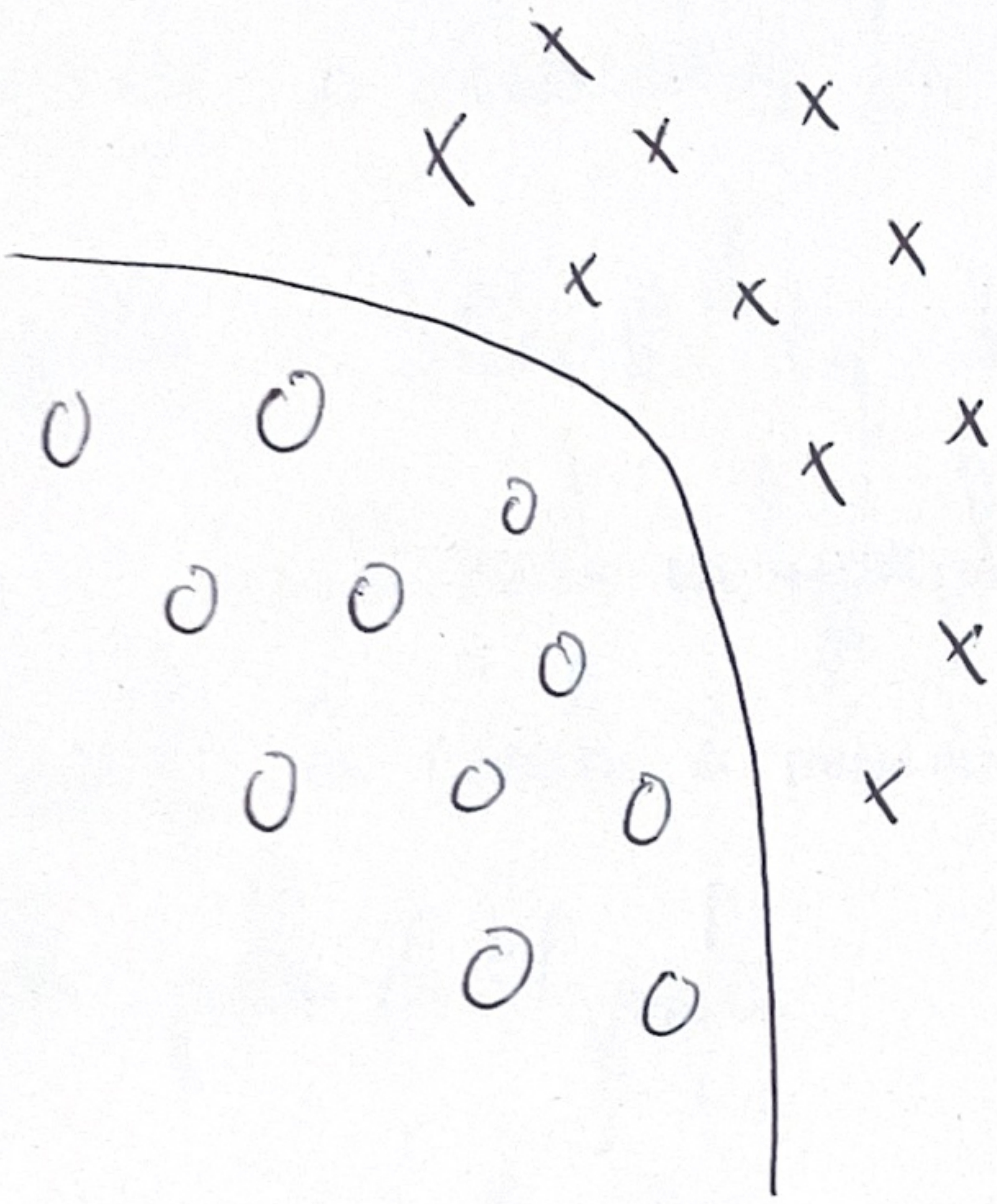
①

Leçon n° 6 : Méthodes à noyaux II

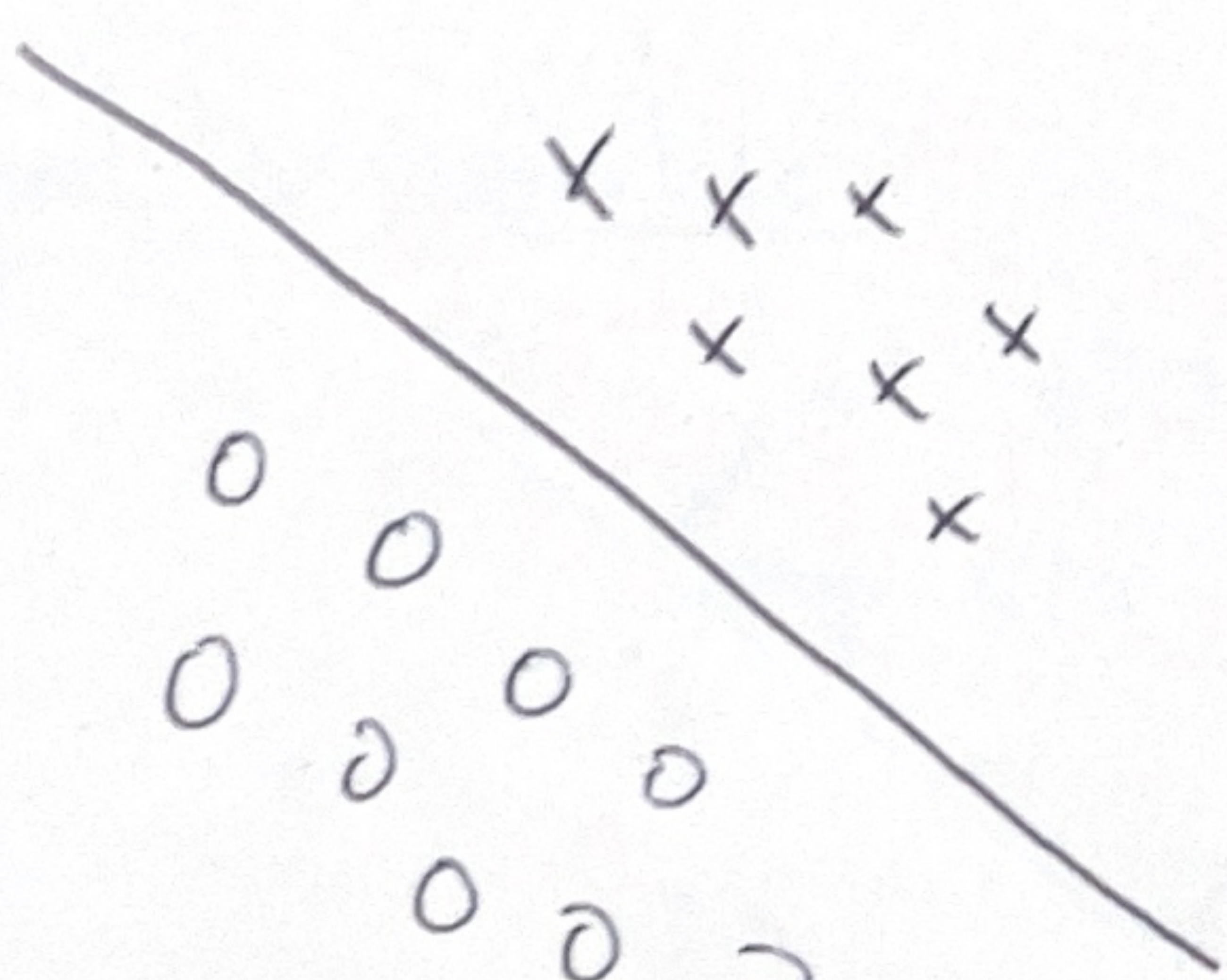


Objectif : Séparer deux classes

Problème : $\hat{\theta}_n \in \underset{\theta}{\arg\min} \frac{1}{n}\sum_{i=1}^{n} \max\left(1 - y_i\langle\theta, \varphi(x_i)\rangle, 0\right) + \frac{\lambda}{2}\|\theta\|^2$

↳ Minimum du risque empirique pour la fonction de perte Hinge + Régularisation.

I. Le cas linéaire séparable



(2)

~~Hypothèse~~

<u>Contexte</u> n observations $(x_i, y_i) \in \mathbb{R}^d \times \{-1, 1\}$

<u>Hypothèse</u> : $\exists \Theta \in \mathbb{R}^d$ tq $\forall i, \; y_i(\theta^T x_i) > 0$

↳ Il existe un hyperplan séparateur.

<u>Objectif</u> : On cherche à maximiser la distance du point le plus proche de l'hyperplan.

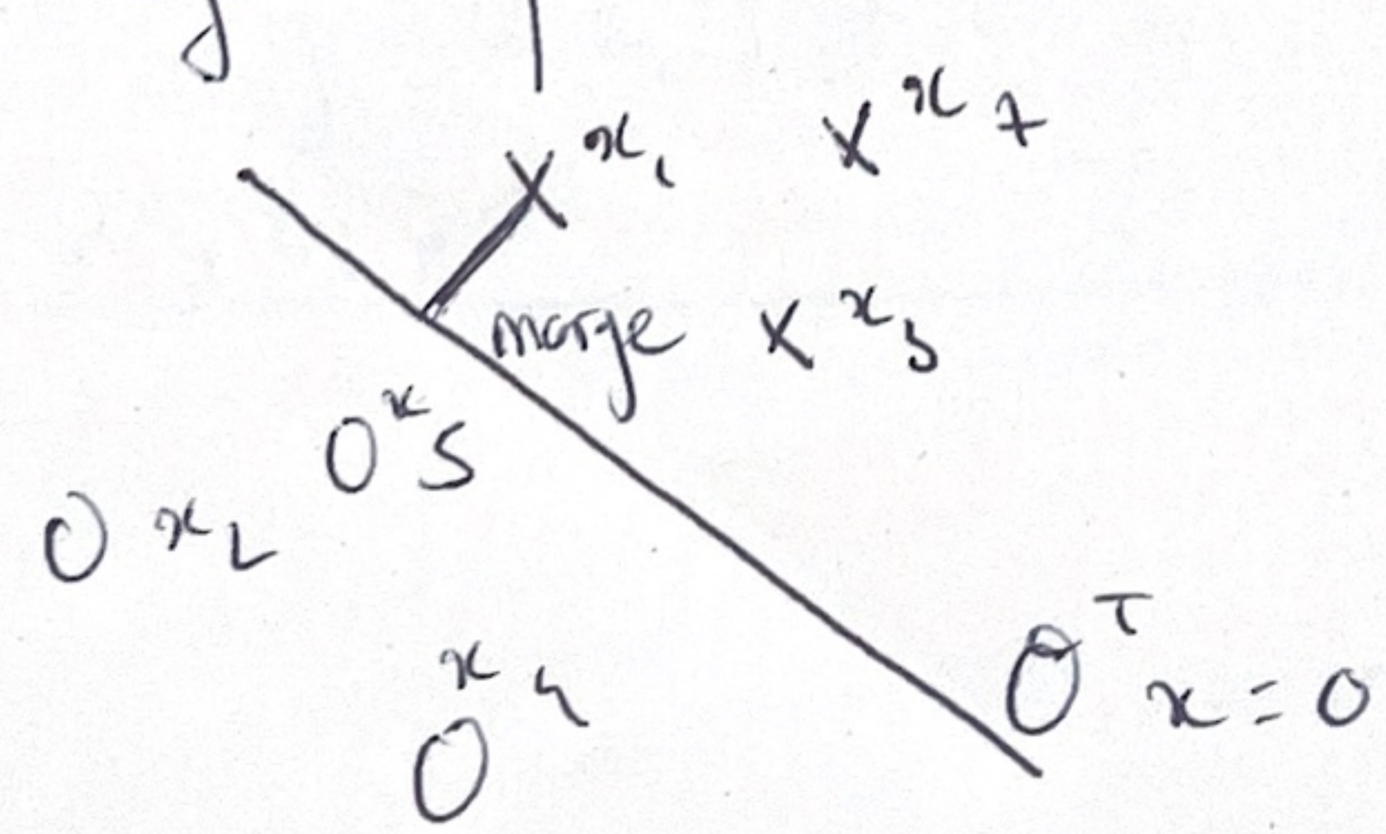
<u>Expression de la distance</u> Soit $H(\theta)$ l'hyperplan d'équation $x^T\theta = 0$.

alors, $\forall x, \; dist(x, H(\theta)) = \dfrac{|\theta^T x|}{\|\theta\|_2}$

↳ en effet, $dist(x, H(\theta))^2 = \|x - P_{H(\theta)}(x)\|^2$

$$= \Big\| \underbrace{\left(\frac{\theta}{\|\theta\|_2}\right)}_{\text{base orthonormée de } H(\theta)^\perp}^T x \Big\|^2$$

<u>Interprétation géométrique</u> :



Marge la plus grande possible

$$\max_{\theta\ \cancel{\text{séparateurs parfaits}}}\ \min_{i\in\{1,\dots,n\}} \frac{y_i(\theta^T x_i)}{\|\theta\|_2}$$

ce qui est équivalent à

$$\boxed{\min_{\theta} \frac{1}{2}\|\theta\|_2^2 \text{ tq } \forall i,\ y_i(\theta^T x_i) \geq 1}$$

II - Cas linéaire non séparable

Dans le cas où le problème n'est pas linéairement séparable (i.e $\forall \theta, \exists i$ tq $y_i(\theta^T x_i) \leq 0$), il est possible d'introduire des variables mesurant la sous-optimalité de chaque contrainte

$$\boxed{\begin{aligned} &\min_{\theta,\xi} \frac{1}{2}\|\theta\|_2^2 + C\sum_{i=1}^{n}\xi_i \\ &\text{tq } \forall i,\ y_i(\theta^T x_i) \geq 1-\xi_i \\ &\quad \forall i,\ \xi_i \geq 0 \end{aligned}}$$

et en notant $\lambda = \frac{1}{nC}$ et $\xi_i := \left(1 - y_i(\theta x_i)\right)_+$, le problème est équivalent à

$$\boxed{\min_{\theta} \frac{1}{n}\sum_{i=1}^{n}\left(1 - y_i(\theta^T x_i)\right)_+ + \frac{\lambda}{2}\|\theta\|_2^2}$$

III - ~~Dual problem~~ Problème dual

$$\boxed{\text{primal : } \begin{array}{l} \min\limits_{\theta, \xi} \frac{1}{2}\|\theta\|_2^2 + C\sum\limits_{i=1}^{n}\xi_i \\ \text{tq } \forall i,\ y_i \theta^T x_i + \xi_i - 1 \geq 0 \\ \quad\ \ \forall i,\ \xi_i \geq 0 \end{array}}$$

problème convexe à contraintes linéaires.

Lagrangien (multiplicateurs de Lagrange $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n \geq 0$)

$$\mathcal{L} = \frac{1}{2}\|\theta\|_2^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\left(y_i\theta^T x_i + \xi_i - 1\right) - \sum_{i=1}^{n}\beta_i\xi_i$$

* Minimisation par rapport à $\xi_1, \dots, \xi_n$:

Le Lagrangien est égal à $-\infty$ si $\exists i$ tq $\alpha_i + \beta_i \neq C$, et ~~est~~ sinon les termes en ξ disparaissent.

⑤ * Minimisation par rapport à θ :

Forme close $\theta^* = \sum_{i=1}^{n} \alpha_i y_i x_i$

Donc le problème dual est

$$\begin{cases} \max\limits_{\alpha \geq 0, \beta \geq 0} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \\ \text{s.t. } \alpha_i + \beta_i = C \end{cases}$$

ce qui est équivalent à

$$\boxed{\begin{aligned} &\max_{\alpha \geq 0} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \\ &\text{s.t. } 0 \leq \alpha_i \leq C \end{aligned}}$$

Problème dual

Par dualité faible Dual $\leq$ Primal

Condition de Slater : Il existe des points strictement faisables donc la condition de Slater s'applique $\Rightarrow$ Il y a dualité forte

(6)

Donc $\boxed{\text{Primal} = \text{Dual}}$

et le dual est un problème de programmation quadratique à contraintes de type "boite".

III La version avec noyaux

Soit H un espace de Hilbert et $\phi : X \to H$.

Soit $K : X \times X \to \mathbb{R}$

$$x, y \mapsto \langle \phi(x), \phi(y) \rangle_H$$

(D'après le théorème d'Aronszajn, les deux définitions sont équivalentes)

Problème : $\hat{\theta}_\lambda \in \underset{\theta \in H}{\text{argmin}} \frac{1}{n} \sum_{i=1}^{n} \left(1 - y_i \langle \theta, \phi(x_i) \rangle_H \right)_+ + \frac{\lambda}{2} \|\theta\|_H^2$

Théorème de représentation :

Il faut chercher $\hat{\theta}_\lambda$ de la forme $\alpha_1 \phi(x_1) + \dots + \alpha_n \phi(x_n)$, ce qui équivaut à chercher une fonction de classification de la forme $x \mapsto \text{sgn}\left(\alpha_1 \langle \phi(x_1), \phi(x) \rangle + \dots + \alpha_n \langle \phi(x_n), \phi(x) \rangle\right)$

Réécriture du problème :

$$\boxed{\hat{\alpha}_\lambda \in \arg\min_{\alpha} \frac{1}{n}\sum_{i=1}^{n}\left(1 - y_i (K\alpha)_i\right)_+ + \frac{\lambda}{2}\alpha^T K \alpha \qquad \text{où } K = \left(k(x_i, x_j)\right)_{i,j}}$$

Introduction des variables "Slack"

$$\boxed{\begin{aligned} &\min_{\alpha, \xi} \ \frac{1}{n}\sum_{i=1}^{n} \xi_i + \frac{\lambda}{2}\alpha^T K \alpha \\ &\text{tq} \quad \forall i,\ \xi_i \geq 1 - y_i (K\alpha)_i \\ &\qquad\ \ \forall i,\ \xi_i \geq 0 \end{aligned}}$$

Lagrangien :

$\mu, \nu \geq 0$ multiplicateurs de Lagrange.

$$L = \frac{1}{n}\sum_{i=1}^{n}\xi_i + \frac{\lambda}{2}\alpha^T K \alpha - \sum_{i=1}^{n}\mu_i\left(y_i (K\alpha)_i + \xi_i - 1\right) - \sum_{i=1}^{n}\nu_i \xi_i$$

$$= \frac{1}{n}\sum_{i=1}^{n}\xi_i + \frac{\lambda}{2}\alpha^T K \alpha - \left(\operatorname{diag}(y)\mu\right)^T K\alpha - (\mu+\nu)^T \xi + \mu^T \mathbb{1}$$

⑧

Minimisation en α $\Rightarrow$ $\alpha = \dfrac{\mathrm{diag}(y)\mu}{\lambda} + \varepsilon$ avec $\varepsilon \in \mathrm{Ker}(K)$ (1)

mais comme ça ne change rien, on peut prendre $\varepsilon = 0$.

Minimisation en ξ $\Rightarrow$ $-\infty$ si $\frac{1}{n} - \mu - \nu \neq 0$

et tous les termes en ξ disparaissent sinon.

d'où

Problème dual : $\begin{cases} \max \sum_{i=1}^{n} \mu_i - \frac{1}{2\lambda} \sum_{i,j=1}^{n} y_i y_j \mu_i \mu_j K(x_i, x_j) \\ \text{tq } \mu \geq 0, \nu \geq 0, \mu + \nu = \frac{1}{n} \end{cases}$

ou encore :

$$\boxed{\begin{array}{l} \max \sum_{i=1}^{n} \mu_i - \frac{1}{2\lambda} \sum_{i,j=1}^{n} y_i y_j \mu_i \mu_j K(x_i, x_j) \\ \text{tq } 0 \leq \mu \leq \frac{1}{n} \end{array}}$$

problème dual

9

Reformulation avec variables primales

D'après (1), il est possible de reformuler le problème dual à l'aide de α :

$$\boxed{\begin{array}{ll} \max\limits_{\alpha} & 2\alpha^T y - \alpha^T K \alpha \\ \text{tq} & 0 \leq y\alpha \leq \frac{1}{2n} \end{array}}$$

Et encore une fois, il y a dualité forte pour les mêmes raisons.