Notes du cours (2)

Slide 7 :

$\forall i$ tq $y_i = 1$ (bleu), on a $\langle w, x_i \rangle + b \geq 1$

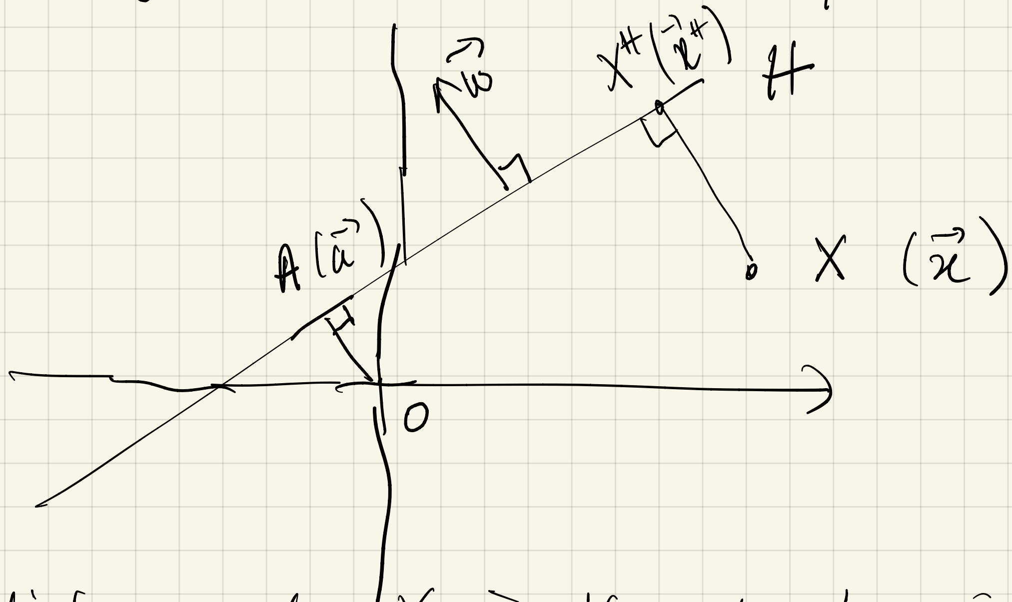
$\forall i$ tq $y_i = -1$ (rouge), on a $\langle w, x_i \rangle + b \leq -1$

donc pour les bleus $\underset{\overset{\shortparallel}{1}}{y_i} \underbrace{(\langle w, x_i \rangle + b)}_{\geq 1} \geq 1$

rouges $\underset{\overset{\shortparallel}{-1}}{y_i} \underbrace{(\langle w, x_i \rangle + b)}_{\leq -1} \geq 1$

$\Rightarrow$ 1 point est correctement classifié $\Leftrightarrow y_i(\langle w, x_i \rangle + b) \geq 1$.

Slide 8 : distance d'un point x à H.



La distance de X à H est donnée par $|XX^H|$

$OX = OA + AX^H + X^H X$ et A et X^H sont dans H

$\Leftrightarrow \vec{x} = \vec{a} + AX^H + X^H X$ donc $AX^H \perp \vec{w}$

$\Rightarrow \langle \vec{x}, \vec{w} \rangle = \langle \vec{a}, \vec{w} \rangle + \langle AX^H, \vec{w} \rangle + \langle X^H X, \vec{w} \rangle$

$= -b + \overset{\shortparallel}{0} + \alpha \langle \vec{w}, \vec{w} \rangle$

$\Rightarrow \langle \vec{x}, \vec{w} \rangle = -b + \alpha \|w\|^2$

$\Rightarrow \alpha \|w\| = \dfrac{\langle \vec{x}, \vec{w} \rangle + b}{\|w\|}$

$\vec{a} \cdot \| \vec{w}$.

$A \in H \Leftrightarrow \langle \vec{a}, \vec{w} \rangle + b = 0$

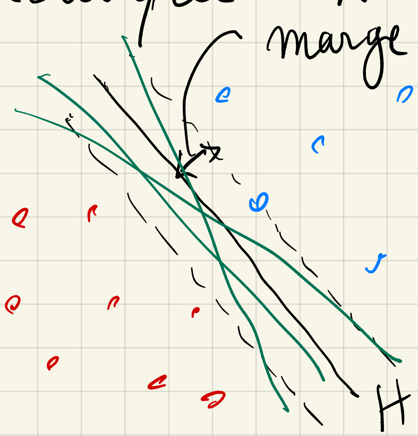
$XX^H \parallel \vec{w}$

Comme $XX^H \parallel \vec{w}$ on a noté $XX^H = \lambda \vec{w}$

et on a dit que

$$d(\vec{x}, H) = |XX^H| = |\lambda \vec{w}| = |\lambda| \, \|\vec{w}\| = \frac{|\langle x, \vec{w}\rangle + b|}{\|\vec{w}\|} \Big)$$

Marge pour le jeu de données et l'hyperplan canonique H.



$$\text{marge} : \min_{1\le i\le n} d(x_i, H)$$

$$= \min_{1\le i\le n} \frac{|\langle x_i, w\rangle + b|}{\|w\|}$$

$$= \frac{1}{\|w\|} \underbrace{\min_{1\le i\le n} |\langle x_i, w\rangle + b|}_{= 1 \text{ pour l'hyperplan } \underline{\underline{\text{canonique}}}}$$

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

Le meilleur hyperplan canonique est celui qui maximise la marge et, comme on est le cas linéairement séparable, on impose que tous les points soient correctement classifiés.

$$\max \frac{1}{\|w\|} \quad \text{s.c.} \quad \forall\, 1\le i\le n \quad y_i\left(\langle x_i, w\rangle + b\right) \geq 1$$

$$\iff \min \|w\| \quad \text{s.c.} \quad \forall\, 1\le i\le n \quad y_i\left(\langle x_i, w\rangle + b\right) \geq 1.$$

⚠ w et b sont choisis de sorte que l'hyperplan est canonique.

Slide 13 on veut résoudre

$$\left\| \min_{w\in\mathbb{R}^d, b\in\mathbb{R}, s\in\mathbb{R}^n} \frac{1}{2}\|w\|_2^2 + C\sum s_i \right.$$

$$\text{s.c. } y_i\left(\langle x_i, w\rangle + b\right) \geq 1 - s_i \text{ et } s_i \geq 0 \quad \forall i = 1, \dots, n.$$

Rappel. $\min_{x\in\mathbb{R}^p} f(x)$ s.c. $g_i(x) \leq 0$ $i = 1, \dots, n$.

$L(x, \alpha) = f(x) + \sum_{i=1}^n \alpha_i g_i(x)$ $\alpha = (\alpha_1, \dots, \alpha_n)$ sont les variables duales.

$D(\alpha) = \inf_x L(x, \alpha)$

$D^* = D(\alpha^*) = \max_{\alpha\geq 0} D(\alpha)$ sous dualité forte $D^* = P^* = f(x^*)$

$f(x^*) = \min f(x)$ s.c. $g_i(x) \leq 0$ $i = 1, \dots, n$.

On a la dualité forte si f, g_i sont convexes et $\exists x$ $g_i(x) < 0$ $i = 1, \dots, n$.

OK dans le cas des SVM.

$$L(w, b, s, \alpha, \beta) = \frac{1}{2}\|w\|_2^2 + C\sum s_i + \sum_{i=1}^n \alpha_i \left(1 - s_i - y_i(\langle x_i, w\rangle + b)\right) - \sum_{i=1}^n \beta_i s_i$$

Conditions KKT : à l'optimum, on a :

$$\nabla_w L(w^*, b^*, s^*, \alpha^*, \beta^*) = 0 = w - \sum_{i=1}^n \alpha_i y_i x_i$$

$$\nabla_b L(w^*, b^*, s^*, \alpha^*, \beta^*) = 0 = -\sum_{i=1}^n \alpha_i y_i$$

$$\nabla_s L(w^*, b^*, s^*, \alpha^*, \beta^*) = 0 = \begin{pmatrix} C \\ \vdots \\ C \end{pmatrix} - \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} - \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}$$

on enlève les * sur les valeurs à l'optimum

Conditions supplémentaires.

$$\alpha_i^* \left(1 - s_i^* - y_i(\langle w^*, x_i\rangle + b^*)\right) = 0 \Leftrightarrow \alpha_i = 0 \text{ ou } \underbrace{1 - s_i - y_i(\langle x_i, w\rangle + b) = 0}_{A}$$

$$\beta_i^* s_i^* = 0 \Leftrightarrow \beta_i = 0 \text{ ou } s_i = 0.$$

On a donc à l'optimum.

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

Seuls les i tq $\alpha_i \neq 0$ apparaissent dans w. On les appellent les vecteurs support

→ SVM : support vectors machine

pour ces vecteurs support, on a :

$y_i(\langle x_i, w\rangle + b) = 1 - s_i$ $\left\langle \begin{array}{l} s_i = 0 \text{ donc } x_i \in \dots \\ s_i > 0 \quad x_i \text{ est un outlier} \leftarrow \alpha_i = C. \end{array} \right.$

Pour un vecteur support, i.e. tq $\underline{\alpha_i \neq 0}$, et tel que $\delta_i = 0$ (c'est un point qui est sur l'un des hyperplans ...)

on a $1 - \delta_i - y_i(\langle x_i, w\rangle + b) = 0$

et $\delta_i = 0$

$\Leftrightarrow 1 - y_i\langle x_i, w\rangle - y_i b = 0$

$\Leftrightarrow y_i b = 1 - y_i\langle x_i, w\rangle$.

$\underset{-1 \text{ ou } 1}{\overset{\shortparallel}{}} \Rightarrow y_i^2 = 1$

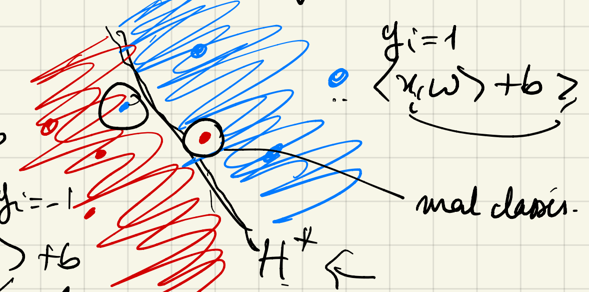
$\Leftrightarrow y_i^2 b = y_i - y_i^2\langle x_i, w\rangle$.

$\Leftrightarrow b = y_i - \langle x_i, w\rangle$ à l'optimum $w = \sum_{i=1}^n \alpha_i y_i x_i$

$\| \; b = y_i - \langle x_i, \sum_{j=1}^n \alpha_j y_j x_j\rangle = y_i - \sum_{j=1}^n \alpha_j y_j \langle x_i, x_j\rangle$

Notre classifieur SVM.

$x \to \text{signe}(\langle x, w\rangle + b) \Leftrightarrow$



Slide 20.

Hinge loss $\Leftrightarrow$ SVM.

$\min \frac{1}{2}\|w\|_2^2 + C\sum \delta_i$ sc $y_i(\langle x_i, w\rangle + b) \geq 1 - \delta_i$ $\|$.

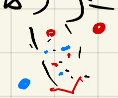
$\delta_i \geq 0$

Si i correctement classifié, $y_i(\langle w, x_i\rangle + b) \geq 1$.
$(\delta_i = 0)$

$\Leftrightarrow 1 - y_i(\langle w, x_i\rangle + b) \leq 0$ et $\delta_i = 0$

dans ce cas $\max(0, 1 - y_i(\langle x_i, w\rangle + b)) = 0 = \delta_i$

Si i est un outlier $y_i(\langle w, x_i\rangle + b) < 1$



$\Leftrightarrow 1 - y_i(\langle w, x_i\rangle + b) > 0$

donc $\max(0, 1 - y_i(\langle w, x_i\rangle + b))$

$= 1 - y_i(\langle w, x_i\rangle + b) = \delta_i$

on a aussi $1 - \delta_i - y_i(\langle w, x_i\rangle + b) = 0$.

$\Leftrightarrow \delta_i = 1 - y_i(\langle w, x_i\rangle + b)$

Notre problème est équivalent :

$$\min_{w,b} \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^n \max(0, y_i(\langle x_i, w\rangle + b)).$$

on n'a plus besoin de contraintes.

$\frac{1}{n}\sum_{i=1}^n \ell(x_i, y_i) + \text{pen}$

$$= \min_{w,b} \frac{1}{n}\sum_{i=1}^n \max(0, y_i(\langle x_i, w\rangle + b)) \text{ hinge loss } + \frac{1}{2Cn}\|w\|_2^2$$

Slide 22 Problème dual : on remplace w^* et b^* par leurs valeurs en fonction des variables duales

$$L(w, b, s, \alpha, \beta) = \frac{1}{2}\|w\|_2^2 + C\sum s_i + \sum \alpha_i (1 - s_i - y_i(\langle x_i, w\rangle + b)) - \sum \beta_i s_i.$$

à l'optimum on a $w^* = \sum_i \alpha_i^* y_i x_i$

$$\alpha_i^* (1 - s_i^* - y_i(\langle x_i, w^*\rangle + b^*)) = 0 \quad (*)$$

$$\beta_i^* s_i^* = 0 \qquad \sum \alpha_i^* y_i = 0 \qquad \alpha_i^* + \beta_i^* = C.$$

(pour simplifier, on enlève les $*$).

$$(**) = \frac{1}{2}\left\|\sum_{i=1}^n \alpha_i y_i x_i\right\|_2^2 + \sum_{i=1}^n (\alpha_i + \beta_i) s_i + \sum_{i=1}^n 0 - \sum_{i=1}^n 0$$

$$= \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle x_i, x_j\rangle + \sum_{i=1}^n \alpha_i s_i + \underbrace{\sum_{i=1}^n \beta_i s_i}$$

Il reste à exprimer s_i en fonction des variables duales.

$$(*) \Leftrightarrow \alpha_i = \alpha_i s_i + \alpha_i y_i (\langle x_i, w\rangle + b).$$

$$\Leftrightarrow \alpha_i s_i = \alpha_i (1 - y_i(\langle x_i, w\rangle + b))$$

On calcule :

$$\sum \alpha_i s_i = \sum \alpha_i - \sum \alpha_i y_i \langle x_i, w\rangle - \sum \alpha_i y_i b.$$

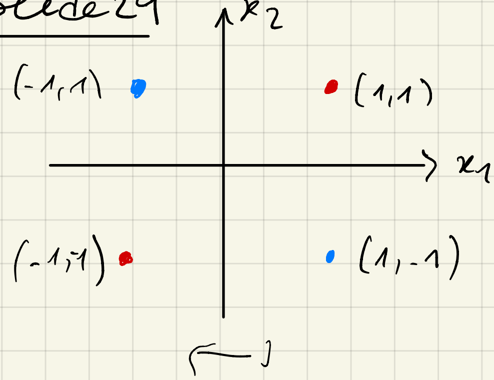
$$= \sum \alpha_i - \sum_{i=1}^n \alpha_i y_i \left\langle x_i, \sum_{j=1}^n \alpha_j y_j x_j \right\rangle - b \underbrace{\sum \alpha_i y_i}_{=0}$$

$$= \sum \alpha_i - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle x_i, x_j\rangle.$$

$$(**) = \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle x_i, x_j\rangle + \sum \alpha_i - \sum\sum \alpha_i \alpha_j y_i y_j \langle x_i, x_j\rangle$$

$$= \sum_{i=1}^n \alpha_i - \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle x_i, x_j\rangle = D(\alpha).$$

Slide 29



problème de classification XOR
exclusive or.

$\iff$ un point est bleu ssi exactement une de ses coordonnées $= +1$

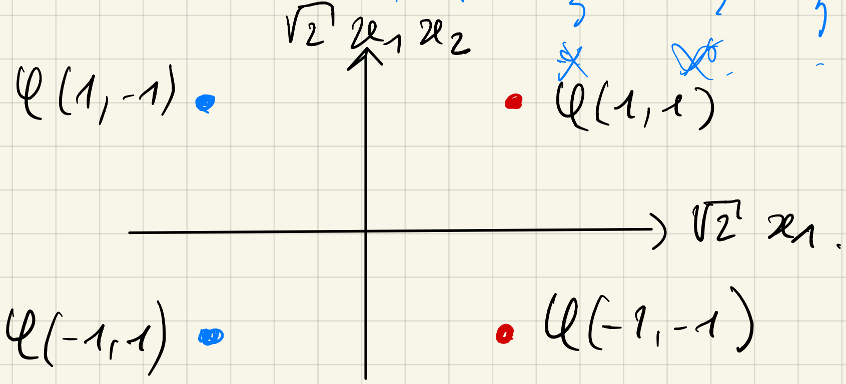
dans $\mathbb{R}^2$ on ne peut trouver d'hyperplan (droite) qui sépare les points rouges des bleus.

On essaie de trouver une transformation φ telle que dans le nouvel espace les points soient séparables par un hyperplan.

$$\varphi(\underset{(x_1,x_2)}{x}) = (x_1^2, x_2^2, \sqrt{2}\,x_1x_2, \sqrt{2}\,x_1, \sqrt{2}\,x_2, 1)$$

$\mathbb{R}^2 \xrightarrow{\varphi} \mathbb{R}^6$. on augmente la dimension.

$$\varphi(1,1) = (1, 1, \sqrt{2}, \sqrt{2}, \sqrt{2}, 1)$$
$$\varphi(-1,-1) = (1, 1, \sqrt{2}, -\sqrt{2}, -\sqrt{2}, 1)$$
$$\varphi(1,-1) = (1, 1, -\sqrt{2}, \sqrt{2}, -\sqrt{2}, 1)$$
$$\varphi(-1,1) = (1, 1, -\sqrt{2}, -\sqrt{2}, \sqrt{2}, 1)$$



c'est linéairement séparable
$\to$ on peut faire passer un hyperplan qui sépare les points rouges des pts bleus.

Slide 30

$x, x' \in \mathbb{R}^2$

$$\langle \varphi(x), \varphi(x') \rangle = \left\langle \begin{pmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}\,x_1x_2 \\ \sqrt{2}\,x_1 \\ \sqrt{2}\,x_2 \end{pmatrix}, \begin{pmatrix} x_1'^2 \\ x_2'^2 \\ \sqrt{2}\,x_1'x_2' \\ \sqrt{2}\,x_1' \\ \sqrt{2}\,x_2' \end{pmatrix} \right\rangle$$

$\langle \ , \ \rangle_{F = \mathbb{R}^6}$

$$= (x_1x_1')^2 + (x_2x_2')^2 + 2\,x_1x_2\,x_1'x_2' + 2x_1x_1' + 2x_2x_2' + 1$$
$$= (x_1x_1' + x_2x_2' + 1)^2 = (\langle x, x' \rangle + 1)^2$$ à nouveau cela dépend de $\langle x, x' \rangle$.

Slide 35.

K et K' 2 noyaux PDS. $K+K'$ est PDS ?

$$K+K'(x,x') = K(x,x') + K'(x,x') = K(x',x) + K'(x',x) = (K+K')(x',x).$$

$$u^T(\mathbb{K}+\mathbb{K}')u = u^T\mathbb{K}u + u^T\mathbb{K}'u \geq 0.$$

$\Rightarrow$ $K+K'$ est PDS

$K \cdot K'$ est PDS ?

$$u^T \mathbb{K}\odot\mathbb{K}' u = \sum_{1\leq i,j\leq N} u_i u_j \mathbb{K}_{ij}\mathbb{K}'_{ij} \quad (*)$$

$$\mathbb{K}\odot\mathbb{K}' = (\mathbb{K}_{ij}\cdot\mathbb{K}'_{ij})_{ij}$$

on sait $\mathbb{K}$ est une matrice définie positive.
il existe M tq $\mathbb{K} = MM^T$

$$\mathbb{K}_{ij} = \sum_{k=1}^{N} M_{ik}M_{kj}$$

$$(*) = \sum_{1\leq i,j\leq N} u_i u_j \sum_{k=1}^{N} M_{ik}M_{kj}\mathbb{K}'_{ij} = \sum_{k=1}^{N} \underbrace{\sum_{i,j} u_i M_{ik}\mathbb{K}'_{ij} u_j M_{kj}}$$

$$z_k = u\odot M_{\cdot k} = \begin{pmatrix} u_i M_{ik} \end{pmatrix}$$

$$= \sum_{k=1}^{N} \underbrace{z_k^T \mathbb{K}' z_k} \geq 0.$$

$K\cdot K'$ est bien PD.

Slides 36 et suivantes

Noyau linéaire. $k(x,x') = \langle x,x'\rangle + c \quad c\geq 0.$

$$\mathbb{K} = (\langle x_i, x_j\rangle + c)_{1\leq i,j\leq n}$$

K est symétrique.

$$u^T\mathbb{K}u = \sum_{1\leq i,j\leq n} u_i u_j K(x_i,x_j) = \sum_{1\leq i,j\leq n} u_i u_j\langle x_i,x_j\rangle + c\sum_{1\leq i,j\leq n} u_i u_j = \left\|\sum_{i=1}^{n} u_i x_i\right\|_2^2 + c\left(\sum_{i=1}^{n} u_i\right)^2 \geq 0. \quad \text{PDS.}$$

$\Rightarrow$ Noyau polynomial $K(x,x') = (\langle x,x'\rangle + c)^q \quad q\in\mathbb{N}^*$

PDS comme puissance du noyau linéaire.