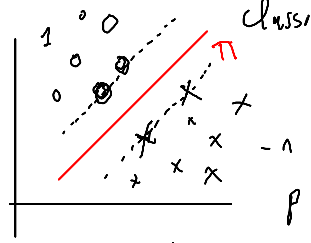
Séparateurs à vaste marge $\underline{SVM}$ (V2AVA)

classification.



p points

(x_i, l_i) $\quad l_i : \begin{matrix} 1 \\ -1 \end{matrix}) \in 2$ classes

$\mathbb{R}^n$ ↳ pts d'apprentissage.

• Séparation Linéaire.

$\Pi : w^T x + b = 0$

$d(x_i, \Pi) = \left| \dfrac{w^T x_i + b}{\|w\|} \right| \swarrow l_i$

$x_i /$

$\underbrace{\dfrac{l_i (w^T x_i + b)}{\|w\|}}_{d(x_i, \Pi)} \geq \dfrac{1}{\|w\|}$ en dehors d'une marge de $\frac{1}{2}$ largeur $\dfrac{1}{\|w\|}$

Max.

$\text{Min } \frac{1}{2} \|w\|^2$ $\quad$ p contraintes : $1 - l_i (w^T x_i + b) \leq 0$

$\forall i \; x_i$ en dehors Marge

$L(w, b, \alpha)$ $\quad w \in \mathbb{R}^n$, $b \in \mathbb{R}$, $\alpha \in \mathbb{R}^p$

$= \underbrace{\frac{1}{2} \|w\|^2} + \sum_{i=1}^{p} \alpha_i (1 - l_i (w^T x_i + b))$

Soit $\alpha_i = 0$ contrainte inactive

Soit $\alpha_i \neq 0$

$l_i (w^T x_i + b) = \dfrac{1}{l_i}$

$\left(b = \dfrac{1}{l_i} - w^T x_i \right)$

$H(\alpha)$ fonction Duale : $\underset{x, b}{\text{Min}} \; L(w, b, \alpha)$

$$\begin{cases} \nabla_w L(w, b, \alpha) = w - \sum_{i=1}^{p} \alpha_i l_i x_i \\ \qquad w(\alpha) = \sum_{i=1,p} \alpha_i l_i x_i \\ \nabla_b L(\,) = -\sum \underline{\alpha_i l_i = 0} \end{cases}$$

$H(\alpha) = -\frac{1}{2} \sum_{i,j} \alpha_i \alpha_j l_i l_j \underbrace{x_i^T x_j} + \underbrace{\sum_{i=1}^{p} \alpha_i} - \cancel{\sum \alpha_i l_i b}$

$\underline{\text{Pb dual}}$: $\text{Max } H(\alpha) = -\frac{1}{2} \alpha^T \underset{p \times p}{A} \alpha + \alpha^T u$

$\underline{\text{avec}}$.

$.\; \alpha \geq 0$

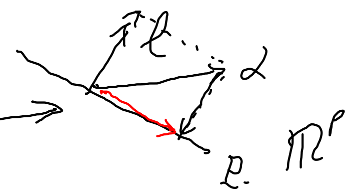
$\underbrace{\alpha^T l} = \sum \alpha_i l_i = 0$

$A_{ij} = l_i l_j x_i^T x_j$

$u = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \mathbb{R}^p$

→ Gradient projeté : ① Projette sur l

② Projette sur $\alpha \geq 0$



Que cherche-t-on ? $\quad w^T x + b = D(x)$

→ on n'a pas besoin de w explicitement.

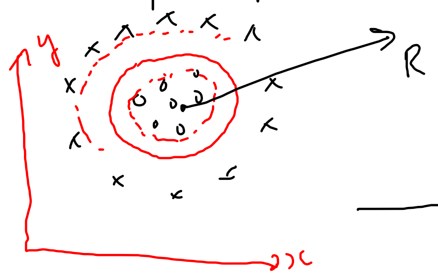
$D(x) = \underbrace{\sum_{i=1}^{p} \alpha_i^* l_i \langle x_i, x \rangle}_{x_i^T . x} + \underbrace{b}_{?}$

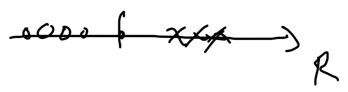
Cherche pts supports : $\alpha_i \geq \varepsilon$ $\quad \varepsilon = 10^{-5}$

$\rightarrow b_i \rightarrow$ Moyenne.

Kernel Trick $\longrightarrow$

2 classes pas séparables.





$\longrightarrow$ Plonger $\{x_i\}$ dans un espace + grand
$\mathbb{R}^n$

Espace de Redescription.

$\to$ Séparer par un Hyperplan.

Idée : $\mathbb{R}^n \longrightarrow$ dim ∞ / Hilbert

Construit à partir des | ol. (1 à p

1. Noyeau $K(x, y)$

$\mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$

- Symétrique
- positif

$K(x,y) = K(y,x)$

$K(x,x) \geqslant 0$

$K(x,y) = x^T y$

$\langle w, x_i \rangle$

$\Downarrow$

prod scalaire

$\perp$, projection $\to$ Géométrie..

$K(x,y) = \exp -\left(\frac{\|x-y\|^2}{\sigma^2}\right)$

$K(x,x) = 1$

$\to$ Noyau proposé dans le TP

$$\underset{\in \mathbb{R}^n}{x} \longrightarrow \left(y \to K(x,y)\right)$$

$\mathbb{R}^n \to \mathbb{R}$

$K(x, \cdot)$

$\to f \in V$

$\tilde{x} \in \mathbb{R}^n \longrightarrow \hat{x}$ fonction.

$V = \mathrm{Vect}\{\hat{x}\}$: Comb. linéaires Finies

$f = \sum_{k=1}^{q} f_i \hat{x}_i$

$g = \sum g_j \hat{x}_j$

f et g : $\langle f, g \rangle = \sum f_i g_j \langle \hat{x}_i, \hat{x}_j \rangle$

V n'est pas complet $\to$ f_n cvs $f \notin V$?

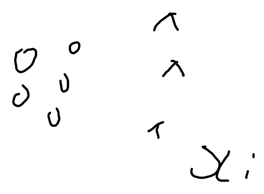
$V \ni g_n$ cvs $g \notin V$

$H = \{$ lim de suites de Cauchy de V $\}$

$\underbrace{\langle f, g \rangle}_{H}$ lim $\underbrace{\langle f_n, g_n \rangle}_{V}$

H : hilbert. $\langle f, g \rangle$

RKHS $\leftarrow$



$(x_i, l_i) \longrightarrow (\hat{x}_i, l_i)$

$\hookrightarrow y \to K(x_i, y)$

$(x_i, l_i) \longrightarrow (\hat{x}_i, l_i)$

$\hookrightarrow y \to K(x_i, y)$

Chercher Hyperplan ... dans g.

$\underbrace{\langle w, g \rangle}_{H} + b = 0$

$$d(g, \Pi) = \left| \frac{\langle w, g \rangle + b}{\|w\|_H} \right|$$

→ exactement pareil :

Contraintes : $1 - \underbrace{\langle w, x_i \rangle}_{H} b \leq 0$

Optim

→ H : min fonc. $\alpha_i \in \mathbb{R}$.

$H(\alpha) = \alpha^T A \alpha + u \alpha$

$A_{ij} = l_i l_j \underbrace{\langle \hat{x}_i, \hat{x}_j \rangle}_{K(x_i, x_j)}$

H

$x_i \to \hat{x}_i$

b. linéaire

$\times$

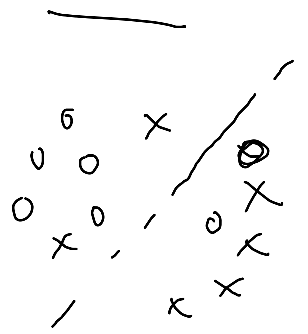
$D(x) = \langle w, \hat{x} \rangle + b$

$w = \sum l_i \alpha_i \hat{x}_i$ $D(x) = \sum l_i \alpha_i \underbrace{\langle \hat{x}_i, \hat{x} \rangle}_{K(x_i, x)} + b$

→ Marges souples

↓

Avoir non H.



① SVM Linéaire → fin de cette séance.

② SVM Kernel + Marges souples.

TP énoncé 2020

Matériel élève → Matlab → Python.