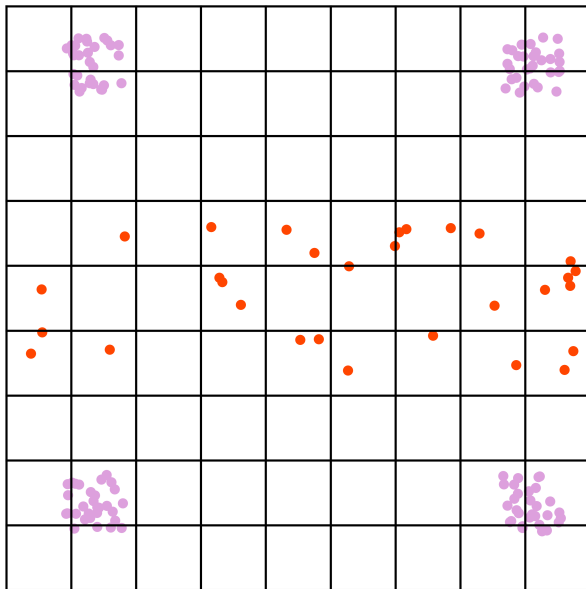


Introduction aux SVM non-linéaires

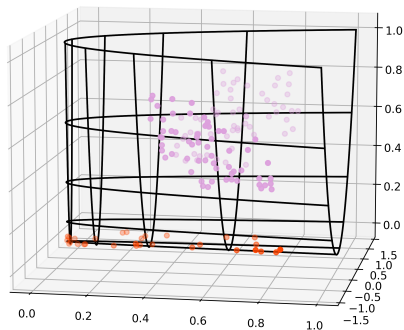
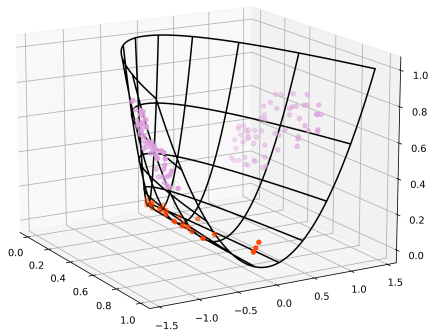
Corentin BERGERON & Pierre LARIVÉ

4 Février 2025

Motivation



Motivation





Méthode *SVM linéaire*

On maximise le lagrangien :

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

soumis aux contraintes :

$$\forall i \in \llbracket 1, l \rrbracket, \alpha_i \geq 0 \text{ et } \sum_{i=1}^l \alpha_i y_i = 0.$$

Le classifieur est alors de la forme :

$$f(x) = \text{sgn} \left(\sum_{i,j=1}^l \alpha_i y_i \langle x_i, x_j \rangle + b \right).$$



Méthode *SVM linéaire*

On maximise le lagrangien :

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

soumis aux contraintes :

$$\forall i \in \llbracket 1, l \rrbracket, \alpha_i \geq 0 \text{ et } \sum_{i=1}^l \alpha_i y_i = 0.$$

Le classifieur est alors de la forme :

$$f(x) = \text{sgn} \left(\sum_{i,j=1}^l \alpha_i y_i \langle x_i, x_j \rangle + b \right).$$



N.B.

Les formules du SVM linéaire dépendent des x_i seulement par des produits scalaires.



Méthode *SVM non linéaire*

- ▶ On cherche un espace de Hilbert $\mathcal{H}$ et une transformation non-linéaire $\Phi : \mathcal{X} \rightarrow \mathcal{H}$ tels que :

$$\forall x, y \in \mathcal{X}, \langle x, y \rangle_{\mathcal{X}} = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}}.$$

- ▶ On applique le SVM linéaire dans $\mathcal{H}$ pour obtenir le classifieur :

$$f(x) = \text{sgn} \left(\sum_{i,j=1}^l \alpha_i y_i \langle \Phi(x_i), \Phi(x_j) \rangle_{\mathcal{H}} + b \right).$$



Méthode *SVM non linéaire*

- ▶ On cherche un espace de Hilbert $\mathcal{H}$ et une transformation non-linéaire $\Phi : \mathcal{X} \rightarrow \mathcal{H}$ tels que :

$$\forall x, y \in \mathcal{X}, \langle x, y \rangle_{\mathcal{X}} = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}}.$$

- ▶ On applique le SVM linéaire dans $\mathcal{H}$ pour obtenir le classifieur :

$$f(x) = \text{sgn} \left(\sum_{i,j=1}^l \alpha_i y_i \langle \Phi(x_i), \Phi(x_j) \rangle_{\mathcal{H}} + b \right).$$



N.B.

- | Problème de la dimension de $\mathcal{H}$.

L'astuce : utiliser une application $k : \mathcal{X}^2 \rightarrow \mathbb{R}$ qui évite de passer par $\mathcal{H}$. Elle vérifie :

$$\forall x_i, x_j \in \mathcal{X}, k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle_{\mathcal{H}}$$

Le lagrangien du SVM dans $\mathcal{H}$ se réécrit :

$$\mathcal{L}(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j k(x_i, x_j).$$

Puis le classifieur obtenu est :

$$f(x) = \operatorname{sgn} \left(\sum_{i,j=1}^l \alpha_i y_i k(x_i, x_j) + b \right).$$

Définition 1 *Noyau défini positif*

Une fonction $k : \mathcal{X}^2 \rightarrow \mathbb{R}$ est un noyau (semi-)défini positif lorsque :

- ▶ (symétrique) $\forall x, y \in \mathcal{X}, k(x, y) = k(y, x)$
- ▶ (positivité) $\forall (a_i) \in \mathbb{R}^N, (x_i) \in \mathcal{X}^N,$

$$\sum_{i,j} a_i a_j k(x_i, x_j) \geq 0$$

Définition 3 *Noyau défini positif*

Une fonction $k : \mathcal{X}^2 \rightarrow \mathbb{R}$ est un noyau (semi-)défini positif lorsque :

- ▶ (symétrique) $\forall x, y \in \mathcal{X}, k(x, y) = k(y, x)$
- ▶ (positivité) $\forall (a_i) \in \mathbb{R}^N, (x_i) \in \mathcal{X}^N,$

$$\sum_{i,j} a_i a_j k(x_i, x_j) \geq 0$$

Définition 4 *RKHS*

Une espace de Hilbert $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$ est un RKHS s'il existe un noyau défini positif k tel que :

- ▶ $\forall x \in \mathcal{X}, k(x, \cdot) \in \mathcal{H}$
- ▶ (propriété de reproduction) $\forall f \in \mathcal{H}, x \in \mathcal{X}, f(x) = \langle f, k(x, \cdot) \rangle_{\mathcal{H}}$

Théorème 1 *Théorème d'existence*

Soit $k : \mathcal{X}^2 \rightarrow \mathbb{R}$ un noyau défini positif. On peut alors lui associer un (unique) RKHS $\mathcal{H}$ dont il est le noyau reproduisant.



N.B.

Dans un RKHS de noyau k , on a $k(x, y) = \langle k(x, \cdot), k(y, \cdot) \rangle_{\mathcal{H}}$.

Ainsi, pour tout noyau défini positif k , en posant $\Phi : x \mapsto k(x, \cdot)$, on a

$$k(x, y) = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}}.$$

Théorème 2 *Condition de Mercer*

Une application $k : \mathcal{X}^2 \longrightarrow \mathbb{R}$ est le noyau reproduisant d'un RKHS i.e. le résultat d'un produit scalaire si et seulement si la condition suivante est vérifiée :

$$\forall g \in L^2, \quad \iint k(x, y)g(x)g(y)dxdy \geq 0.$$



Méthode SVM non-linéaires

- ▶ On part de l'étude des SVM linéaires avec la formule

$$f(x) = \operatorname{sgn} \left(\sum_{i,j=1}^l \alpha_i y_i \langle x_i, x_j \rangle + b \right).$$

- ▶ On applique à notre jeu de données une transformation Φ , afin d'obtenir la formule

$$f(x) = \operatorname{sgn} \left(\sum_{i,j=1}^l \alpha_i y_i \langle \Phi(x_i), \Phi(x_j) \rangle_{\mathcal{H}} + b \right).$$

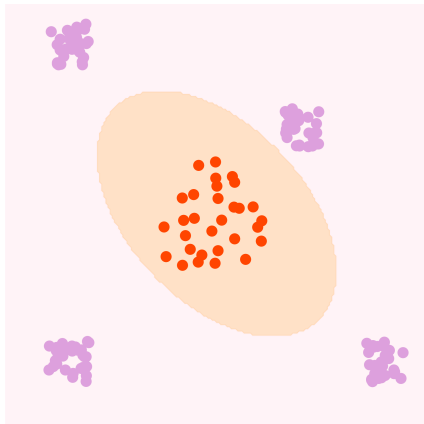
- ▶ Cependant, l'utilisation en pratique de Φ est compliquée (dim infinie), on utilise donc des noyaux reproduisants $k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle_{\mathcal{H}}$.

$$f(x) = \operatorname{sgn} \left(\sum_{i,j=1}^l \alpha_i y_i k(x_i, x_j) + b \right)$$

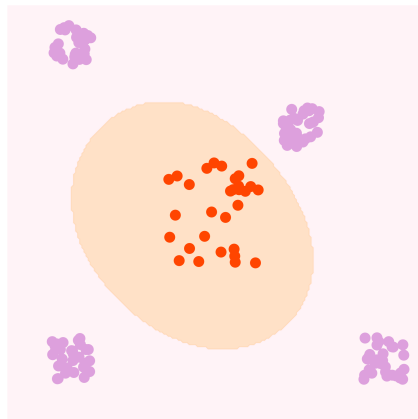
- ▶ La théorie des RKHS nous permet alors de dire qu'un tel k existe si et seulement si k est défini positif.

Exemples pour visualiser

Noyau polynomial (degré 2)

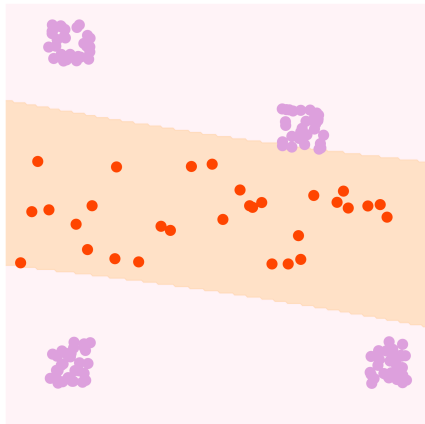


Noyau rbf

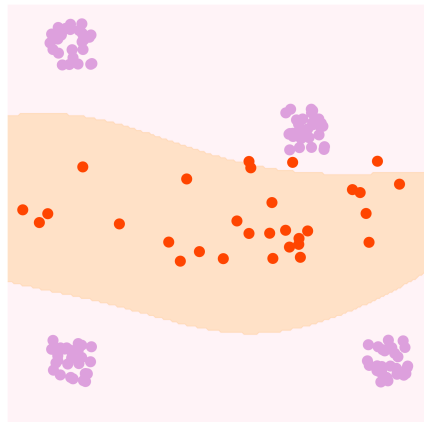


Exemples pour visualiser

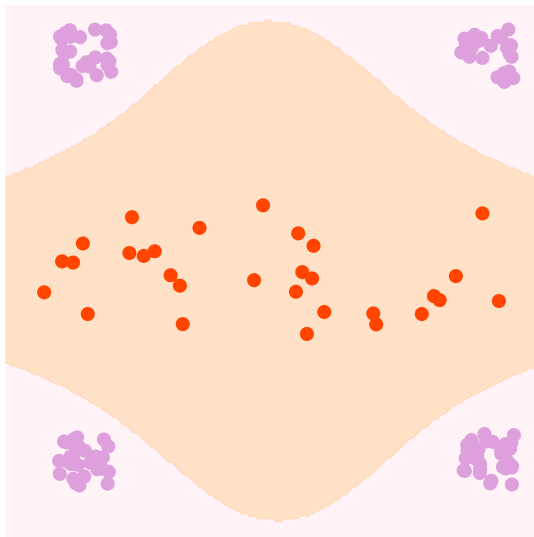
Noyau polynomial (degré 2)



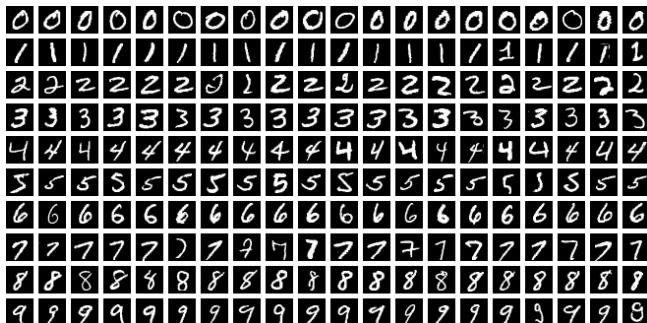
Noyau rbf



Noyau polynomial (degré 4)



Exemple d'application (banger en approche)



Situation : on veut que l'ordinateur reconnaisse le chiffre écrit sur l'image qu'on lui donne.



Exemple 1 *Classification MNIST*

Cadre :

- ▶ $\mathcal{X} = [0, 1]^{28 \times 28} = [0, 1]^{784}$
- ▶ $\mathcal{Y} = \{0, 1\}$ ou $\{0, \dots, 9\}$
- ▶ Noyaux utilisés : linéaire, polynomial, rbf

Méthode employée :

- ▶ On sépare le jeu de données en 2 : une partie d'entraînement et une de test ;
- ▶ On entraîne le SVM sur les données d'entraînement ;
- ▶ On vérifie les prédictions sur les données de test.

Noyau	Taux de réussite (binaire)	Taux de réussite (multi)
linéaire	0.9996	0.9351
polynomial	0.9993 ($d = 2$)	0.9739 ($d = 3$)
rbf	0.9997	0.9764

Merci pour votre écoute !