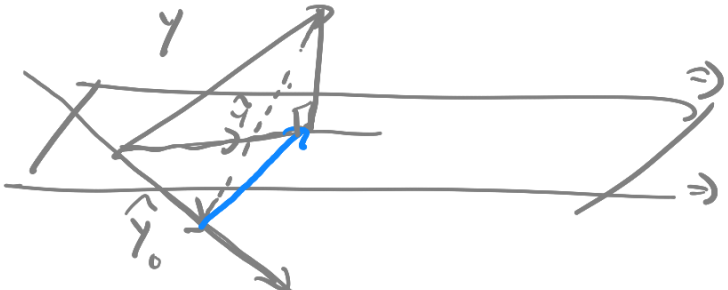
$H_0 : \mathbb{E}(Y) \in \mathcal{J}(X_0)$, $H_1 : \mathbb{E}(Y) \notin \mathcal{J}(X_0)$.

$$F = \frac{\|\hat{Y} - \hat{Y}_0\|^2 / q}{\|Y - \hat{Y}\|^2 / n-p} \quad \text{où} \quad \hat{Y} = P_X Y \quad \hat{Y}_0 = P_{X_0} Y.$$

$\hat{Y} - \hat{Y}_0 \perp Y - \hat{Y}$



$$\Rightarrow \|Y - \hat{Y}_0\|^2 = \|Y - \hat{Y}\|^2 + \|\hat{Y} - \hat{Y}_0\|^2$$

$$\Rightarrow \|\hat{Y} - \hat{Y}_0\|^2 = \|Y - \hat{Y}_0\|^2 - \|Y - \hat{Y}\|^2$$

$$= SSE_0 - SSE$$

$$F = \frac{(SSE_0 - SSE)/q}{SSE / n-p}$$

Cas particulier : si $H_0 : \beta_j = 0$ $H_1 : \beta_j \neq 0$.

$$T = \frac{\hat{\beta}_j}{\hat{\sigma}_{\hat{\beta}_j}} \qquad \text{sous } H_0 : \frac{\hat{\beta}_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim T(n-p).$$

Test : si $|T| \geq t_{n-p}(1-\alpha/2) \Rightarrow H_0$ rejected

si $|T| < t_{n-p}(1-\alpha/2) \Rightarrow H_0$ acceptée.

Goodness of fit : $H_0 : \underbrace{\mathbb{E}(Y) = \mu \in \mathbb{R}\mathbb{1}}_{(\beta_2 = \dots = \beta_p = 0)}$. $H_1 : \exists i \in [2;p]$, $\beta_i \neq 0$.

$$F = \frac{\|P_X Y - P_{\mathbb{1}} Y\|^2 / p-1}{\|Y - P_X Y\|^2 / n-p}$$

sous H_0, $F \sim F_{p-1, n-p}$.

$F > f_{p-1,n-p}(1-\alpha) \Rightarrow H_0$ rejected.

$F < f_{p-1,n-p}(1-\alpha) \Rightarrow H_0$ acceptée.

Rq :

$$P_{\mathbb{1}} Y = \frac{\langle Y, \mathbb{1} \rangle}{\langle \mathbb{1}, \mathbb{1} \rangle} \mathbb{1}$$

$$= \bar{y} \mathbb{1}.$$

$$F = \frac{\|\hat{Y} - \bar{Y}\mathbb{1}\|^2 / p-1}{\|Y - \hat{Y}\|^2 / n-p.}$$

$$\begin{cases} \|Y - \hat{Y}\|^2 + \|\hat{Y} - \bar{Y}\mathbb{1}\|^2 \\ \quad = \|Y - \bar{Y}\mathbb{1}\|^2. \end{cases}$$

$$\hookrightarrow \|Y - \hat{Y}\|^2 = \|Y - \bar{Y}\mathbb{1}\|^2 - \|\hat{Y} - \bar{Y}\mathbb{1}\|^2.$$

$$= \frac{\|\hat{Y} - \bar{Y}\mathbb{1}\|^2}{\|Y - \bar{Y}\mathbb{1}\|^2 - \|\hat{Y} - \bar{Y}\mathbb{1}\|^2} \cdot \frac{n-p}{p-1}$$

$$= \frac{R^2}{1 - R^2} \frac{n-p}{p-1}$$

(en divisant numérateur et dénominateur par $\|Y - \bar{Y}\mathbb{1}\|^2$).

Chap 5 : Sélection de modèles.

Thm 5.1 $S \subset \{1, \dots, p\}$. (ex : $S = \{2, 5, 8\}$)

Notation : $X_S = [(X_k)_{k \in S}]$ (ex : $[X_2\ X_5\ X_8] = X_S$).

$X_{S+1} = [X_S, X_k]$ où $k \notin S$.

$H_0 : \mathbb{E}(Y) \in \mathcal{J}(X_S)$, $H_1 : \mathbb{E}(Y) \in \mathcal{J}(X_{S+1})$.

$$F_1 = \frac{SSE(S) - SSE(S+1) / 1}{SSE(S+1) / n - (|S|+1)} \begin{cases} > f_{1, n-|S|-1}(1-\alpha) \Rightarrow H_1 \text{ accepté} \\ \leq \Rightarrow H_0 \text{ accepté} \end{cases}$$

Forward : F-test transition :

Si le modèle s'écrit : X_I où $I \subset [1:p]$ (ex : $p = 10$, $I = \{2, 5, 8\}$)

alors $\forall k \notin I$, on fait le test :

$$H_0 : \mathbb{E}(Y) \in \mathcal{J}(X_I) \quad , \quad H_1 : \mathbb{E}(Y) \in \mathcal{J}(X_I \cup \{X_k\}).$$

- **Cas 1** : toutes les p-valeurs ($\forall k \notin I$) sont $\geq 0{,}05$.
  alors on ne rajoute aucun X_k.
- **Cas 2** : Sinon, on choisit d'ajouter X_k correspondant.

à la plus petite p-valeur dans ts les tests.

Backward F-test transition :

le modèle s'écrit X_I où $I \subset \{1, \dots, p\}$.

on veut choisir un $k \in I$ et voir si on passe à $X_{I \setminus \{k\}}$.

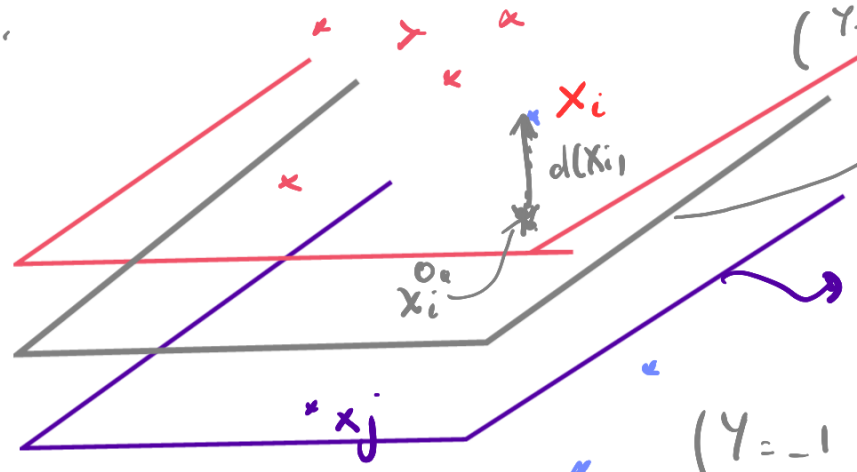
$\forall k \in I$, $H_0 : \mathbb{E}(Y) \in \mathcal{J}(X_{I \setminus \{k\}})$ $\quad$ $H_1 : \mathbb{E}(Y) \in \mathcal{J}(X_I)$.

Cas 1 : $\exists$ 1 p-valeur $\geq 0{,}05$., alors on enlève X_k.

où $k = \underset{k}{\text{Arg max}}$ p-valeur (k)

Cas 2 : sinon : on reste avec X_I.

SVM.



$(Y=1)$ $\rightarrow$ $w^T X + b = 1$. H_1.

H_0 : $w^T X + b = 0$.

$\rightarrow$ $w^T X + b = -1$. H_{-1}.

$(Y=-1)$

$\|x_i - x_i^0\| \overset{?}{=} \frac{1}{\|w\|}$.

$w^T X + b = 0 \Leftrightarrow \frac{w^T}{\|w\|} X = \frac{-b}{\|w\|}$

$$\begin{cases} x_i \in H_1 \\ x_i^0 \in H_0 \\ x_i - x_i^0 \in H_0^\perp \Rightarrow x_i - x_i^0 = \lambda \frac{w}{\|w\|} \end{cases}$$

$Mq : |\lambda| = \|x_i - x_i^0\| \overset{?}{=} \frac{1}{\|w\|}$.

$\rightarrow \begin{cases} w^T x_i + b = 1 \\ w^T x_{i0} + b = 0 \end{cases} \Rightarrow \underbrace{w^T (x_i - x_i^0) = 1}$

$w^T \lambda \frac{w}{\|w\|} = 1.$

$\Rightarrow$

$\Rightarrow |\lambda| \frac{\|w\|^2}{\|w\|} = 1 \Rightarrow |\lambda| \|w\| = 1$

$\Rightarrow |\lambda| = \frac{1}{\|w\|}$

$d(H_1, H_0) = \frac{1}{\|w\|}.$

$d(H_1, H_{-1}) = \frac{2}{\|w\|}.$

$\forall i \in \mathbb{N}$

Sous contrainte : $\begin{cases} \forall i\ w^T x_i + b \geq 1, & y_i = 1. \\ \forall i\ w^T x_i + b \leq -1, & y_i = -1. \end{cases}$

ie : $y_i (w^T x_i + b) \geq 1.$

On cherche : $\underset{w}{\text{Argmax}} \frac{2}{\|w\|}.$ sous contrainte : $\forall i \in \mathbb{N}$, $y_i (w^T x_i + b) \geq 1.$

ie : $\underset{w}{\text{Argmin}} \underbrace{\|w\|^2}$ sous contrainte : $\begin{cases} \forall i \in \mathbb{N}, \\ (*)\ y_i (w^T x_i + b) \geq 1. \end{cases}$

(sous hyp de séparabilité stricte (par un hyperplan)).

On remplace $(*)$ (trop forte) par des ξ_i (slack variable) tq : $\begin{cases} \xi_i + y_i (w^T x_i + b) \geq 1. \\ \xi_i \geq 0. \end{cases}$ ($y_i (w^T x_i + b) \leq 0$)

De plus, en posant $\varphi(x) = \text{Sgn}(w^T x + b)$.

$\mathbb{1}_{y_i \neq \varphi(x_i)} = \mathbb{1}_{y_i (w^T x + b) \leq 0} \leq \mathbb{1}_{\xi_i \geq 1.} \leq \xi_i$

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{y_i \neq \varphi(x_i)} \leq \frac{1}{n}\sum_{i=1}^{n} \xi_i$$

Donc on va chercher à minimiser : $\frac{1}{n}\sum_{i=1}^{n} \xi_i + \lambda \|w\|^2$.

$\forall (\xi_i, w, b)$

sous contrainte : $\xi_i \geq 0$, $\xi_i + y_i(w^T x_i + b) \geq 1$.

$\xi_i \geq \max(0, 1 - y_i(w^T x_i + b))$.

$\xi_i \geq (1 - y_i(w^T x_i + b))_+$.

$$\underset{w,b}{\text{Argmin}} \; \frac{1}{n}\sum_{i=1}^{n} \underbrace{(1 - y_i(w^T x_i + b))_+}_{l(y_i, w^T x_i + b)} + \underbrace{\lambda_n \|w\|^2}_{\text{pénalité ridge}}.$$

$f_{\hat{w},\hat{b}}(x_i) = \hat{w}^T x_i + \hat{b}$.

où $l(y, a) = (1 - ya)^+$.

↙ Hinge loss.

$\varphi(x) = \text{sgn}\left(f_{\hat{w},\hat{b}}(x)\right)$.

$$\left\{ \underset{f \in F = \{f : \text{mesurables}\}}{\text{Argmin}} \mathbb{E}\left[(1 - Y f(X))_+\right] \right. \quad \text{et} \quad \varphi(x) = \text{sgn}(f(x)).$$

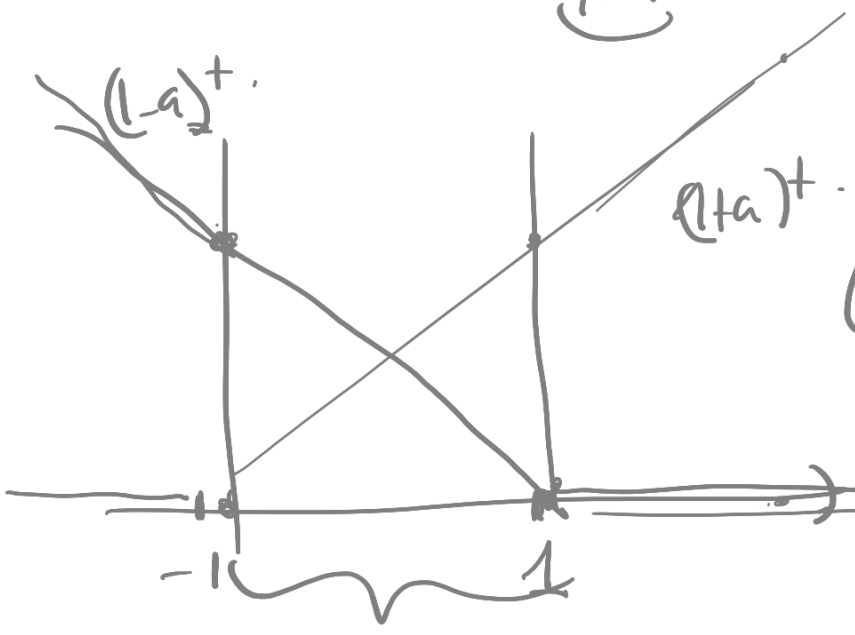
$$\mathbb{E}\left[(1 - Y f(X))_+\right] = \mathbb{E}\left[\mathbb{E}\left[(1 - Y f(X))^+ \mid X\right]\right].$$

$$\geq \mathbb{E}\left[\inf_{a \in \mathbb{R}} \mathbb{E}\left((1-Ya)^+ \mid X\right) \right]$$

$$\mathbb{E}\left((1-Ya)^+ \mid X\right)$$

$$= (1-a)^+ \mathbb{P}(Y=1 \mid X) + (1+a)^+ (1-\eta(X))$$

$\eta(X)$



$$\min_{a \in [-1,1]} (1-a)\eta + (1+a)(1-\eta)$$

$$(1-a)\,\|$$

$$\min_{a \in [-1,1]} a(1-2\eta) + \ldots$$

$\eta < 1/2 \leftarrow 1-2\eta > 0$ $\quad \operatorname{Argmin}_{a \in [-1,1]} a(1-2\eta) = -1$

$\eta > 1/2 \leftarrow 1-2\eta < 0$ $\quad \operatorname{Argmin}_{a \in [-1,1]} a(1-2\eta) = 1$

on en tire :

optimal Bayes classifier $\in \operatorname{argmin}_a \mathbb{E}\left((1-Ya)^+ \mid X\right)$