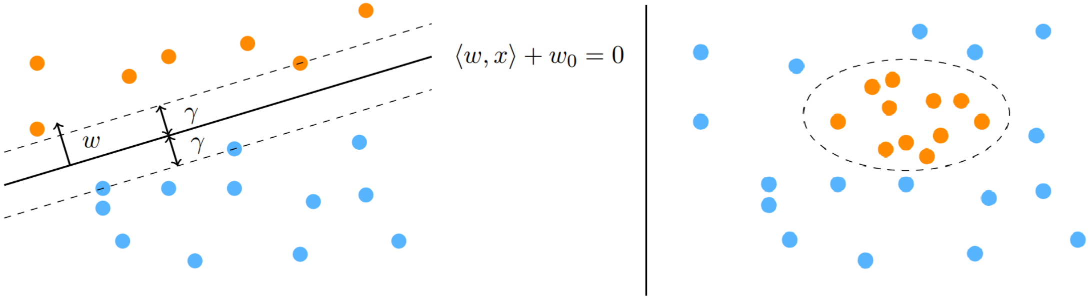
Espaces de Hilbert à noyau reproduisant et SVM non linéaires pour la classification supervisée binaire.

DONNART Titouan, RINGUEDE Marceau

Novembre 2022

Introduction



Plan

Plan

Espaces de Hilbert

Définition

Soit H un espace vectoriel sur $\mathbb{R}$ (*resp.* sur $\mathbb{C}$), $\langle \cdot \mid \cdot \rangle$ un produit scalaire (*resp.* hermitien) sur H et $\|\cdot\|$ la norme associée.

$(H, \langle \cdot \mid \cdot \rangle, \|\cdot\|)$ est **un espace de Hilbert** si $(H, \|\cdot\|)$ est complet.

Théorème de représentation de Riesz

Théorème de représentation de Riesz

Soit $(H, \|\cdot\|, \langle . \mid . \rangle)$ un espace de Hilbert et f une forme linéaire continue définie sur H. Alors il existe un unique vecteur $y \in H$ tel que, pour tout x :

$$f(x) = \langle x \mid y \rangle \,. \quad \text{De plus :} \quad \|f\|_{\mathcal{L}_c(H,\mathbf{K})} = \|y\|.$$

Plan

Noyaux définis positifs

Définition

Une application : $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ est un **noyau défini positif** sur l'ensemble $\mathcal{X}$ si :

- Elle est symétrique : pour tout $(x, y) \in \mathcal{X}$ on a $K(x, y) = K(y, x)$.
- Pour tous $N \in \mathbb{N}$, $(x_1, x_2, \ldots, x_N) \in \mathcal{X}^N$ et $(a_1, a_2, \ldots, a_N) \in \mathbb{R}^N$ on a :

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j K(x_i, x_j) \geq 0.$$

Exemples

Exemples

- Si $\mathcal{X} = \mathbb{R}^d$ et que $K : \mathcal{X}^2 \to \mathbb{R}$ est définie par :

$$K(x, y) = \langle x \mid y \rangle\,, \quad \forall (x, y) \in \mathcal{X}^2.$$

  Alors K est un noyau défini positif appelé le **noyau linéaire**

- Plus généralement, si $\mathcal{X}$ est un ensemble quelconque, $\phi : \mathcal{X} \to \mathbb{R}^d$ et que $K : \mathcal{X}^2 \to \mathbb{R}$ est définie par :

$$K(x, y) = \langle \phi(x) \mid \phi(y) \rangle \quad \forall (x, y) \in \mathcal{X}^2.$$

  Alors K est un noyau défini positif.

Exemples

Exemples

- Si $\mathcal{X} = [0,1]$ et que $K : [0,1]^2 \to \mathbb{R}$ est définie par :

$$K(x,y) = \min(x,y), \quad \forall (x,y) \in \mathcal{X}^2.$$

Alors K est un noyau défini positif appelé **noyau histogramme**.

- Si $\mathcal{X} = \mathbb{R}$ et que $K : \mathbb{R}^2 \to \mathbb{R}$ est définie par :

$$K(x,y) = (1+xy)^d, \quad \forall (x,y) \in \mathcal{X}^2 \quad \text{et où } d \in \mathbb{N}^*.$$

Alors K est un noyau défini positif appelé **noyau polynomial**.

Construction de noyaux définis positifs

Propriétés

Si $(K_n)_{n\in\mathbb{N}}$ est une suite de noyaux définis positifs sur $\mathcal{X}$ alors :

- Pour tous $\lambda_1, \lambda_2 \geq 0$, $\lambda_1 K_1 + \lambda_2 K_2$ est un noyau défini positif.
- L'application $K_1 \times K_2$ est un noyau défini positif.
- Si pour tout $(x, y) \in \mathcal{X}^2$ la suite $(K_n(x, y))_{n\in\mathbb{N}}$ admet une limite $K(x, y)$. Alors K est un noyau défini positif.

Construction de noyaux définis positifs

Propriété

Si K_1 et K_2 sont des noyaux définis positifs sur $\mathcal{X}_1$ et $\mathcal{X}_2$ respectivement alors :

- La somme directe : $K_1 \oplus K_2 = K_1 + K_2$ est un noyau défini positif sur $(\mathcal{X}_1 \times \mathcal{X}_2) \times (\mathcal{X}_1 \times \mathcal{X}_2)$.
- Le produit tensoriel : $K_1 \otimes K_2 = K_1 \times K_2$ est un noyau défini positif sur $(\mathcal{X}_1 \times \mathcal{X}_2) \times (\mathcal{X}_1 \times \mathcal{X}_2)$.

Noyau Gaussien

Exemple

Si $\mathcal{X} = \mathbb{R}^d$ et que $K : \mathcal{X}^2 \to \mathbb{R}$ est définie par :

$$K(x, y) = e^{\frac{-\|x-y\|^2}{2\sigma^2}}, \quad \forall (x, y) \in \mathcal{X}^2.$$

Avec $\sigma \in \mathbb{R}$. Alors K est un noyau défini positif appelé **noyau gaussien**.

Plan

Espace de Hilbert à noyau reproduisant (RKHS)

Définition

Soit $(H, \|\cdot\|, \langle . \mid . \rangle)$ un espace de Hilbert de fonctions $f : \mathcal{X} \to \mathbb{R}$. On dit que H est un **espace de Hilbert à noyau reproduisant** si pour tout $x \in \mathcal{X}$ l'application

$$\delta_x : \begin{cases} H \to \mathbb{R} \\ f \mapsto f(x) \end{cases} \quad \text{est continue.}$$

Noyau reproduisant

Définition

Soit $(H, \|\cdot\|, \langle . \mid . \rangle)$ un espace de Hilbert de fonctions $f : \mathcal{X} \to \mathbb{R}$. On dit qu'une application $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ est un **noyau reproduisant** de H si :

- $\forall x \in \mathcal{X}, K(\cdot, x) \in H$,
- $\forall x \in \mathcal{X}, \forall f \in H, \langle f \mid K(\cdot, x) \rangle = f(x)$.

Quelques propriétés des RKHS

Théorème

Soit $(H, \|\cdot\|, \langle . \mid . \rangle)$ un espace de Hilbert de fonctions $f : \mathcal{X} \to \mathbb{R}$.

- Si un noyau reproduisant existe, alors il est **unique**.
- Un noyau reproduisant existe **si et seulement si** H est un espace de Hilbert à noyau reproduisant.
- Un noyau reproduisant est un noyau défini positif.

Exemple de RKHS

Exemple

Soit $\mathcal{X} = \mathbb{R}^d$. Soit H l'ensemble des formes linéaires (continues) sur $\mathbb{R}^d$. D'après le théorème de Riesz on peut identifier H à $\mathbb{R}^d$ c'est donc un espace de Hilbert muni du produit scalaire :

$$\langle f \mid g \rangle_H = \langle y_f \mid y_g \rangle_{\mathbb{R}^d}$$

Où y_f et y_g sont les vecteurs associés à f et g par le théorème de Riesz. On vérifie que H est un RKHS ayant le noyau linéaire $K(x, y) = \langle x \mid y \rangle_{\mathbb{R}^d}$ comme noyau reproduisant.

Plan

Plan

Construction d'un RKHS

Théorème d'Aronszajn

Soit $\mathcal{X}$ un espace quelconque. Si $K : \mathcal{X}^2 \mapsto \mathbb{R}$ est un noyau défini positif, alors il existe un RKHS H qui possède K comme noyau reproduisant.

Preuve

- Soit H_0 le sous-espace vectoriel de $\mathbb{R}^{\mathcal{X}}$ engendré par les fonctions $(K_x)_{x\in\mathcal{X}}$ où :

$$K_x : \begin{cases} \mathcal{X} \longrightarrow \mathbb{R} \\ \ t \longmapsto K(x,t). \end{cases}$$

- Pour $f, g \in H_0$ s'écrivant :

$$f = \sum_{i=1}^{n} a_i K_{x_i} \quad g = \sum_{j=1}^{m} b_j K_{y_j}$$

Montrons que la quantité suivante définit bien un produit scalaire sur H_0 :

$$\langle f \mid g \rangle_{H_0} = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j K(x_i, y_j).$$

Preuve

- Soit H_0 le sous-espace vectoriel de $\mathbb{R}^{\mathcal{X}}$ engendré par les fonctions $(K_x)_{x\in\mathcal{X}}$ où :

$$K_x : \begin{cases} \mathcal{X} \longrightarrow \mathbb{R} \\ \quad t \longmapsto K(x,t). \end{cases}$$

- Pour $f, g \in H_0$ s'écrivant :

$$f = \sum_{i=1}^{n} a_i K_{x_i} \quad g = \sum_{j=1}^{m} b_j K_{y_j}$$

Montrons que la quantité suivante définit bien un produit scalaire sur H_0 :

$$\langle f \mid g \rangle_{H_0} = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j K(x_i, y_j).$$

Preuve

- $(H_0, \langle \cdot \mid \cdot \rangle_{H_0})$ un espace pré-hilbertien.
- Soit H l'espace de Hilbert obtenu en complétant H_0 par les limites des suites de Cauchy (dans H_0).
- On étend le produit scalaire de H_0 par passage à la limite, en un produit scalaire $\langle \cdot \mid \cdot \rangle_H$.
- On vérifie que K est bien un noyau reproduisant sur l'espace de Hilbert H.

Preuve

- $(H_0, \langle \cdot \mid \cdot \rangle_{H_0})$ un espace pré-hilbertien.
- Soit H l'espace de Hilbert obtenu en complétant H_0 par les limites des suites de Cauchy (dans H_0).
- On étend le produit scalaire de H_0 par passage à la limite, en un produit scalaire $\langle \cdot \mid \cdot \rangle_H$.
- On vérifie que K est bien un noyau reproduisant sur l'espace de Hilbert H.

Preuve

- $(H_0, \langle \cdot \mid \cdot \rangle_{H_0})$ un espace pré-hilbertien.
- Soit H l'espace de Hilbert obtenu en complétant H_0 par les limites des suites de Cauchy (dans H_0).
- On étend le produit scalaire de H_0 par passage à la limite, en un produit scalaire $\langle \cdot \mid \cdot \rangle_H$.
- On vérifie que K est bien un noyau reproduisant sur l'espace de Hilbert H.

Preuve

- $(H_0, \langle \cdot \mid \cdot \rangle_{H_0})$ un espace pré-hilbertien.
- Soit H l'espace de Hilbert obtenu en complétant H_0 par les limites des suites de Cauchy (dans H_0).
- On étend le produit scalaire de H_0 par passage à la limite, en un produit scalaire $\langle \cdot \mid \cdot \rangle_H$.
- On vérifie que K est bien un noyau reproduisant sur l'espace de Hilbert H.

Théorème de Mercer

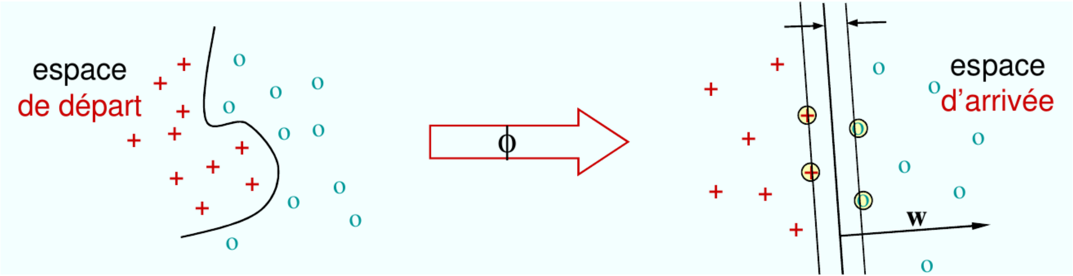
Théorème de Mercer

Soit $\mathcal{X}$ un compact de $\mathbb{R}^d, d \in \mathbb{N}^*$. Si $K : \mathcal{X}^2 \to \mathbb{R}$ est un noyau défini positif, alors il existe un espace de Hilbert $(H, \langle \cdot \mid \cdot \rangle, \|\cdot\|)$ et une application $\phi : \mathcal{X} \to H$ telle que pour tous $x, y \in \mathcal{X}$:

$$K(x, y) = \langle \phi(x) \mid \phi(y) \rangle .$$

Plan

Principe

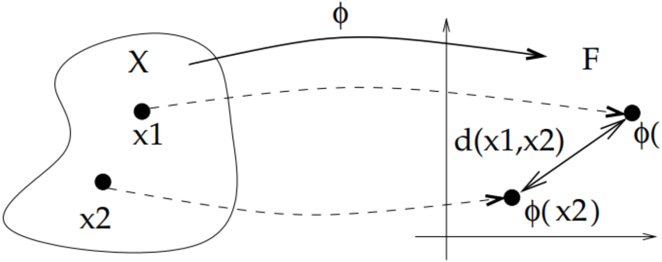


Si notre problème ne fait qu'apparaître que des produits scalaires entre les échantillons de données, comme

$$K(x, y) = \langle \phi(x) \mid \phi(y) \rangle ,$$

on peut directement l'étudier implicitement dans notre espace d'arrivée H.

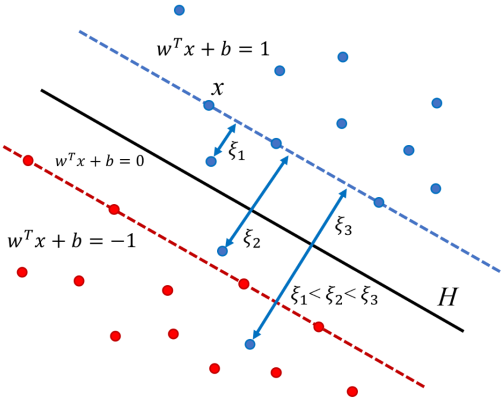
Exemple : calcul de distances



$$\forall (x_1, x_2) \in \mathcal{X} : d(x_1, x_2)^2 = K(x_1, x_1) + K(x_2, x_2) - 2K(x_1, x_2).$$

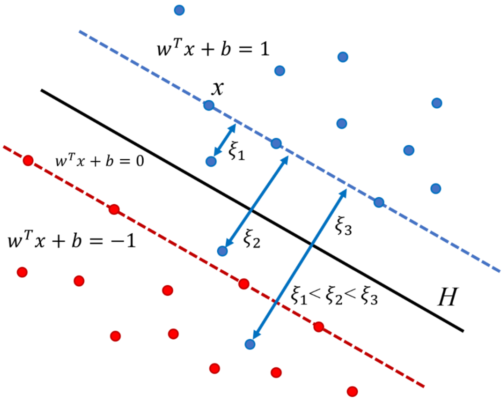
Plan

Reprise des SVM linéaires



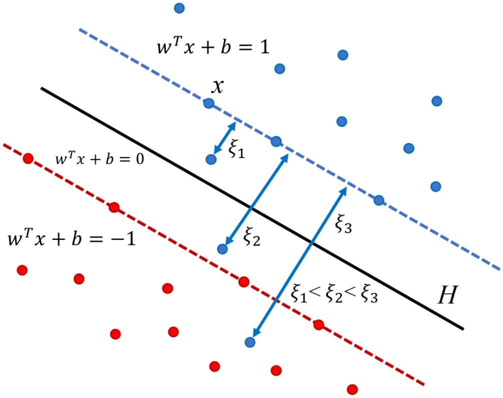
- $\mathcal{X} = \{(x_i, y_i), i \in \{1, \ldots, n\}\}$.
- $w = \sum_{i=1}^{n} \alpha_i y_i x_i, \quad f(x) = \sum_{i=1}^{n} \alpha_i y_i x_i^T x + b.$
- $\begin{cases} \max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j, \\ C \geq \alpha_i \geq 0, \quad \text{et} \quad \sum_{i=1}^{n} \alpha_i y_i = 0. \end{cases}$
- Vecteurs de support : $y_i f(x_i) = 1$.

Reprise des SVM linéaires



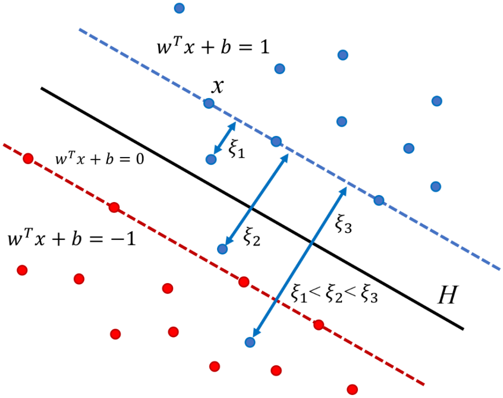
- $\mathcal{X} = \{(x_i, y_i), i \in \{1, \dots, n\}\}$.
- $w = \sum_{i=1}^{n} \alpha_i y_i x_i, \quad f(x) = \sum_{i=1}^{n} \alpha_i y_i x_i^T x + b.$
- $$\begin{cases} \max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j, \\ C \geq \alpha_i \geq 0, \quad \text{et} \quad \sum_{i=1}^{n} \alpha_i y_i = 0. \end{cases}$$
- Vecteurs de support : $y_i f(x_i) = 1$.

Reprise des SVM linéaires



- $\mathcal{X} = \{(x_i, y_i), i \in \{1, \dots, n\}\}$.
- $w = \sum_{i=1}^{n} \alpha_i y_i x_i, \quad f(x) = \sum_{i=1}^{n} \alpha_i y_i x_i^T x + b.$
- $$\begin{cases} \max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j, \\ C \geq \alpha_i \geq 0, \quad \text{et} \quad \sum_{i=1}^{n} \alpha_i y_i = 0. \end{cases}$$
- Vecteurs de support : $y_i f(x_i) = 1$.

Reprise des SVM linéaires



- $\mathcal{X} = \{(x_i, y_i), i \in \{1, \dots, n\}\}$.
- $w = \sum_{i=1}^{n} \alpha_i y_i x_i, \quad f(x) = \sum_{i=1}^{n} \alpha_i y_i x_i^T x + b.$
- $$\begin{cases} \max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j, \\ C \geq \alpha_i \geq 0, \quad \text{et} \quad \sum_{i=1}^{n} \alpha_i y_i = 0. \end{cases}$$
- Vecteurs de support : $y_i f(x_i) = 1$.

Méthode de résolution

Il existe un KRHS H tel que le problème dual s'écrive :

$$\begin{cases} \max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \langle \phi(x_i) \mid \phi(x_j) \rangle, \\ C \geq \alpha_i \geq 0, \forall i \in \{1, \ldots, n\}, \quad \sum_{i=1}^{n} \alpha_i y_i = 0. \end{cases}$$

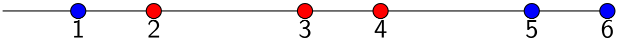
Donc en notant K le noyau reproduisant de H :

$$\begin{cases} \max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ C \geq \alpha_i \geq 0, \ \forall i \in \{1, \ldots, n\}, \\ \sum_{i=1}^{n} \alpha_i y_i = 0. \end{cases}$$

La fonction de décision est donnée par :

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b.$$

Exemple simple

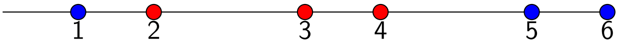


- On va utiliser un noyau polynôme de degré 2 : $K(x, y) = (xy + 1)^2$.
- On prend $C = 100$.

On considère les éléments :

- $(x_1 = 1, y_1 = 1)$,
- $(x_2 = 2, y_2 = -1)$,
- $(x_3 = 4, y_3 = -1)$,
- $(x_4 = 5, y_4 = -1)$,
- $(x_5 = 7, y_5 = +1)$,
- $(x_6 = 8, y_6 = +1)$,

Exemple simple



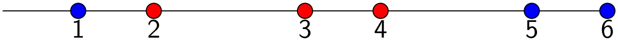
- $k(x, y) = (xy + 1)^2$.
- $C = 100$.

On cherche les (α_i) qui vérifient :

$$\begin{cases} \max_{\alpha} \sum_{i=1}^{6} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{6} \alpha_i \alpha_j y_i y_j (1 + x_i x_j)^2, \\ \\ 0 \leqslant \alpha_i \leqslant 100, \ \forall i \in \{1, \ldots, 6\}, \quad \sum_{i=1}^{6} \alpha_i y_i = 0. \end{cases}$$

On trouve $\alpha_1 = 4.53, \quad \alpha_2 = 4.08, \quad \alpha_3 = 0, \quad \alpha_4 = 0, \quad \alpha_5 = 0.25, \quad \alpha_6 = 0.$

Exemple simple



- $k(x, y) = (xy + 1)^2$.
- $C = 100$.

On a $\alpha_1 = 4.53, \quad \alpha_2 = 4.08, \quad \alpha_3 = 0, \quad \alpha_4 = 0, \quad \alpha_5 = 0.25, \quad \alpha_6 = 0.$
Donc :

$$f(x) = 4.53(1 + x)^2 - 4.08(2x + 1)^2 + 0.25(7x + 1)^2 + b.$$

Or f vérifie :

$$f(1) = 1 = f(7).$$

Finalement :

$$f(x) = 0.46x^2 - 3.76x + 4.3.$$

Mise en place sous Python