Machines à vecteurs supports et dualité

Samuel Gallay et Tristan Delbende

15 novembre 2022

Un peu de dualité

Problème primal

Minimiser $f(x)$ avec $h_i(x) \leq 0$ pour $1 \leq i \leq n$.

Lagrangien

$$L(x,u) = f(x) + \sum_{i=1}^{n} u_i h_i(x)$$

Fonction duale de Lagrange

$$g(u) = \min_x L(x,u)$$

Problème dual

Maximiser $g(u)$ avec $u_i \geq 0$ pour $1 \leq i \leq n$.

Théorème : dualité faible

$$g^* \leq f^*$$

Peut-on avoir dualité forte, *i.e.* $g^* = f^*$?

Conditions de Slater, et de Karush Kuhn et Tucker

Conditions de Slater

Les h_i et f sont convexes, et il existe x_0 tel que $h_i(x_0) < 0$ pour $0 \leq i \leq n$.

Théorème : dualité forte

Sous les conditions de Slater, $g^* = f^*$.

On se place dans les conditions de Slater, et on suppose que f et les h_i sont $\mathscr{C}^1$.

Théorème : KKT

x^* et u^* satisfont les conditions de KKT *ssi* x^* et u^* sont solutions du problème primal et dual.

Conditions de Karush Kuhn et Tucker

Stationarité :

$$\nabla f(x) + \sum_{i=1}^{n} u_i \nabla h_i(x) = 0$$

Complémentarité :

$$u_i h_i(x) = 0 \text{ pour tout } i$$

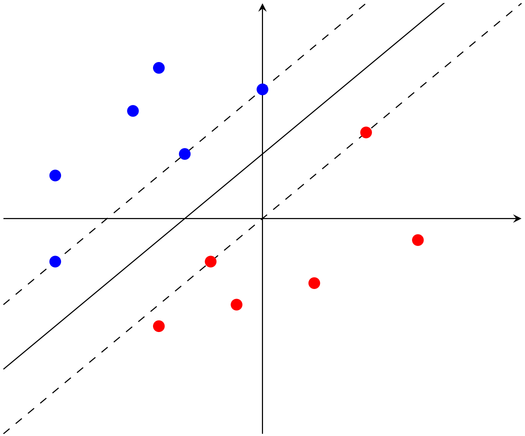
Faisabilité primale :

$$h_i(x) \leq 0 \text{ pour tout } i$$

Faisabilité duale :

$$u_i \geq 0 \text{ pour tout } i$$

Présentation du problème



Problème initial

Données : $(x_1, y_1) \dots (x_n, y_n) \in \mathbb{R}^d \times \{+1, -1\}$.
But : les séparer par un «bon» hyperplan H.

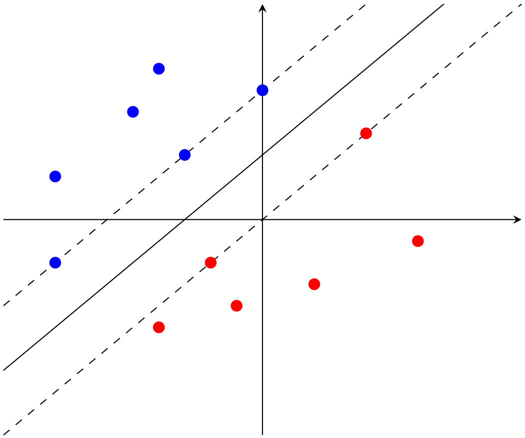
Données linéairement séparable

Hypothèse : existence d'un hyperplan qui sépare les données.

Marge d'un hyperplan H : $\langle w, x \rangle + b = 0$

$$m(H) = \min_i d(H, x_i) = \min_i \frac{|\langle w, x_i \rangle + b|}{\|w\|}$$

Présentation du problème II



Problème initial formalisé

Maximiser $m(H)$ où $H : \langle w, x \rangle + b = 0$
avec $\forall i, y_i(\langle w, x_i \rangle + b) > 0$.

Reformulation équivalente

Minimiser $\frac{1}{2}\|w\|^2$
avec $\forall i, 1 - y_i(\langle w, x_i \rangle + b) \leq 0$.

Dualité forte

On constate que les conditions de Slater sont vérifiées.

Dualité dans ce cas particulier I

Problème dual

Maximiser $\min_{w,b} L(w, b, u)$ sur u avec $\forall i, u_i \geq 0$

Problème dual équivalent I

Maximiser $L(w, b, u)$ sur w, b, et u
avec $\forall i, u_i \geq 0$ et $\partial_{w,b} L(w, b, u) = 0$.

Problème dual équivalent II

Maximiser $\sum_i u_i - \frac{1}{2} \left\| \sum_i u_i y_i x_i \right\|^2$
avec $\forall i, u_i \geq 0$ et $\sum_i u_i y_i = 0$.

$$L(w, b, u) = \frac{1}{2}\|w\|^2 + \sum_{i=1}^{n} u_i(1 - y_i(\langle w, x_i \rangle + b))$$

$$= \frac{1}{2}\|w\|^2 + \sum_{i=1}^{n} u_i - \left\langle w, \sum_i u_i y_i x_i \right\rangle + b \sum_i u_i y_i$$

Reformulation des contraintes :

$$w = \sum_i u_i y_i x_i \text{ et } \sum_i u_i y_i = 0$$

Alors

$$L(w, b, u) = -\frac{1}{2} \left\| \sum_i u_i y_i x_i \right\|^2 + \sum_i u_i$$

Récupération de la solution au problème primal

Existence de solutions

S'il existe au moins un point de chaque couleur, alors le problème primal possède une solution.

Programmation convexe quadratique

Le problème dual possède une solution *ssi* le problème primal possède une solution.

Problème dual équivalent II (rappel)

Maximiser $\sum_i u_i - \frac{1}{2}\left\|\sum_i u_i y_i x_i\right\|^2$

avec $\forall i, u_i \geq 0$ et $\sum_i u_i y_i = 0$.

Soit u^* une solution au problème dual, et (w^*, b^*) solution au problème primal.

Expression de w^* et b^* en fonction de u^*

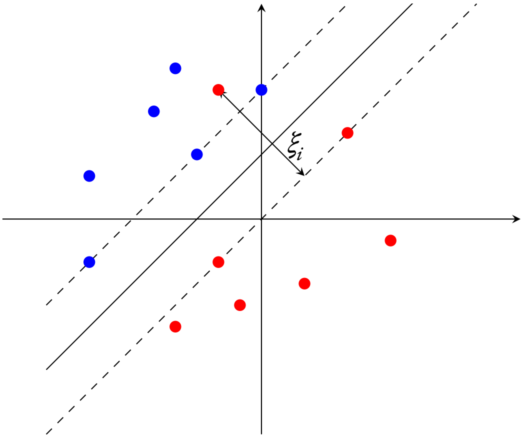
Stationarité : $w^* = \sum_i u_i^* y_i x_i$ et $\sum_i u_i y_i = 0$.

Complémentarité : $u_i^*(1 - y_i(\langle w^*, x_i \rangle + b^*)) = 0$.

Avantages du problème dual

Contraintes plus simples, permet l'utilisation de noyaux...

Cas des données non séparables



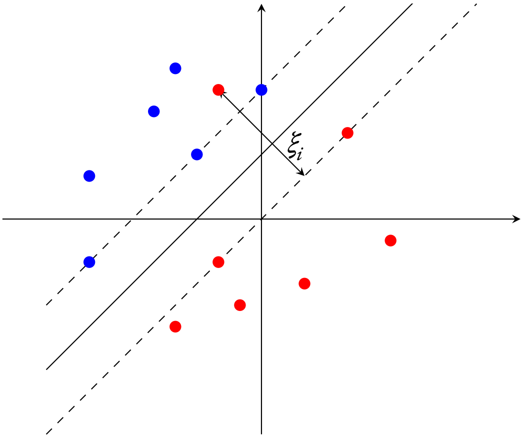
Problème initial

Données : $(x_1, y_1) \dots (x_n, y_n) \in \mathbb{R}^p \times \{+1, -1\}$.
But : les séparer par un hyperplan H en autorisant des erreurs.

Variables ressorts

On introduit les variables ressorts ξ_i qui modélisent l'écart entre x_i et la marge quand les x_i sont du mauvais côté de l'hyperplan.

Cas des données non séparables



Nouveau problème primal

Minimiser $\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i^d$
avec $\forall i$, $1 - y_i(\langle w, x_i\rangle + b) - \xi_i \leq 0$ et $\xi_i \geq 0$.

Dualité forte

On constate que les conditions de Slater sont vérifiées.

Existence de solution primale

Il existe une solution primale.

Problème dual

Problème dual simplifié, $d = 1$

$$\text{Maximiser} \sum_{i=1}^{n} u_i - \frac{1}{2} \left\| \sum_i u_i y_i x_i \right\|^2$$

$$\text{avec } 0 \leq u_i \leq C \text{ et } \sum_{i=1}^{n} u_i y_i = 0$$

Problème dual simplifié, $d = 2$

$$\text{Maximiser} \sum_{i=1}^{n} u_i - \frac{1}{2} \left\| \sum_i u_i y_i x_i \right\|^2 - \frac{1}{4C} \sum_i u_i^2$$

$$\text{avec } 0 \leq u_i \text{ et } \sum_{i=1}^{n} u_i y_i = 0$$

Dans les cas $d = 1$ ou $d = 2$

Le problème dual possède une solution *ssi* le problème primal possède une solution.

Récupération de la solution primale

$$w = \sum_i u_i y_i x_i$$

Il est aussi possible de récupérer b.