$\boxed{\text{Cours du } 8/11/2023}$

$$x^*(j) = \frac{x(j) - m(j)}{\sigma(j)}$$

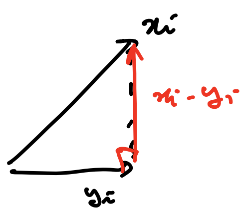
variable j

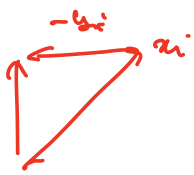
$$d^2(x^*, y^*) = \sum_{j=1}^{p} \left[x^*(j) - y^*(j)\right]^2$$

$$= \sum_{j=1}^{p} \left[\frac{x(j)-m(j)}{\sigma(j)} - \frac{y(j)-m(j)}{\sigma(j)}\right]^2$$

$$= \sum_{j=1}^{p} \frac{(x(j)-y(j))^2}{\sigma^2(j)}$$

$$= \underbrace{\left[x(1)-y(1), \dots, x(p)-y(p)\right]}_{(x-y)^T} \underbrace{\begin{pmatrix} 1/\sigma^2(1) & & \\ & \ddots & \\ & & \frac{1}{\sigma^2(p)} \end{pmatrix}}_{M} \underbrace{\begin{pmatrix} x(1)-y(1) \\ \vdots \\ x(p)-y(p) \end{pmatrix}}_{(x-y)}$$



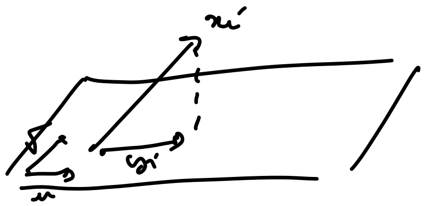


$$\langle x_i - y_i, y_i \rangle = 0 \iff \boxed{\langle x_i, y_i \rangle = \langle y_i, y_i \rangle}$$

$$\frac{1}{n}\sum_{i=1}^{n} \|x_i - y_i\|^2 = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\langle x_i - y_i, x_i - y_i \rangle}_{\underbrace{\langle x_i - y_i, x_i \rangle}_{\langle x_i, x_i \rangle - \langle y_i, x_i \rangle} - \underbrace{\langle x_i - y_i, y_i \rangle}_{0}}$$

donc $\displaystyle \frac{1}{n}\sum_{i=1}^{n} \|x_i - y_i\|^2 = \frac{1}{n}\sum \|x_i\|^2 - \frac{1}{n}\sum_{i=1}^{n} \|y_i\|^2$

donc minimiser $\frac{1}{n}\sum_{i=1}^{n} \|x_i - y_i\|^2 \iff \boxed{\text{maximiser } \frac{1}{n}\sum_{i=1}^{n} \|y_i\|^2}$



$$y_i = (x_i^T u)\,u + (x_i^T v)\,v$$

on cherche les vecteurs u et v qui maximisent $\boxed{\frac{1}{n}\sum_{i=1}^{n} \|y_i\|^2}$

on va d'abord chercher le vecteur u de norme 1 (i.e. $u^T u = 1$) qui maximise $\frac{1}{n}\sum_{i=1}^{n} (x_i^T u)^2$

$$\frac{1}{n}\sum_{i=1}^{n} (u^T x_i)(x_i^T u)$$

$$\boxed{u^T \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i^T\right)}_{\text{matrice de covariance des données}} u}$$

$x_1 = \begin{pmatrix} \\ \\ \end{pmatrix}$,

$x_2 = \begin{pmatrix} \\ \\ \end{pmatrix}$

$$\underbrace{\begin{pmatrix} x_i(1) \\ x_i(2) \\ x_i(3) \\ x_i(4) \end{pmatrix}}_{x_i \in \mathbb{R}^4} \underbrace{\begin{pmatrix} x_i(1) & x_i(2) & x_i(3) & x_i(4) \end{pmatrix}}_{x_i^T} = \begin{pmatrix} x_i^2(1) & x_i(1)x_i(2) & & \\ & x_i^2(2) & & \\ & & \ddots & \\ & & & \ddots \end{pmatrix}$$

$\boxed{\text{Le premier axe principal est le vecteur } u \text{ de norme } 1 \ (u^T u = 1) \text{ qui maximise } u^T \Sigma u \text{ où } \Sigma \text{ est la matrice de covariance des données}}$

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i x_i^T} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2(1) \right.$$

LAGRANGIEN (optimisation sous contrainte) paramètre de Lagrange

$$L(u) = \underbrace{u^T \Sigma u}_{\text{critère que l'on cherche à optimiser}} - \lambda \underbrace{(u^T u - 1)}_{\text{contrainte égalité}}$$

$u^T u = 1 \iff \boxed{u^T u - 1 = 0}$

$\boxed{\lambda > 0 \text{ est le paramètre de Lagrange}}$

La solution recherchée vérifie $L'(u) = 0$

$$L'(u) = \begin{pmatrix} \frac{\partial L}{\partial u_1} \\ \vdots \\ \frac{\partial L}{\partial u_p} \end{pmatrix} \qquad \boxed{\frac{\partial\, u^T \Sigma u}{\partial u} = 2 \Sigma u}$$

$$L'(u) = 0 \implies 2\Sigma u - \lambda 2 u = 0$$

$$\boxed{\Sigma u = \lambda u}$$

donc u est un vecteur propre de la matrice Σ (associé à la valeur propre λ)

Σ symétrique réelle $\implies$ p valeurs propres et p vecteurs propres $u_1, \dots, u_p$

on recherche à maximiser $u^T \Sigma u$

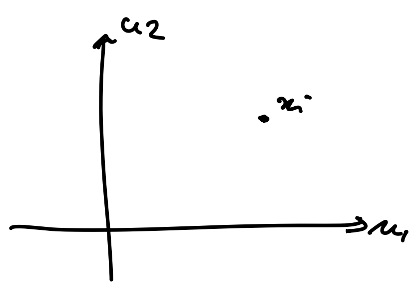
si $\Sigma u = \lambda u \implies u^T \Sigma u = u^T(\lambda u) = \lambda \underbrace{u^T u}_{1} = \lambda$

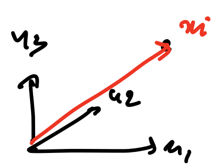
$\boxed{\text{plus } \lambda \text{ est grand, plus le critère } u^T\Sigma u \text{ est grand}}$

$\underbrace{\frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i}}_{\in]0,1[}$ s'appelle le pouvoir discriminant de l'axe i = $\boxed{\text{inertie de l'axe } i}$

<ins>individu i</ins> : $\begin{pmatrix} x_i^T u_1 \\ x_i^T u_2 \end{pmatrix}$

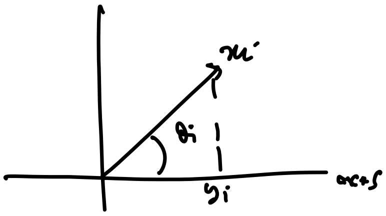
$x_i^T u_3$





<ins>recherche de v</ins>

$$L(v) = v^T \Sigma v - \lambda (v^T v - 1) - \mu (v^T u)$$

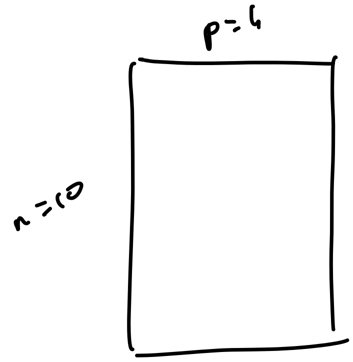


$$\boxed{\cos^2\theta_i =}$$

Cours du 9/11/2023

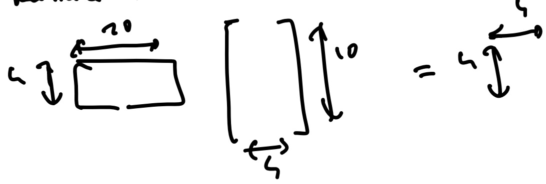
$$v'_j = \begin{pmatrix} \dfrac{x_1(j) - m(j)}{\sqrt{n}\;\sigma(j)} \\ \\ \dfrac{x_n(j) - m(j)}{\sqrt{n}\;\sigma(j)} \end{pmatrix}$$

$$\|v'_j\|^2 = \sum_{k=1}^{n} \frac{\left(x_k(j) - m(j)\right)^2}{n\,\sigma^2(j)} = \frac{1}{\sigma^2(j)} \underbrace{\frac{1}{n}\sum_{k=1}^{n}\left(x_k(j) - m(j)\right)^2}_{\sigma^2(j)} = 1$$



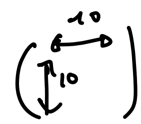
__ACP des individus.__

matrice de covariance $\frac{1}{n} X^T X$



__ACP des variables__

matrice de covariance $\frac{1}{p} X X^T$



$$\langle v'_j, v'_k \rangle = \sum_{i=1}^{n} \frac{1}{n} \frac{x_i(j) - m(j)}{\sigma(j)} \frac{x_i(k) - m(k)}{\sigma(k)}$$

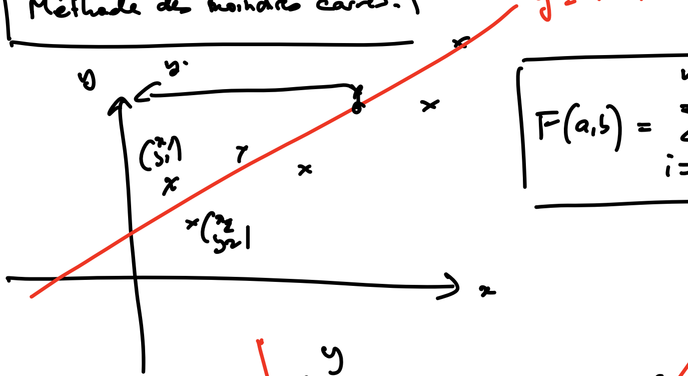
$$= \frac{\frac{1}{n}\sum_{i=1}^{n}\left[x_i(j) - m(j)\right]\left[x_i(k) - m(k)\right]}{\sigma(j)\,\sigma(k)}$$

= coefficient de corrélation entre la variable j et la variable k

$= \cos(v'_k, v'_j)$

Méthode des moindres carrés.

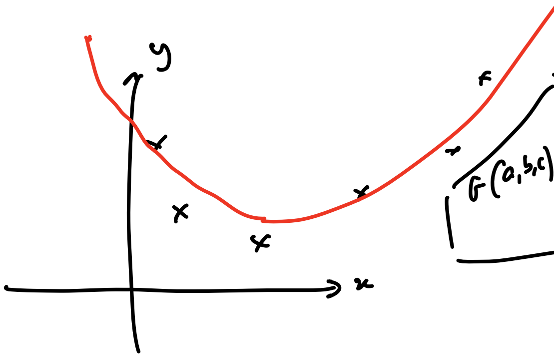
$y = ax + b$



$$\boxed{y_i \approx a x_i + b}$$

$$\boxed{F(a,b) = \sum_{i=1}^{n} (y_i - a x_i - b)^2} \qquad \beta = \begin{pmatrix} a \\ b \end{pmatrix}$$

$$y = ax^2 + bx + c = (x^2 \quad x \quad 1)\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \boxed{f(x)\,\beta}$$



$$\boxed{F(a,b,c) = \sum_{i=1}^{n} (y_i - a x_i^2 - b x_i - c)^2}$$

$$\beta = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

Exemple 1

$f(t) = a\sqrt{t} + bt$ $\qquad$ t = jour

$\beta = \begin{pmatrix} a \\ b \end{pmatrix}$ $\qquad$ $f(t)$ = nombre de serviettes vendues

1) $\left| \begin{array}{ll} t = 1 & f(t) = 2 \\ t = 4 & f(t) = 10 \end{array} \right.$

$$\begin{cases} 2 = a + b \\ 10 = 2a + 4b \end{cases} \Rightarrow \boxed{\begin{array}{l} b = 3 \\ a = -1 \end{array}}$$

Rq $\begin{pmatrix} a \\ b \end{pmatrix}$ vérifie $A\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 2 \\ 10 \end{pmatrix}$ avec $A = \begin{pmatrix} 1 & 1 \\ 2 & 4 \end{pmatrix}$

donc $\boxed{\begin{pmatrix} a \\ b \end{pmatrix} = A^{-1}\begin{pmatrix} 2 \\ 10 \end{pmatrix}}$

Erreur de modèle = erreur des moindres carrés

$$\sum_t (f(t) - a\sqrt{t} - bt)^2 = [f(1) - a - b]^2 + [f(4) - 2a - 4b]^2$$

$$= \boxed{0} \qquad \left| \begin{array}{l} f(1) = a + b \\ f(4) = 2a + 4b \end{array} \right.$$

2) $t = 9 \quad f(t) = 18$

erreur de modèle
$$\sum_t \left(f(t) - a\sqrt{t} - bt\right)^2 = \left(f(1) - a - b\right)^2 + \left(f(4) - 2a - 4b\right)^2 + \left(f(9) - 3a - 9b\right)^2$$

$$= 0 + 0 + (18 + 3 - 27)^2$$

$a = -1$
$b = 3$

$$= \boxed{36}$$

3) $\boxed{\text{Résolution du problème au sens des moindres carrés}}$

$$\boxed{\begin{aligned} 2 &\simeq a + b \\ 10 &\simeq 2a + 4b \\ 18 &\simeq 3a + 9b \end{aligned}}$$

$t = 1 \quad f(t) = 2$

$t = 4 \quad f(t) = 10$

$t = 9 \quad f(t) = 18$

$(a\sqrt{t} + bt)$

$$A \underbrace{\begin{pmatrix} a \\ b \end{pmatrix}}_{\beta} \simeq b = \begin{pmatrix} 2 \\ 10 \\ 18 \end{pmatrix}$$

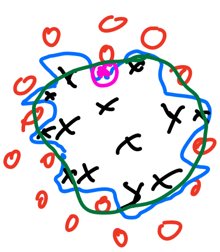
$$A = \begin{pmatrix} 1 & 1 \\ 2 & 4 \\ 3 & 9 \end{pmatrix}$$

la solution des moindres carrés est le vecteur β qui minimise $\|A\beta - b\|^2$:

$$\boxed{\hat{\beta} = (A^T A)^{-1} A^T b}$$

$\boxed{\text{Cours du 13/11/2023}}$

Surapprentissage



Classifieur Bayésien

Loi a priori

$$P(\omega_k) \begin{cases} \frac{1}{K} \quad k = 1, \dots, K & \text{cas équiprobable} \\ \frac{n_k}{n} \quad \text{avec } n = \sum_{k=1}^{K} n_k & \end{cases}$$

probabilités a priori estimées à l'aide des vecteurs de la base d'apprentissage

Exemple 1

classe ω_1 $\quad x = 0 \quad z = x + n$

classe ω_2 $\quad x = 1 \quad z = x + n$

$n \sim N(0, \sigma^2)$

quelle est la règle de décision du classifieur Bayésien ?

Probabilités a priori

$$\boxed{\begin{array}{ll} P_0 = P[x=0] & = P(\omega_1) \\ P_1 = P[x=1] = 1 - P_0 & = P(\omega_2) \end{array}}$$

lois de z conditionnellement aux classes ω_1 et ω_2

loi de $z \mid \omega_1$: $\quad x = 0 \Rightarrow z = 0 + n = n \sim N(0, \sigma^2) = f(z \mid \omega_1)$

loi de $z \mid \omega_2$: $\quad x = 1 \Rightarrow z = 1 + n \sim N(1, \sigma^2) = f(z \mid \omega_2)$

classifieur de Bayes

$d(z) = a_1 \iff P(\omega_1 \mid z) \geq P(\omega_2 \mid z)$

on affecte à la classe ω_1

$$\iff \frac{f(z \mid \omega_1) P(\omega_1)}{\cancel{f(z)}} \geq \frac{f(z \mid \omega_2) P(\omega_2)}{\cancel{f(z)}}$$

$$\iff \cancel{\frac{1}{\sqrt{2\pi\sigma^2}}} \exp\left(\frac{-z^2}{2\sigma^2}\right) P_0 \geq \cancel{\frac{1}{\sqrt{2\pi\sigma^2}}} \exp\left(-\frac{(z-1)^2}{2\sigma^2}\right) P_1$$

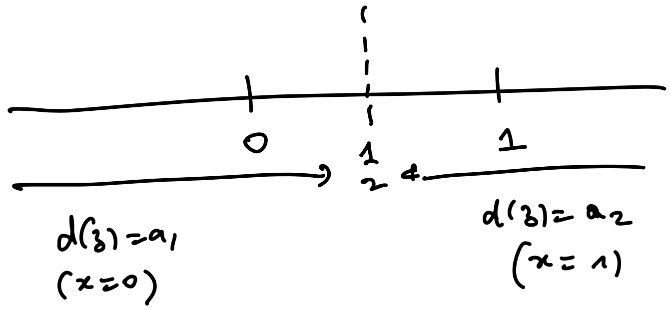
$$\iff \cancel{\frac{-z^2}{2\sigma^2}} + \ln P_0 \geq -\frac{(\cancel{z^2} - 2z + 1)}{2\sigma^2} + \ln P_1$$

on prend le ln

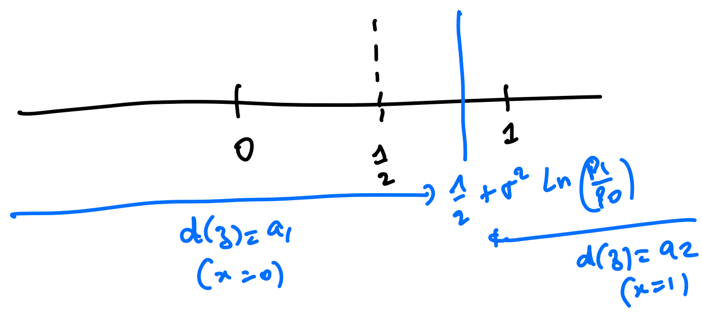
$$\iff \sigma^2(\ln P_0 - \ln P_1) \geq z - \frac{1}{2}$$

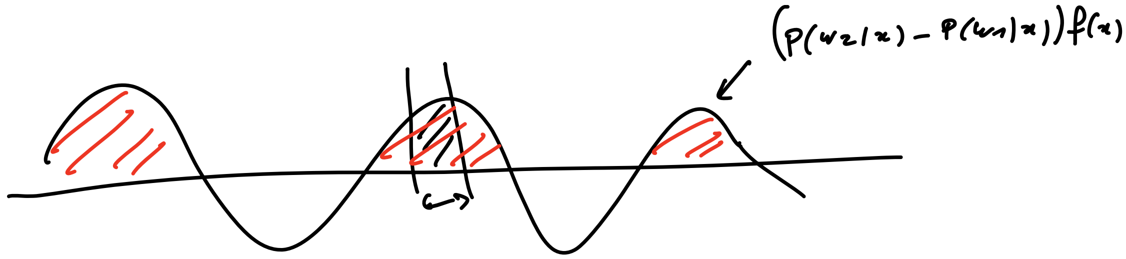
$$\iff \boxed{z \leq \frac{1}{2} + \sigma^2\left(\ln\left(\frac{P_0}{P_1}\right)\right)}$$

Si $P_0 = P_1$ (cas équiprobable) $\quad d(z) = a_1 \iff \boxed{z \leq \frac{1}{2}}$



Si $P_0 > P_1 \Rightarrow \frac{P_0}{P_1} > 1 \Rightarrow \ln\left(\frac{P_0}{P_1}\right) > 0$





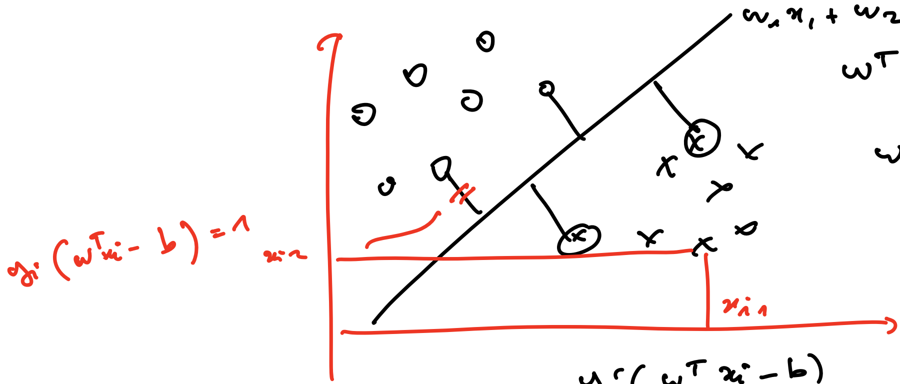
$$I = \int_{R_2} \underbrace{P[w_2|x) - P(w_1|x)]}_{\geq 0} f(x)\, dx$$

La région R_2 qui maximise I est définie par

$$\boxed{R_2 = \{ x \,/\, P(w_2|x) \geq P(w_1|x) \}}$$

Cours du 29/11/2023

SVM : on cherche à maximiser la distance minimale = marge du classifieur



$w_1 x_1 + w_2 x_2 - b = 0$

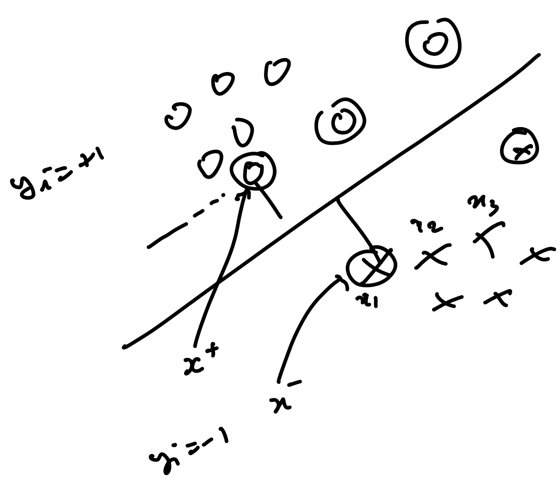
$w^T x - b = 0$

$w = \begin{pmatrix} w_1 \\ \vdots \\ w_p \end{pmatrix}$

Po. $\quad \gamma_i = \dfrac{y_i(w^T x_i - b)}{\|w\|}$ distance entre x_i et l'hyperplan d'équation $w^T x - b = 0$

Pour avoir une seule équation de la droite solution, on va se fixer une contrainte :

Contrainte globale : $\boxed{y_i(w^T x_i - b) \geq 1 \quad \forall i}$



$w^T x - b = 0$

$$\boxed{w = \sum_{i=1}^{n} \alpha_i y_i x_i}$$

b !

$$\boxed{\begin{aligned} w^T x^+ - b &= 1 \quad (1) \\ w^T x^- - b &= -1 \quad (2) \end{aligned}}$$

$(1) + (2) \Rightarrow \quad w^T(x^+ + x^-) - 2b = 0$

$$\boxed{b = \frac{1}{2} w^T (x^+ + x^-)}$$

$$L(\tilde{w},\alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{n} \alpha_i \left[(w^T x_i - b) y_i - 1 \right]$$

$$\tilde{w} = \begin{pmatrix} w \\ b \end{pmatrix} \qquad \frac{\partial L}{\partial b} = 0 \Rightarrow \boxed{\sum_{i=1}^{n} \alpha_i y_i = 0} \quad (*)$$

$$\frac{\partial L}{\partial w} = 0 \Rightarrow \boxed{w = \sum_{i=1}^{n} \alpha_i y_i x_i}$$

$$\frac{1}{2} \underbrace{\left(\sum_{i=1}^{n} \alpha_i y_i x_i^T \right)}_{w^T} \underbrace{\left(\sum_{j=1}^{n} \alpha_j y_j x_j \right)}_{w} + b \underbrace{\sum_{i=1}^{n} \alpha_i y_i}_{=0 \text{ d'après } (*)} + \sum \alpha_i$$

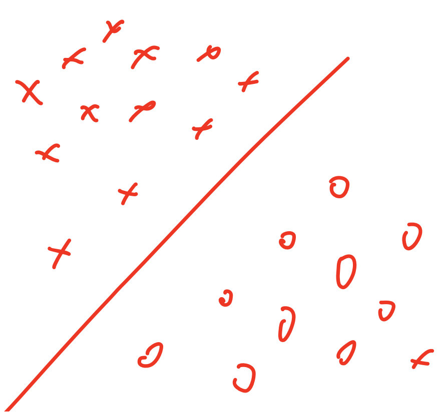
$$- \sum_{i=1}^{n} \alpha_i y_i \Big(\underbrace{\sum_{j=1}^{n} \alpha_j y_j x_j^T x_i}_{w^T x_i} \Big)$$

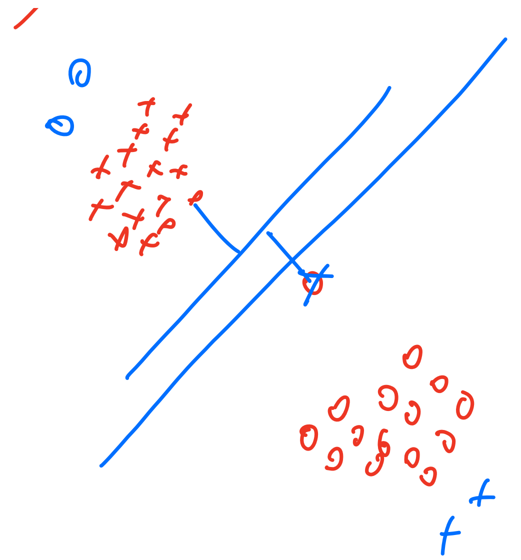
$$\left(\sum_i a_i\right)\left(\sum_j b_j\right) = \sum_{i,j} a_i b_j$$

On obtient

$$\frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_j^T x_i$$

$$\boxed{\begin{array}{l} -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_j^T x_i + \sum_{i=1}^{n} \alpha_i \\ \sum_{i=1}^{n} \alpha_i y_i = 0 \qquad \alpha_i \geq 0 \end{array}}$$

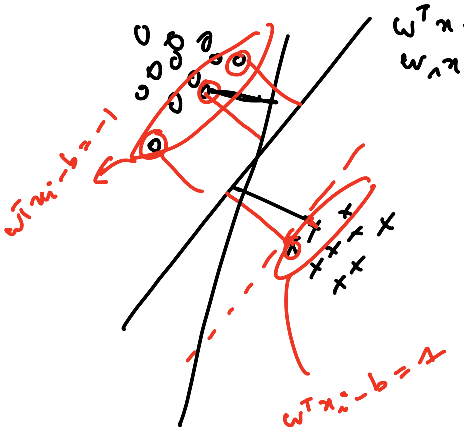


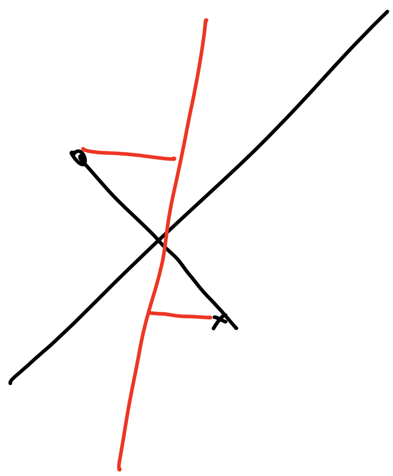


Cours du 6/12/2023

$w^T x - b = 0$

$w_1 x_1 + w_2 x_2 - b = 0$





Cours AD du 20/12/2023

	Weight	Size	Age	Class
x_1	L	S	Y	P
x_2	L	S	Y	P
x_3	L	T	Y	P
x_4	L	T	O	F
x_5	L	T	O	P
x_6	L	T	O	F

x_7	H	S	O	F
x_8	H	S	Y	F
x_9	L	S	O	P
x_{10}	H	T	O	P

1re itération

Weight gauche $\{x_1, \dots, x_6, x_9\}$ droite $\{x_7, x_8, x_{10}\}$

P P P F P F P F F P

$$G_g = 2 \times \frac{2}{7} \times \frac{5}{7} \qquad G_d = 2 \times \frac{1}{3} \times \frac{2}{3}$$

$$P_g G_g + P_d G_d = \frac{\cancel{7}}{\cancel{10}} \times \frac{2}{\cancel{7}} \times \frac{\cancel{10}}{7} + \frac{\cancel{3}}{10} \times \frac{4}{\cancel{3}} \times \frac{1}{3}$$

$$= \frac{2}{7} + \frac{2}{15} \simeq 0.42$$

Size gauche $\{x_1, x_2, x_7, x_8, x_9\}$ droite $\{x_3, x_4, x_5, x_6, x_{10}\}$

P P F F P P F P F P

$$G_g = 2 \times \frac{2}{5} \times \frac{3}{5} \qquad G_d = 2 \times \frac{3}{5} \times \frac{2}{5}$$

$$P_g G_g + P_d G_d = \frac{6}{25} + \frac{6}{25} = \frac{12}{25} \approx 0.48$$

Age gauche $\{x_1, x_2, x_3, x_8\}$ droite $\{x_4, x_5, x_6, x_7, x_9, x_{10}\}$

P P P F F P F F P P

$$G_g = 2 \times \frac{1}{4} \times \frac{3}{4} \qquad G_d = 2 \times \frac{1}{2} \times \frac{1}{2}$$

$$P_g G_g + P_d G_d = \frac{1}{\cancel{2}} \times \frac{3}{4} \times \frac{\cancel{2}}{5} + \frac{1}{2} \times \frac{3}{5}$$

$$= \frac{3}{20} + \frac{3}{10} \sim 0.45$$

On garde la variable avec l'indice de gini global le plus petit !!

ici, on choisit weight à la 1ère itération

BOOTSTRAP

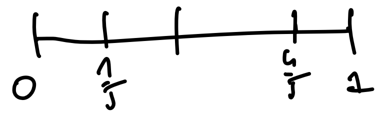
$x_1 = 1.2$ $\quad x_2 = 2.3$ $\quad x_3 = 3.4$ $\quad x_4 = 4.8$ $\quad x_5 = 6.3$

$$\mathcal{X} = \{x_1, \dots, x_5\}$$

$X \to x_1, \; x_2, \; \dots, \; x_5$

$P(X = x_1) = p_1 = \frac{1}{5}$

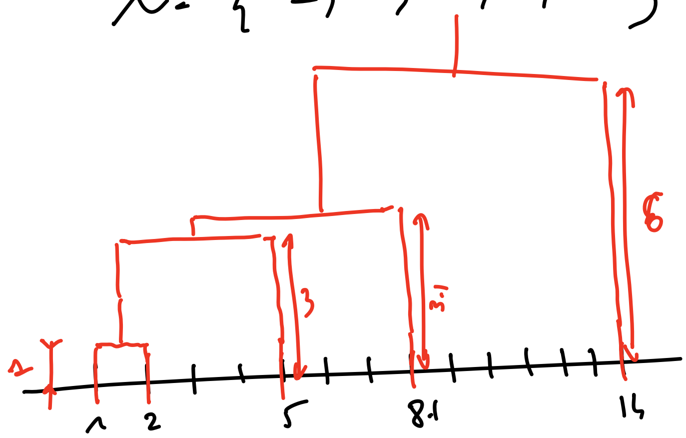
$\vdots$



$$\mathcal{X}' = \{x_2, x_2, x_2, x_4, x_5\}$$

classification hiérarchique

$$\mathcal{X} = \{1, 2, 5, 8.1, 14\}$$



dendogramm

$$d(\{1,2\}, \{5\}) = \min \{d(1,5), d(2,5)\}$$

distance entre groupes

FIN