - 1}\sum_{i=1}^{n}(X_i - \hat{\mu}_k)(X_i - \hat{\mu}_k)^T \mathbb{1}\{Y_i = k\}$$

LDA:

$$\hat{\Sigma} = \frac{1}{n-K}\left((n_1 - 1)S_1 + \dots + (n_K - 1)S_K\right)$$