SVM et programmation quadratique

Données : $\{(a_i, b_i)\}_{i=1..m}$ $\quad a_i \in \mathbb{R}^d \quad b_i \in \{-1,+1\}$

But : Trouver un modèle qui associe chaque a_i à b_i, et donc "sépare les +1 des -1"

① Cas linéaire

On considère le modèle $a \mapsto a^T x$ paramétré par x.

On cherche $x \in \mathbb{R}^d$ tel que

$$a_i^T x > 1 \quad \text{si} \quad b_i = 1$$

$$a_i^T x < -1 \quad \text{si} \quad b_i = -1$$

Fonction de perte "hinge loss"

$$\ell(a_i, b_i; x) = \max(1 - b_i a_i^T x, 0)$$

✓ Si $a_i^T x > 1$ et $b_i = 1$, $\ell(a_i, b_i; x) = 0$

✓ Si $a_i^T x < -1$ et $b_i = -1$, $\ell(a_i, b_i; x) = 0$

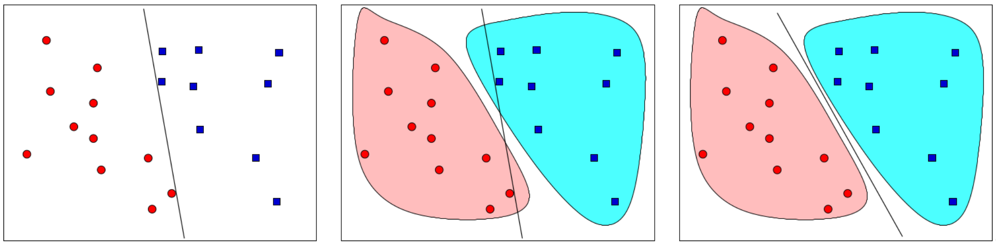
✗ Si $a_i^T x > 1$ et $b_i = -1$, $\ell(a_i, b_i; x) > 2$

✗ Si $a_i^T x < -1$ et $b_i = 1$, $\ell(a_i, b_i; x) > 2$

? Si $|a_i^T x| \leq 1$, $\ell(a_i, b_i; x) \in [0, 2]$

Problème de classification : SVM linéaire

$$(\mathrm{SVM}) \quad \underset{x \in \mathbb{R}^d}{\text{minimiser}} \quad \underbrace{\frac{1}{n}\sum_{i=1}^{n}}_{\text{Moyenne sur toutes les données}} \underbrace{\max(1 - b_i a_i^T x, 0)}_{\text{fonction de perte (erreur au point } i)} + \underbrace{\frac{\lambda}{2}\|x\|^2}_{\text{Régularisation (permet de séparer plus nettement les données)}}$$



$\underbrace{\hspace{6cm}}_{\text{Une Solution pour } \lambda = 0}$ $\underbrace{\hspace{3cm}}_{\text{La Solution pour } \lambda > 0}$

(Wright & Recht 2022)

Résoudre le problème (SVM)

a) Ajouter des variables

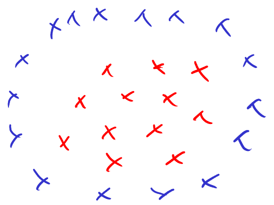
$$\underset{\substack{x \in \mathbb{R}^d \\ t \in \mathbb{R}^n}}{\text{minimiser}} \quad \frac{1}{n}\sum_{i=1}^{n} t_i + \frac{\lambda}{2}\|x\|^2 \quad \text{s.c.} \quad \begin{array}{ll} t_i \geq 1 - b_i a_i^T x & \forall i = 1..n \\ t_i \geq 0 & \forall i = 1..n \end{array}$$

↳ Les variables t_i remplacent le maximum.
⇒ La fonction objectif dépend linéairement de t
⇒ Rajoute des contraintes linéaires <u>en t et x</u>

↳ L'ensemble est un programme quadratique

b) Résoudre le problème via les points intérieurs

Extension aux séparateurs non linéaires



Idée : Se ramener au cas linéaire en se plaçant dans le bon espace
$\Rightarrow$ kernel SVM

Modèle : $a \mapsto \Psi(a)^T x$ où $\Psi(a)$ est une fonction non linéaire de a
$x \in \mathbb{R}^{\hat{d}}$ avec souvent $\hat{d} > d$
(polynômes, RBF

Problème de SVM

$$\underset{x \in \mathbb{R}^{\hat{d}}}{\text{minimiser}} \quad \frac{1}{m} \sum_{i=1}^{m} \max(1 - b_i \Psi(a_i)^T x, 0) + \frac{\lambda}{2} \|x\|^2$$

Reformulation en QP

$$\underset{\substack{x \in \mathbb{R}^{\hat{d}} \\ t \in \mathbb{R}^m}}{\text{minimiser}} \quad \frac{1}{m} \sum_{i=1}^{m} t_i + \frac{\lambda}{2} \|x\|^2 \quad \text{s.c.} \quad t_i \geq 1 - b_i \Psi(a_i)^T x, \quad t_i \geq 0 \quad i = 1..m$$

⊕ Toujours un programme quadratique
⊕ En passant au dual, on peut même réduire le coût de calcul