

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ ABDELHAMID BEN BADIS DE MOSTAGANEM
FACULTÉ DES SCIENCES EXACTES ET DE L'INFORMATIQUE
DÉPARTEMENT DE MATHÉMATIQUES ET INFORMATIQUE



Mémoire de fin d'étude

Master LMD

Pour l'obtention du diplôme de Master en Mathématiques

Option "Modélisation, Contrôle et Optimisation"

présentée par

Naima BENMEGHNIA

le 21 Mai 2017

Noyaux à base de polynômes orthogonaux pour les machines à vecteur de support (SVMs).

Encadreur	Abdessamad AMIR
Président	Omar BELHAMITI
Examinatrice	Naima ABLAOUI

Résumé

Dans ce mémoire, on a donné une vision purement mathématique des SVMs, cette méthode de classification est basée sur la recherche d'un hyperplan qui permet de séparer au mieux les ensembles des données. On a exposé le cas linéairement séparable et non-linéairement séparables qui nécessite l'utilisation des fonctions noyaux. Parmi les fonction noyaux, nous sommes particulièrement intéressés par les noyaux à base de polynômes orthogonaux de Jacobi, Hermite et Laguerre. Ce multiple choix de noyaux rend les SVMs plus intéressants et surtout plus riches puisqu'on peut toujours chercher de nouveaux noyaux qui peuvent être mieux adaptés à la tâche qu'on veut accomplir.

Mots clés : Les machines à vecteur de support, Ensemble d'apprentissage, Fonction noyau, Polynômes orthogonaux.

Remerciements

Tous mes remerciements s'adressent tous d'abord à tout puissant ALLAH, d'avoir guidé mes pas vers le chemin de savoir.

Je tiens à remercier mon encadreur Mr. Amir Abdsamad qui m'a aidé et conseillé durant ce travail.

Mes remerciements vont également aux membre de jury Mr. Omar Belhamiti et Mme Naima Ablaoui pour m'avoir honoré par leur évaluation de ce travail.

Un grand merci va également à la doctorante Abbassa Nadira et Dr. Bouzid Leila pour leur disponibilité et leur gentillesse.

En fin, un énorme merci à ma famille, mes amis et tous ceux qui m'ont aidé de près ou de loin.

Table des matières

Résumé	i
Remerciements	ii
Introduction	2
1 Formulation mathématique des SVMs	4
1 Cas des données linéairement séparable	4
1.1 Marge dure	5
1.2 Calcul de la marge	5
1.3 Marge souple	6
1.4 Formulation de problème dual	6
1.5 Vecteur de support	7
2 Cas des données non linéairement séparable	7
2.1 La transformation de problème	8
2 Les noyaux	10
3 Noyaux à base de polynômes orthogonaux	13
1 généralités	13
2 Exemples classiques des polynômes orthogonaux	14
2.1 Polynômes d'Hermite	14
2.2 Polynômes de Laguerre	16
2.3 Polynômes de Jacobi	18
3 La construction d'une fonction noyau par les polynômes orthogonaux . .	20
3.1 Noyau des polynômes orthogonaux	20
3.2 Noyau à base de polynômes orthogonaux pour les SVMs	22
4 Tests numériques	23
1 Méthode d'évaluation	23
2 Résultats	25
Conclusion	27

Liste des figures

1.1	Les hyperplans séparateurs	5
1.2	La marge souple	7
1.3	La transformation Φ pour le cas non linéairement séparable	8
4.1	Les données <i>twospirals</i>	24
4.2	Organigramme de la méthode d'évaluation	24
4.3	Les résultats obtenus.	25
4.4	La courbe de décision du noyau d'Hermite dans l'espace des entrées	25
4.5	La courbe de décision du noyau de Laguerre dans l'espace des entrées . . .	26
4.6	La courbe de décision du noyau de Jacobi dans l'espace des entrées	26

Introduction

Les machines à vecteurs de support, appelés aussi les séparateurs à vaste marge (en anglais, support vector machines (SVMs)), sont des algorithmes d'apprentissage, devenus très populaire depuis leurs apparitions dans le début des années 1990 par Vladimir Vapnik [11]. Les SVMs sont dans leurs origine utilisés pour la classification et la régression. Aujourd'hui, ils sont utilisés dans différents domaines de la science et d'ingénierie, tels que la Biologie, Bio-informatique, Reconnaissance des items, Diagnostic médical, Robotique Le succès de ces algorithmes est souvent justifié par les solides bases mathématiques sur lesquelles sont fondés. Les SVMs sont basés sur deux idées clés, qui permettent de traiter les problèmes de discrimination de données linéairement et non linéairement séparable en formulant un problème d'optimisation quadratique; La première idée est la notion de marge maximale, Les SVMs représentent un modèle discriminant qui tente à minimiser l'erreur de classification tout en maximisant la marge séparant les données de deux classes. La deuxième idée, est l'astuce noyau qui concerne beaucoup plus les données non linéairement séparables. Ces dernières sont projetées dans un espace de plus grand dimension appelé espace des caractéristiques de façon à ce que les données deviennent linéairement séparables, dans lequel opère plus facilement la séparation des deux classes. Un autre intérêt est l'identification des vecteurs de support qui sont les seules données nécessaires dans le calcul de l'hyperplan de séparation.

L'avantage des fonctions noyaux, réside d'abord dans les propriétés d'analyse fonctionnelle que ce type de fonctions possède telles que, les espaces noyaux reproduisant et l'universalité [10], qui permettent d'assurer la séparation dans l'espace des caractéristiques et faire tous les calculs directement à partir des données de l'espace d'entrée, sans évoquer la fonction de transformation très difficile à identifier. A partir de quelques propriétés de base, il est toujours possible de définir d'autres noyaux, ceci donne la possibilité aux praticiens de choisir le noyau donnant les meilleures performances numériques.

Récemment, une attention particulière a été donnée à la construction des noyaux sur la base des concepts de polynômes orthogonaux qui ont connu une utilisation répandue dans tous les domaines de la science et de l'ingénierie [3]. Ce travail a pour objectif, de proposer des fonctions noyaux pour les SVMs basées sur les polynômes orthogonaux d'Hermite, Laguerre et Jacobi. Les résultats obtenus montrent bien que cette classe de noyaux pourra conquérir les noyaux les plus utilisés dans le cadre des SVMs.

Ce mémoire est organisé en quatre chapitres. Le premier chapitre expose le fondement mathématique des SVMs vus comme un problème d'optimisation quadratique sous contraintes, ce problème peut être simplement énoncé comme suit : maximiser la marge entre les deux classes, tout en tolérant un minimum d'erreurs. Les notions de l'hyperplan optimal et l'astuce noyau réglant le problème de données linéairement et non linéairement séparable y sont introduites. Le deuxième chapitre présente le cadre théorique des noyaux et la manière de les construire ainsi que les noyaux les plus fréquemment utilisée pour les SVMs, on termine ce chapitre par le théorème classique de Mercer qui donne une caractérisation des noyaux sans passer par l'espace des caractéristiques. Dans le troisième chapitre, on commence par donner une première définition générale de polynômes orthogonaux [9], en suite, on présente des propriétés importantes qui caractérise trois familles classiques qui sont : les polynômes de Jacobi, Laguerre et Hermite [2], enfin, on termine par une description détaillée sur la manière de construire un noyau en se basant sur ces classes des polynômes orthogonaux [12],[8],[7],[5]. Dans le quatrième chapitre, on réalise une étude comparative entre les différent fonction noyau (d'Hermite, Laguerre et Jacobi) tout en effectuant des tests numériques sur un exemple académique de données non linéairement séparable appelé *twospirals data* [4], afin de bien évaluer les performances de chaque fonction noyau utilisée.

En fin, une conclusion générale et des perspectives sont évoquées.

Chapitre 1

Formulation mathématique des SVMs

Les SVMs sont un ensemble de techniques d'apprentissages destinées à résoudre des problèmes de classification et de régression. Dans ce travail on s'intéresse uniquement à la classification binaire, où les données viennent uniquement de deux classes différentes (+1 ou -1). L'idée des SVMs est de chercher une fonction classificatrice qui va séparer un ensemble d'apprentissage (pour lequel les données sont étiquetées), en maximisant la distance entre ces deux classes. Dans un premier lieu on s'intéressera aux données linéairement séparables.

1 Cas des données linéairement séparable

Soit $\{(x^i, y^i)\}_{i=1}^n$ un ensemble de n données d'apprentissage x^i et leurs étiquettes correspondantes y^i avec $x^i \in \mathbb{R}^d$ et $y^i = \pm 1$. On supposera que l'enveloppe convexe des deux classes d'ensemble est disjoint, le but est de chercher un hyperplan de séparation donné par la fonction

$$f(x) = \omega^T x + b,$$

où ω est un vecteur de $\mathbb{R}^d$ et b un réel. La fonction de décision f , peut être exprimé comme suit :

$$\begin{cases} f(x^i) = \omega^T x^i + b > 0, \text{ pour } y^i = +1, \text{ avec } i = 1, \dots, n \\ f(x^i) = \omega^T x^i + b < 0, \text{ pour } y^i = -1, \text{ avec } i = 1, \dots, n \end{cases}$$

on peut lui associer l'hyperplan séparateur :

$$H_0 := \{x \in \mathbb{R}^d : \omega^T x + b = 0\},$$

Il peut exister une infinité d'hyperplan permettant de séparer les deux classes, voir la figure 1.1. Le but des SVMs est de maximiser la distance entre les points les plus proche de l'hyperplan séparateur, cette distance est appelée "marge".

1. CAS DES DONNÉES LINÉAIREMENT SÉPARABLE

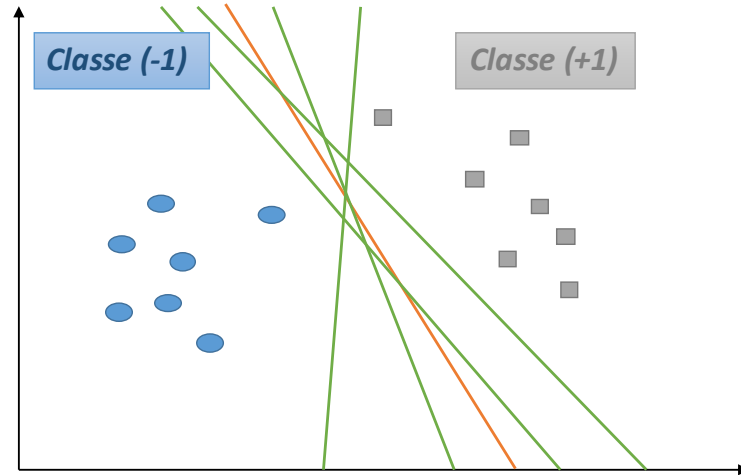


FIGURE 1.1 – Les hyperplans séparateurs

1.1 Marge dure

Signifie que tous les points sont bien classés, c-à-d $f(x^i) \geq +1$, pour $y^i = +1$ et aucun point x^i ne satisfait $\omega^T x^i + b = 0$, c-à-d $x^i \notin H_0$. Il convient alors d'utiliser la fonction de décision suivante :

$$\begin{aligned} f(x^i) &\geq +1, \text{ pour } y^i = +1, \text{ avec } i = 1, \dots, n \\ f(x^i) &\leq -1, \text{ pour } y^i = -1, \text{ avec } i = 1, \dots, n \end{aligned}$$

qu'on peut rassembler dans l'unique inégalité :

$$y^i(\omega^T x^i + b) \geq 1, i = 1, \dots, n$$

1.2 Calcul de la marge

Rappelons que la distance entre un point x quelconque et l'hyperplan H_0 est donnée par :

$$d(x, H_0) = \|x - P_{H_0}(x)\| = \frac{|f(x)|}{\|\omega\|} = \frac{|\omega^T x + b|}{\|\omega\|},$$

les hyperplans H_1 et H_2 définis par :

$$\begin{aligned} H_1 &: = \{x \in \mathbb{R}^d : \omega^T x + b = +1\} \\ H_2 &: = \{x \in \mathbb{R}^d : \omega^T x + b = -1\} \end{aligned}$$

sont parallèles à H_0 et sont appelés les hyperplans canoniques. La distance entre un point x situé dans H_1 (ou H_2) et H_0 est donnée par

$$d(x, H_0) = \frac{1}{\|\omega\|}.$$

1. CAS DES DONNÉES LINÉAIREMENT SÉPARABLE

Donc la marge (la distance entre les deux hyperplans H_1 et H_2) vaut $\frac{2}{\|\omega\|}$. La maximisation de cette quantité revient à minimiser l'inverse $\frac{\|\omega\|}{2}$, notons qu'il est plus facile de minimiser $\frac{\|\omega\|^2}{2}$ plutôt que $\frac{\|\omega\|}{2}$, finalement notre problème peut être formulé comme suit :

$$\begin{cases} \min_{\omega, b} \frac{1}{2} \omega^T \omega \\ y^i (\omega^T x^i + b) \geq 1, \text{ avec } i = 1, \dots, n \end{cases}$$

1.3 Marge souple

Les données d'apprentissage peuvent être mal classées (bruitées). La maximisation de la marge de séparation introduite précédemment peut conduire à un hyperplan qui sépare mal les données, afin de régler ce problème, on introduit la notion de classifieur à marge souple en introduisant des variables d'écart $\xi_i (i = 1, \dots, n)$ dans les contraintes qui deviennent :

$$\begin{cases} \omega^T x^i + b + \xi_i \geq +1, \text{ pour } y^i = +1, \text{ avec } i = 1, \dots, n \\ \omega^T x^i + b - \xi_i \leq -1, \text{ pour } y^i = -1, \text{ avec } i = 1, \dots, n \end{cases}$$

les ξ_i sont des réels positifs ($\xi_i = 0$, si x^i est bien classé; $\xi_i > 0$, si x^i est mal classé), la somme $\sum_{i=1}^n \xi_i$ est incorporée dans la fonction objectif afin de minimiser l'influence de la variable ξ sur les autres variables de décision ω et b , on obtient alors le problème suivant :

$$\begin{cases} \min \frac{1}{2} \omega^T \omega + c \sum_{i=1}^n \xi_i \\ y^i (\omega^T x^i + b) + \xi_i \geq 1, i = 1, \dots, n \\ \xi_i \geq 0, i = 1, \dots, n \end{cases}, \quad (1.1)$$

où c est un paramètre qui balance entre la maximisation de la marge et la minimisation de l'influence de ξ sur l'erreur de classification.

1.4 Formulation de problème dual

Le problème (1.1) est appelé Problème Primal SVM introduit par V. Vapnik [11], afin de lui associé son dual, on introduit les multiplicateurs de Karush Kuhn Tucker (KKT) $\alpha_i \geq 0$ et $\beta_i \geq 0$, pour chaque contraintes $i = 1, \dots, n$. Le Lagrangien de (1.1) est donné par :

$$L(\omega, b, \xi, \alpha, \beta) = \frac{1}{2} \omega^T \omega + c \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i (y_i (\omega^T x^i + b) + \xi_i - 1) - \sum_{i=1}^n \beta_i \xi_i.$$

Le problème (1.1) est convexe, la fonction objectif et les contraintes sont de classe C^∞ alors les conditions de Karush Kuhn Tucker (KKT) sont nécessaires et suffisantes [6] :

$$\left\{ \begin{array}{l} \frac{\partial L(\omega, b, \xi, \alpha, \beta)}{\partial \omega} = 0 \Rightarrow \omega - \sum_{i=1}^n \alpha_i y_i x^i = 0 \Rightarrow \omega = \sum_{i=1}^n \alpha_i y_i x^i \\ \frac{\partial L(\omega, b, \xi, \alpha, \beta)}{\partial b} = 0 \Rightarrow \sum_{i=1}^n \alpha_i y_i = 0 \\ \frac{\partial L(\omega, b, \xi, \alpha, \beta)}{\partial \xi_i} = 0 \Rightarrow c - \alpha_i - \beta_i = 0 \Rightarrow c = \alpha_i + \beta_i \Rightarrow c \geq \alpha_i, \text{ pour } i = 1, \dots, n \\ \alpha_i ((y_i (\omega^T x^i + b) + \xi_i - 1)) = 0, \text{ pour } i = 1, \dots, n \\ \beta_i \xi_i = 0, \text{ pour } i = 1, \dots, n \end{array} \right. \quad (1.2)$$

2. CAS DES DONNÉES NON LINÉAIREMENT SÉPARABLE

en réinjectant ces conditions dans le Lagrangien, on obtient le problème dual :

$$\begin{cases} \max -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i y_i \alpha_j y_j (x^i)^T x^j + \sum_{i=1}^n \alpha_i \\ \sum_{i=1}^n \alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq c, i = 1, \dots, n \end{cases}, \quad (1.3)$$

c'est aussi un problème quadratique, dépendant d'une seule variable α , avec une seule contrainte égalité et n contraintes en boîte.

1.5 Vecteur de support

Les vecteurs de support sont les données dont le multiplicateur de K-K-T est non nul ($\alpha_i > 0$). Les conditions de complémentarité de K-K-T et la contrainte de boîte nous permettent de distinguer deux types de vecteurs de support. Quand, $0 < \alpha_i < c$ alors $\beta_i > 0$ ce qui implique que $\xi_i = 0$, d'où $y_i(\omega^T x^i + b) = 1$, on déduit que x^i est un vecteur de support bien classé. Par ailleurs, si $\alpha_i = c$, on aura dans ce cas $\beta_i = 0$, dans le cas où $\xi_i > 0$, on déduit que x^i est un vecteur de support mal classé. Notons que lorsque $\alpha_i = 0$, alors on a forcément $\beta_i = c$, ce qui donne $\xi_i = 0$ d'où x^i est bien classé, la figure 1.2 illustre la situation. Des conditions de K-K-T (1.2) et la résolution du problème dual (1.3) on peut déduire le vecteur normal $\omega = \sum_{i=1}^n \alpha_i y_i x^i$. Un grand nombre de termes de cette somme sont nuls ($\alpha_i = 0$). Seuls les vecteurs de support contiennent les informations nécessaires à définir l'hyperplan de séparation.

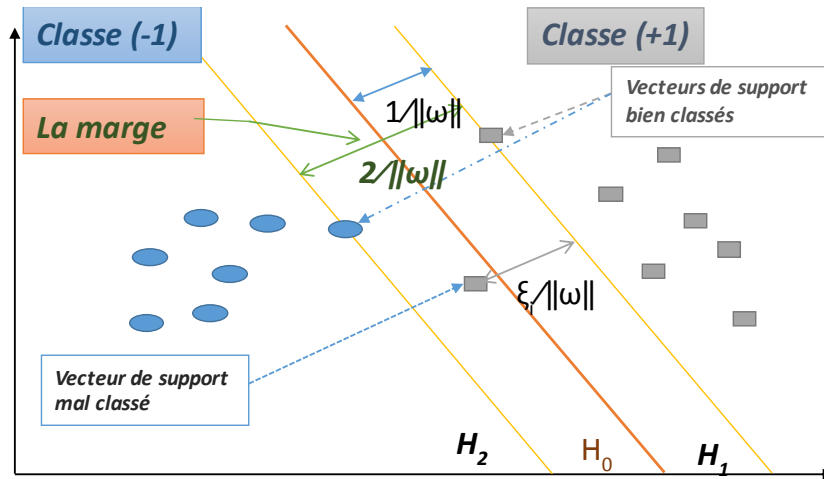


FIGURE 1.2 – La marge souple

2 Cas des données non linéairement séparable

En pratique, les données sont en général non linéairement séparable. L'idée est d'envoyer les données dans un autre espace de grand dimension (la figure 1.3), appelé espace caractéristique afin de pouvoir faire une séparation linéaire.

2. CAS DES DONNÉES NON LINÉAIREMENT SÉPARABLE

2.1 La transformation de problème

On définit une transformation Φ par :

$$\begin{aligned} \Phi: \mathbb{R}^d &\rightarrow H \\ x &\mapsto \Phi(x) \end{aligned}$$

ou H est l'espace caractéristique.

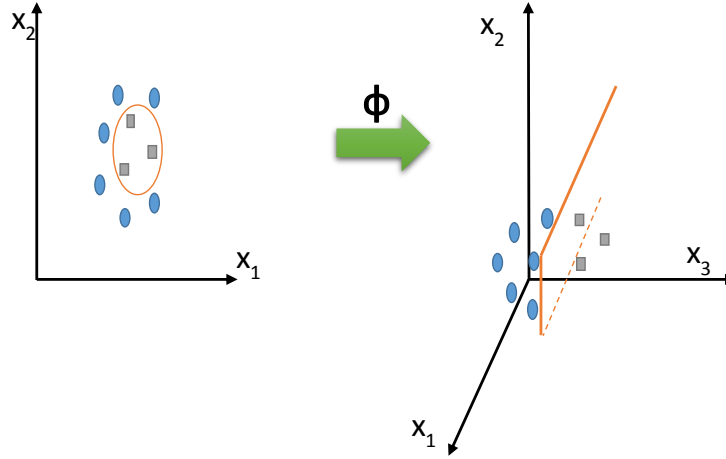


FIGURE 1.3 – La transformation Φ pour le cas non linéairement séparable

On considère maintenant l'ensemble $(\Phi(x^i), y_i) \in H \times \{-1, +1\}$, tout ce qui reste à faire serait de résoudre le problème dans H en remplaçant $\langle x^i, x^j \rangle$ par $\langle \Phi(x^i), \Phi(x^j) \rangle$, $i, j = 1, \dots, d$. On doit donc résoudre

$$\begin{cases} \max -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i y_i \alpha_j y_j \langle \Phi(x^i), \Phi(x^j) \rangle + \sum_{i=1}^n \alpha_i \\ \sum_{i=1}^n \alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq c, \quad i = 1, \dots, n \end{cases}, \quad (1.4)$$

Le classifieur f deviendra

$$f(x) = \sum_{i=1}^n \alpha_i y_i \langle \Phi(x^i), \Phi(x) \rangle_H + b$$

Notons que dans le problème (1.4), la fonction objectif dépend uniquement du $\langle \Phi(x^i), \Phi(x^j) \rangle$, la connaissance explicite de Φ n'est pas requise. Alors, au lieu de chercher la transformation Φ , il suffit d'introduire la fonction k

$$\begin{aligned} k: \mathbb{R}^d \times \mathbb{R}^d &\rightarrow \mathbb{R} \\ (x, x') &\mapsto k(x, x') = \langle \Phi(x), \Phi(x') \rangle_H \end{aligned}$$

appelée fonction noyau. Grâce au concept de noyau, il est possible de réécrire le problème dual de cette manière :

$$\begin{cases} \max -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i y_i \alpha_j y_j k(x^i, x^j) + \sum_{i=1}^n \alpha_i \\ \sum_{i=1}^n \alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq c, \quad i = 1, \dots, n \end{cases},$$

2. CAS DES DONNÉES NON LINÉAIREMENT SÉPARABLE

que l'on pourra aussi l'écrire sous la forme matricielle :

$$\begin{cases} \min \frac{1}{2} \alpha^T D K D \alpha - e^T \alpha \\ \alpha^T y = 0 \\ 0 \leq \alpha \leq c \end{cases},$$

où $\alpha = [\alpha_i]_{i=1}^n$, $y = [y^i]_{i=1}^n$, $e = (1, \dots, 1)^T \in \mathbb{R}^n$, $D = \text{diag}(y)$, $K = [k(x^i, x^j)]_{i,j=1}^n$ (appelée matrice de Gram).

Dans la pratique, le choix du noyau k joue un rôle important. Dans le chapitre suivant, on présente le cadre théorique de la fonction noyau.

Chapitre 2

Les noyaux

Il est à signaler que la théorie des fonctions noyaux existée avant l'apparition des SVMs. La maîtrise de cette théorie est fondamentale pour la bonne compréhension du rôle que joue cette classe de fonction dans la séparation non linéaire des données. Dans ce chapitre on va s'intéresser à quelques résultats dont on aura besoin pour la suite. Uniquement les noyaux à valeurs réels seront considérés, alors qu'en général il peuvent être à valeurs complexe. La plupart des résultats seront donnés sans démonstration, le lecteur intéressé peut consulter l'ouvrage [10].

Soit X un ensemble non vide, une fonction k est dit fonction noyau s'il existe un $\mathbb{R}$ -espace de Hilbert H et une transformation $\Phi : X \rightarrow H$ telle que :

$$\begin{aligned} k : X \times X &\rightarrow \mathbb{R} \\ (x, x') &\longmapsto k(x, x') = \langle \Phi(x'), \Phi(x) \rangle_H \end{aligned}$$

H s'appelle espace caractéristique.

Il n'y a pas de bijection entre les fonctions noyaux et les espaces caractéristiques, en effet, soit $X = \mathbb{R}$, et $k(x, x') = xx'$ pour tout $x, x' \in \mathbb{R}$, on définit la projection $\Phi : x \rightarrow (\frac{x}{\sqrt{2}}, \frac{x}{\sqrt{2}})$, alors,

$$\langle \Phi(x), \Phi(x') \rangle_{\mathbb{R}^2} = \frac{x}{\sqrt{2}} \frac{x'}{\sqrt{2}} + \frac{x}{\sqrt{2}} \frac{x'}{\sqrt{2}} = xx' = k(x, x'),$$

mais il est facile de voir que si on définit la transformation $\Phi : x \rightarrow (\frac{x}{\sqrt{3}}, \frac{x}{\sqrt{3}}, \frac{x}{\sqrt{3}})$, on obtient :

$$\langle \Phi(x), \Phi(x') \rangle_{\mathbb{R}^3} = \frac{x}{\sqrt{3}} \frac{x'}{\sqrt{3}} + \frac{x}{\sqrt{3}} \frac{x'}{\sqrt{3}} + \frac{x}{\sqrt{3}} \frac{x'}{\sqrt{3}} = xx' = k(x, x'),$$

par conséquent, une fonction noyau peut être associée à plusieurs espaces caractéristiques. Il est possible de construire des noyaux à partir des noyaux déjà existants, en utilisant des propriétés algébriques sur les espaces, les résultats sont donnés comme suit :

Lemme 2.1 (Somme des noyaux) [10] Soient k, k_1, k_2 des noyaux définis sur X , et $\alpha \geq 0$. Alors, αk et $k_1 + k_2$ sont aussi des noyaux définis sur X .

On considère maintenant le produit des noyaux.

Lemme 2.2 (Produit des noyaux) [10] Soient k_1, k_2 deux noyaux définis sur X_1, X_2 respectivement. Alors, $k_1 \cdot k_2$ est un noyau défini sur $X_1 \times X_2$, en particulier, si $X_1 = X_2$, alors $k(x, x') = k_1(x, x')k_2(x, x')$, pour tout $x, x' \in X_1$ est un noyau défini sur X_1 .

À partir des résultats précédent, on peut citer les exemple des noyaux les plus utilisés.

Lemme 2.3 (Noyau polynomial) [10] Soient $m \in \mathbb{N}$, $d \in \mathbb{N}^*$ et $c \geq 0$. Alors, la fonction k définis par

$$k(x, x') = (\langle x, x' \rangle_{\mathbb{R}^d} + c)^m \text{ pour tout } x, x' \in \mathbb{R}^d,$$

est un noyau sur $\mathbb{R}^d$ appelé noyau polynomial, en particulier, si $c = 0$ et $m = 1$, alors ce noyaux sera dit noyau linéaire.

On se basant sur la définition d'une fonction noyau et les propriétés de l'espace de Hilbert ℓ_2 , on peut montrer le résultat suivant :

Théorème 2.1 [10] soit X un ensemble non vide, et $f_n : X \rightarrow \mathbb{R}$, $n \in \mathbb{N}$, des fonctions telles que $(f_n(x)) \in \ell_2$, pour tout $x \in X$. alors,

$$k(x, x') = \sum_{i=0}^{\infty} f_i(x) f_i(x'), \quad x, x' \in X,$$

définit un noyau dans X .

On considère maintenant les fonctions développables en série de Taylor.

Théorème 2.2 [10] soient $\mathring{B}_r(\mathbb{R})$ la boule ouvert de $\mathbb{R}$ de centre r ($r \in [0, \infty]$), et $f : \mathring{B}_r(\mathbb{R}) \rightarrow \mathbb{R}$ une fonction développable en série de Taylor

$$f(x) = \sum_{n=0}^{\infty} a_n x^n, \quad x \in \mathring{B}_r(\mathbb{R}).$$

si $a_n \geq 0$ pour tout $n \in \mathbb{N}$, alors $k(x, x') = f(\langle x, x' \rangle_{\mathbb{R}^d}) = \sum_{n=0}^{\infty} a_n (\langle x, x' \rangle_{\mathbb{R}^d})^n$, $x, x' \in \mathring{B}_{\sqrt{r}}(\mathbb{R}^d)$ avec $d \in \mathbb{N}^*$ définit un noyau sur $\mathring{B}_{\sqrt{r}}(\mathbb{R}^d)$. Dans ce cas, on dit que k est un noyau de type Taylor.

À l'aide du lemme précédent, on peut déduire quelques exemples de noyaux, par exemple : pour $d \in \mathbb{N}^*$, pour tout $x, x' \in \mathbb{R}^d$, on a $k(x, x') := \exp(\langle x, x' \rangle_{\mathbb{R}^d})$ est un noyau sur $\mathbb{R}^d$ appelé noyau exponentiel. Le lemme suivant est aussi important pour construire des noyaux à partir d'autres.

Lemme 2.4 [10] Etant donnée un noyau k sur X , $\tilde{X}$ un ensemble, et une transformation $A : \tilde{X} \rightarrow X$. Alors, $\tilde{k} := k(A(x), A(x'))$, $x, x' \in \tilde{X}$, est un noyau sur $\tilde{X}$. En particulier, si $\tilde{X} \subset X$, alors $k|_{\tilde{X} \times \tilde{X}}$ est un noyau.

On peut introduire maintenant un des noyaux les plus utilisés en pratique, il est donné par la proposition suivante :

Proposition 2.1 (Noyau RBF) [10] Pour $d \in \mathbb{N}$, $\gamma > 0$, $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ et $x' = (x'_1, \dots, x'_d) \in \mathbb{R}^d$, on définit :

$$k_\gamma(x, x') = \exp(-\gamma^{-2} \sum_{i=1}^d (x_i - x'_i)^2).$$

Alors, k_γ est un noyau sur $\mathbb{R}^d$ appelé fonction à base radial (RBF) ou bien noyau Gaussien, il peut s'écrire aussi :

$$k_\gamma(x, x') = \exp(-\frac{\|x - x'\|_2^2}{\gamma^2}), \quad x, x' \in \mathbb{R}^d.$$

On présente à présent une caractérisation pratique des noyaux, d'abord, on a besoin de la définition suivante :

Définition 2.1 [10] Une fonction $k : X \times X \rightarrow \mathbb{R}$ est dite semi-défini positive si, pour tout $n \in \mathbb{N}$, pour tous $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ et pour tous $x_1, \dots, x_n \in X$, on a

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j k(x_i, x_j) \geq 0,$$

k est dit défini positive si, pour tout $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$, tel que $\alpha \neq 0$, on a

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j k(x_i, x_j) > 0, \quad x_1, \dots, x_n \in X,$$

finally, k est dit symétrique si, $k(x, x') = k(x', x)$, pour tout $x, x' \in X$.

Si on définit la matrice de Gram K d'un noyau k par rapport aux éléments $x_1, \dots, x_n \in X$ comme $[K_{ij}]_{i,j=1,\dots,n} = [k(x_i, x_j)]_{i,j=1,\dots,n}$, alors la définition précédente est équivalente à la semi-définie positivité (au sens matriciel) de K pour tout $x_1, \dots, x_n \in X$. Un dernier résultat qui nous permet de caractériser les fonctions noyaux sans passer explicitement par l'espace caractéristiques est donné par le théorème suivant :

Théorème 2.3 (Théorème de Mercer) [10] Une fonction $k : X \times X \rightarrow \mathbb{R}$ est un noyau si et seulement si elle est symétrique et semi-définie positive.

Dans le chapitre suivant nous verrons une classe de noyaux basée sur la notion de polynômes orthogonaux.

Chapitre 3

Noyaux à base de polynômes orthogonaux

Les polynômes orthogonaux jouent un rôle très important dans la théorie des espaces de Hilbert. Dans ce chapitre, on va introduire certains polynômes orthogonaux s'obtenant à partir d'un problème de Sturm-Liouville [3]. En particulier, notre attention sera centrée sur les polynômes d'Hermite, Laguerre et Jacobi. Ensuite, on fait une description de la construction d'une fonction noyau à l'aide de ces classes de polynômes. On commence d'abord par donner une première définition très générale d'une suite des polynômes orthogonaux et les propriétés qui la caractérisent.

1 généralités

soit $-\infty \leq a \leq b \leq +\infty$, on introduit une fonction poids $w :]a, b[\rightarrow \mathbb{R}^+$, avec $w \in L^1(]a, b[)$, on définit l'espace de Hilbert

$$L_w^2(]a, b[) = \left\{ f :]a, b[\rightarrow \mathbb{R} \mid \int_a^b f^2(x) w(x) dx < +\infty \right\}$$

muni d'un produit scalaire

$$\langle f, g \rangle_w = \int_a^b f(x) g(x) w(x) dx, f, g \in L_w^2(]a, b[),$$

soit la famille des polynômes $(P_i)_{i \geq 0}$ telle que chaque P_i , $i \in \mathbb{N}$ est un polynôme de degré i : On dit que $(P_i)_{i \geq 0}$ est une famille orthogonale sur $]a, b[$ par rapport au poids w si $\langle P_i, P_j \rangle_w = 0$, $\forall i, j \in \mathbb{N}$, $i \neq j$. Il est intéressant de savoir que les polynômes orthogonaux peuvent s'obtenir à l'aide d'une relation de récurrence, particulièrement lorsqu'on veut faire du calcul numérique.

2 Exemples classiques des polynômes orthogonaux

Dans cette section, on mentionne quelques exemples classiques et leurs caractéristiques principales [9],[2].

2.1 Polynômes d'Hermite

on considère la fonction poids définie par

$$w(x) = \exp\left(-\frac{x^2}{2}\right), \quad x \in \mathbb{R},$$

w est de classe $C^\infty(\mathbb{R})$ et on a

$$w'(x) = -x \exp\left(-\frac{x^2}{2}\right), \quad w''(x) = (x^2 - 1) \exp\left(-\frac{x^2}{2}\right), \quad w^{(3)}(x) = (-x^3 + 3x) \exp\left(-\frac{x^2}{2}\right), \dots$$

il est clair que la dérivée de w de n'importe quelle ordre est le produit entre $\exp(-\frac{x^2}{2})$ et un polynôme en x , pour tout $n \in \mathbb{N}$, on définit

$$H_n(x) = (-1)^n \exp\left(\frac{x^2}{2}\right) \frac{d^n}{dx^n} w(x), \quad (3.1)$$

alors,

$$\frac{d^n}{dx^n} w(x) = (-1)^n \exp\left(-\frac{x^2}{2}\right) H_n(x), \quad (3.2)$$

et donc,

$$\frac{d^{n+1}}{dx^{n+1}} w(x) = (-1)^{n+1} \exp\left(-\frac{x^2}{2}\right) [-xH_n(x) + H'_n(x)], \quad (3.3)$$

d'autre part,

$$\frac{d^{n+1}}{dx^{n+1}} w(x) = (-1)^{n+1} \exp\left(-\frac{x^2}{2}\right) (-H_{n+1}(x)), \quad (3.4)$$

par identification entre (3.3) et (3.4), on obtient la relation de récurrence

$$\begin{cases} H_0(x) = 1, \\ H_{n+1}(x) = xH_n(x) - H'_n(x), \end{cases} \quad (3.5)$$

cette dernière formule assure bien que $H_n(x)$ est un polynôme de degré n , les $H_n(x)$, $\forall n \in \mathbb{N}$, sont appelés les polynômes d'Hermite, et la représentation (3.1) est appelée formule de Rodrigues des polynômes d'Hermite.

Orthogonalité

Les polynômes d'Hermite sont orthogonaux par rapport à la fonction poids w sur $\mathbb{R}$, en effet, soient $n, m \in \mathbb{N}$, telles que $n \geq m$

$$\langle H_n, H_m \rangle_w = \int_{-\infty}^{+\infty} H_n(x) H_m(x) w(x) dx = (-1)^n \int_{-\infty}^{+\infty} H_m(x) w^{(n)}(x) dx,$$

2. EXEMPLES CLASSIQUES DES POLYNÔMES ORTHOGONAUX

par simple intégration par partie,

$$\begin{aligned} u(x) &= H_m(x) & u'(x) &= H'_m(x) \\ v'(x) &= w^{(n)}(x) & v(x) &= w^{(n-1)}(x) \end{aligned} ,$$

on a,

$$H_m(x) w^{(n-1)}(x) \Big|_{-\infty}^{+\infty} = (-1)^{n-1} H_m(x) H_{n-1}(x) \exp\left(-\frac{x^2}{2}\right) \Big|_{-\infty}^{+\infty} = 0,$$

donc

$$\langle H_n, H_m \rangle_w = (-1)^{n+1} \int_{-\infty}^{+\infty} H'_m(x) w^{(n-1)}(x) dx,$$

en répétant le processus $(m-1)$ -fois, on obtient

$$\langle H_n, H_m \rangle_w = (-1)^{n+m} \int_{-\infty}^{+\infty} H_m^{(m)}(x) w^{(n-m)}(x) dx,$$

si $n > m$: par une autre intégration par partie, $H_m^{(m+1)}(x) = 0$ car $H_m^{(m)}(x) = m!$ et donc, $\langle H_n, H_m \rangle_w = 0$, d'où, l'orthogonalité de $H_m(x)$ et $H_n(x)$ par rapport à $\exp(-\frac{x^2}{2})$ sur $\mathbb{R}$, si $n \neq m$: on a

$$\|H_m\|_w^2 = \int_{-\infty}^{+\infty} [H_m(x)]^2 w(x) dx = \int_{-\infty}^{+\infty} H_m^{(m)}(x) w(x) dx = m! \int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2}\right) dx = m! \sqrt{2\pi},$$

finalemt,

$$\langle H_n, H_m \rangle_w = m! \sqrt{2\pi} \delta_{nm},$$

avec δ est le symbole de Kronecker.

Relation de récurrence

la fonction w satisfait l'équation

$$w'(x) + xw(x) = 0,$$

par dérivation successive, n -fois, les deux membres de l'égalité

$$w^{(n+1)}(x) + xw^{(n)}(x) + nw^{(n-1)}(x) = 0,$$

en utilisant la formule (3.2), on obtient

$$[-H_{n+1}(x) + xH_n(x) - nH_{n-1}(x)](-1)^n \exp\left(-\frac{x^2}{2}\right) = 0,$$

d'où, chaque trois polynômes d'Hermite d'ordre successive vérifie la relation de récurrence suivante

$$\begin{cases} H_0(x) = 1, \\ H_1(x) = x, \\ H_{n+1}(x) = xH_n(x) - nH_{n-1}(x). \end{cases} \quad (3.6)$$

L'équation différentielle vérifiée

Par comparaison entre les deux formules (3.5) et (3.6), on a

$$H'_n(x) = nH_{n-1}(x),$$

d'une part

$$H'_{n+1}(x) = (n+1)H_n(x),$$

et d'autre part, en dérivant les deux membre de l'égalité (3.5)

$$H'_{n+1}(x) = -H''_n(x) + xH'_n(x) + H_n(x),$$

d'où, les polynômes d'Hermite sont solution de l'équation différentielle suivante

$$-H''_n(x) + xH'_n(x) - nH_n(x) = 0,$$

qui peut être mise sous forme d'un problème de Sturm-Liouville

$$\frac{d}{dx} \left[e^{-\frac{x^2}{2}} \frac{d}{dx} H_n(x) \right] + n e^{-\frac{x^2}{2}} H_n(x) = 0.$$

Les premeir termes de polynômes d'Hermite

$$\begin{aligned} H_0(x) &= 1, \\ H_1(x) &= x, \\ H_2(x) &= x^2 - 1, \\ H_3(x) &= x^3 - 3x, \\ H_4(x) &= x^4 - 6x^2 + 3, \\ H_5(x) &= x^5 - 10x^3 + 15x. \end{aligned}$$

2.2 Polynômes de Laguerre

on considère la fonction poids

$$w(x) = x^\alpha e^{-x}, \quad x \in [0, +\infty[\text{ et } \alpha \in \mathbb{R},$$

w est intégrable sur $[0, +\infty[$ pour $\alpha > -1$ car; au voisinage de 0 : $x^\alpha e^{-x}$ est équivalente à x^α qui est intégrable pour $\alpha > -1$, et au voisinage de $+\infty$: $x^\alpha e^{-x}$ est équivalente à e^{-x} qui est intégrable pour tout α , on définit la fonction Φ , pour tout $x \in [0, +\infty[$ et $\alpha > -1$, par

$$\Phi_n(x) = x^{\alpha+n} e^{-x},$$

2. EXEMPLES CLASSIQUES DES POLYNÔMES ORTHOGONAUX

comme $\Phi_n \in C^\infty([0, +\infty[)$, alors, on applique la formule de Leibniz pour calculer la dérivée n -ième de Φ

$$\begin{aligned}\Phi_n^{(n)}(x) &= \sum_{i=0}^n C_n^i (x^{\alpha+n})^{(i)} (e^{-x})^{(n-i)}, \text{ avec } C_n^i = \frac{n!}{i!(n-i)!}, \\ \Phi_n^{(n)}(x) &= [(-1)^n x^n + \dots + (\alpha+n) \dots (\alpha+1)] x^\alpha e^{-x},\end{aligned}$$

de cette dernière formule, on remarque que $\Phi^{(n)}$ est le produit de w et un polynôme de degré n , on peut définir maintenant les polynômes de Laguerre par la formule de Rodrigues

$$L_n^\alpha(x) = (-1)^n x^{-\alpha} e^x \frac{d^n}{dx^n} \Phi_n(x),$$

alors,

$$\frac{d^n}{dx^n} \Phi_n(x) = (-1)^n x^\alpha e^{-x} L_n^\alpha(x),$$

qui nous donne directement la relation suivante

$$L_{n+1}^\alpha(x) = (x - \alpha - n - 1) L_n^\alpha(x) - x L_n^{\alpha'}(x),$$

L'orthogonalité

soient $\alpha > -1$, $n, m \in \mathbb{N}$, $n \geq m$

$$\langle L_n^\alpha, L_m^\alpha \rangle_w = \int_0^\infty L_n^\alpha(x) L_m^\alpha(x) w(x) dx = (-1)^n \int_0^\infty L_m^\alpha(x) \Phi_n^{(n)}(x) dx$$

après m -intégration par partie successive, on va obtenir

$$\int_0^\infty L_n^\alpha(x) L_m^\alpha(x) w(x) dx = (-1)^{n+m} \int_0^\infty (L_m^\alpha)^{(m)}(x) \Phi_n(x) dx$$

si $n > m$: par une autre intégration par partie et comme $(L_m^\alpha)^{(m)}(x) = m!$, on a

$$\int_0^\infty L_n^\alpha(x) L_m^\alpha(x) w(x) dx = 0,$$

d'où l'orthogonalité des polynômes de Laguerre par rapport à la fonction poids w sur $]0, +\infty[$ pour tout $\alpha > -1$, si $n \neq m$:

$$\|L_n^\alpha\|_w^2 = \int_0^\infty [L_n^\alpha(x)]^2 w(x) dx = n! \int_0^\infty x^{\alpha+n} e^{-x} dx = n! \Gamma(n + \alpha + 1),$$

Γ représente la fonction gamma d'Euler. Finalement

$$\langle L_n^\alpha, L_m^\alpha \rangle_w = n! \Gamma(n + \alpha + 1) \delta_{nm}.$$

Équation différentielle et la formule de récurrence

Pour tout $x \in]0, +\infty[$ et $\alpha > -1$, $n \in \mathbb{N}$, on a

$$\frac{w'(x)}{w(x)} = \frac{\alpha - x}{x}$$

alors les polynômes de Laguerre satisfaites l'équation différentielle suivante [2]

$$xL_n^{\alpha}{}''(x) + (\alpha + 1 - x)L_n^{\alpha'}(x) + nL_n^{\alpha}(x) = 0,$$

qu'elle est équivalente à

$$\frac{d}{dx}[x^{\alpha+1}e^{-x}\frac{d}{dx}L_n(x)] + x^{\alpha+1}e^{-x}nL_n^{\alpha}(x) = 0,$$

ils vérifient aussi la relation de récurrence suivante [2]

$$L_{n+1}^{\alpha}(x) - (x - \alpha - 2n - 1)L_n^{\alpha}(x) + n(\alpha + n)L_{n-1}^{\alpha}(x) = 0.$$

les premiers termes de polynômes de Laguerre

$$\begin{aligned} L_0^{\alpha}(x) &= 1, \\ L_1^{\alpha}(x) &= x - \alpha - 1, \\ L_2^{\alpha}(x) &= x^2 - 2(\alpha + 2)x + (\alpha + 2)(\alpha + 1). \end{aligned}$$

Il est clair que quand a et b sont finis, on peut toujours ramener l'intervalle $[a, b]$ à $[-1, 1]$ par changement de variable affine

$$x \rightarrow \frac{2(x-a)}{b-a} - 1, x \in [a, b], \quad (3.7)$$

et c'est le cas de l'exemple suivant.

2.3 Polynômes de Jacobi

On définit la fonction poids $w(x) = (1-x)^{\alpha}(1+x)^{\beta}$, $x \in]-1, 1[$, qui est intégrable pour $\alpha, \beta > -1$, on considère la fonction

$$\Phi_n(x) = (1-x)^{\alpha+n}(1+x)^{\beta+n},$$

on définit les polynômes de Jacobi par la formule de Rodrigues

$$P_n^{(\alpha, \beta)}(x) = \frac{(-1)^n}{2^n n!} (1-x)^{-\alpha} (1+x)^{-\beta} \frac{d^n}{dx^n} ((1-x)^{\alpha+n} (1+x)^{\beta+n}),$$

un raisonnement similaire à celui fait à la section précédente nous permet de conclure les résultats suivants :

L'orthogonalité

Les polynômes de Jacobi sont orthogonaux par rapport à w sur $] -1, 1[$, c.-à-d.

$$\int_{-1}^{+1} P_n^{(\alpha, \beta)}(x) P_m^{(\alpha, \beta)}(x) w(x) dx = 0, \text{ pour } n \neq m,$$

et

$$\|P_n^{(\alpha, \beta)}\|_w^2 = \frac{2^{\alpha+\beta+1} \Gamma(n+\alpha+1) \Gamma(n+\beta+1)}{(2n+\alpha+\beta+1) n! \Gamma(n+\alpha+\beta+1)}.$$

Relation de récurrence

Une propriété importante des polynômes de Jacobi est que chaque trois termes vérifient la relation de récurrence suivante :

$$\begin{cases} P_0^{(\alpha, \beta)} = 1, \\ P_1^{(\alpha, \beta)} = \frac{1}{2} [2(\alpha+1) + (\alpha+\beta+2)(x-1)], \\ P_{n+1}^{(\alpha, \beta)}(x) = (a_n x + b_n) P_n^{(\alpha, \beta)}(x) \\ \quad - c_n P_{n-1}^{(\alpha, \beta)}(x), n \geq 1, \end{cases}$$

avec

$$\begin{cases} a_n = \frac{(2n+\alpha+\beta+1)(2n+\alpha+\beta+2)}{2(n+1)(n+\alpha+\beta+1)}, \\ b_n = \frac{(2n+\alpha+\beta+1)(\alpha^2-\beta^2)}{2(n+1)(n+\alpha+\beta+1)(2n+\alpha+\beta)}, \\ c_n = \frac{(\alpha+n)(\beta+n)(2n+\alpha+\beta+2)}{(n+1)(n+\alpha+\beta+1)(2n+\alpha+\beta)}. \end{cases}$$

Équation différentielle

Il est possible de définir ces familles des polynômes par une équation différentielle

$$(1-x^2)P_n^{(\alpha, \beta)''}(x) + (\beta-\alpha-(\alpha+\beta+2)x)P_n^{(\alpha, \beta)'}(x) + n(n+\alpha+\beta+1)P_n^{(\alpha, \beta)}(x) = 0,$$

qui est équivalent au problème de Sturm-Liouville

$$-(1-x)^{-\alpha}(1+x)^{-\beta} \frac{d}{dx} [(1-x)^{\alpha+1}(1+x)^{\beta+1} \frac{d}{dx} P_n^{(\alpha, \beta)}(x)] - n(n+\alpha+\beta+1)P_n^{(\alpha, \beta)}(x) = 0.$$

Les premiers de ces polynômes

$$\begin{aligned} P_0^{(\alpha, \beta)}(x) &= 1, \\ P_1^{(\alpha, \beta)}(x) &= \frac{1}{2} [2(\alpha+1) + (\alpha+\beta+2)(x-1)], \\ P_2^{(\alpha, \beta)}(x) &= \frac{1}{8} [4(\alpha+1)(\alpha+2) + 4(\alpha+\beta+3)(\alpha+2)(x-1) + (\alpha+\beta+3)(\alpha+\beta+4)(x-1)^2]. \end{aligned}$$

3. LA CONSTRUCTION D'UNE FONCTION NOYAU PAR LES POLYNÔMES ORTHOGONAUX

Remarque 3.1 Les sélections spéciales de α et β portent des noms spéciaux :

- $\alpha = \beta = 0$: polynômes de Legendre,
- $\alpha = \beta = -\frac{1}{2}$: polynômes de chebychev (de premier espèce),
- $\alpha = \beta = +\frac{1}{2}$: polynômes de chebychev (de seconde espèce),
- $\alpha = \beta$: polynômes ultrasphériques.

Ces trois familles des polynômes orthogonaux possèdent une propriétés très intéressante [3] : chacun forme une base orthogonale de l'espace $L_w^2(a, b)$

3 La construction d'une fonction noyau par les polynômes orthogonaux

Dans cette partie, on construit une fonction noyau à base des polynômes d'Hermite, Laguerre et Jacobi [12].

3.1 Noyau des polynômes orthogonaux

Soient $\{\Phi_n\}_{n \in \mathbb{N}}$ ensemble des polynômes orthogonaux définis sur $L_w^2(a, b)$, donc tout fonction $f \in L_w^2(a, b)$ peut s'écrire comme suit

$$f(x) = \sum_{n=0}^{\infty} c_n \Phi_n(x), \quad x \in]a, b[, \quad \text{où } c_n = \frac{\langle f, \Phi_n \rangle_w}{\|\Phi_n\|_w^2}, \quad (3.8)$$

$$\begin{aligned} f(x) &= \sum_{n=0}^{\text{inf ty}} \frac{\langle f, \Phi_n \rangle_w}{\|\Phi_n\|_w^2} \Phi_n(x) \\ &= \sum_{n=0}^{\text{inf ty}} \int_a^b f(t) \Phi_n(t) w(t) \frac{\Phi_n(x)}{\|\Phi_n\|_w^2} dt \\ &= \int_a^b \sum_{n=0}^{\text{inf ty}} \frac{\Phi_n(t) \Phi_n(x)}{\|\Phi_n\|_w^2} f(t) w(t) dt \end{aligned}$$

on peut exprimer la fonction noyau comme suit

$$K(x, t) = \sum_{i=0}^N \frac{\Phi_i(t) \Phi_i(x)}{\|\Phi_i\|_w^2}, \quad x, t \in]a, b[\quad (3.9)$$

qu'on peut le voir comme produit scalaire entre deux vecteurs de $\mathbb{R}^{N+1}$, c.-à-d.

$$K(x, t) = \langle \Phi(x), \Phi(t) \rangle_{\mathbb{R}^{N+1}}$$

3. LA CONSTRUCTION D'UNE FONCTION NOYAU PAR LES POLYNÔMES ORTHOGONAUX

telle que Φ est la transformation défini par $\Phi :]a, b[\rightarrow \mathbb{R}^{N+1}$, $\Phi(x) = (\frac{\Phi_0(x)}{\|\Phi_0\|_w}, \frac{\Phi_1(x)}{\|\Phi_1\|_w}, \dots, \frac{\Phi_N(x)}{\|\Phi_N\|_w})$.

Il reste à montrer que K est effectivement une fonction noyau, on a

$$\begin{aligned} \int_a^b \int_a^b K(x, t) f(x) f(t) w(x) dx w(t) dt &= \int_a^b \int_a^b \sum_{n=0}^N \frac{\Phi_n(t) \Phi_n(x)}{\|\Phi_n\|_w^2} f(x) w(x) f(t) w(t) dx dt \\ &= \sum_{n=0}^N \left[\int_a^b \frac{\Phi_n(x) f(x)}{\|\Phi_n\|_w} w(x) dx \int_a^b \frac{\Phi_n(t) f(t)}{\|\Phi_n\|_w} w(t) dt \right] \\ &= \sum_{n=0}^N \left[\int_a^b \frac{\Phi_n(x) f(x)}{\|\Phi_n\|_w} w(x) dx \right]^2 \geq 0 \end{aligned}$$

donc K vérifie bien les conditions de Mercer, d'où K est une fonction noyau sur $]a, b[$.
À l'aide de la formule (3.9) et Pour N = 2, on peut citer les noyaux à base des polynômes orthogonaux.

Noyau d'Hermite

$$\begin{aligned} K(x, y) &= \sum_{i=0}^2 \frac{H_i(x) H_i(y)}{\|H_i\|_w^2} \\ &= \frac{1}{\sqrt{2\pi}} [1 + xy + \frac{1}{2}(x^2 - 1)(y^2 - 1)]. \end{aligned}$$

Noyau de Laguerre

$$\begin{aligned} K(x, y) &= \sum_{i=0}^2 \frac{L_i^\alpha(x) L_i^\alpha(y)}{\|L_i^\alpha\|_w^2} \\ &= \frac{1}{\Gamma(\alpha+1)} \left[1 + \frac{(-x+\alpha+1)(-y+\alpha+1)}{(\alpha+1)} \right. \\ &\quad \left. + \frac{[x^2 - 2(\alpha+2)x + (\alpha+2)(\alpha+1)][y^2 - 2(\alpha+2)y + (\alpha+2)(\alpha+1)]}{8(\alpha+1)(\alpha+2)} \right]. \end{aligned}$$

Noyau de Jacobi

$$\begin{aligned}
 K(x, y) &= \sum_{i=0}^2 \frac{P_i^{(\alpha, \beta)}(x) P_i^{(\alpha, \beta)}(y)}{\|P_i^{(\alpha, \beta)}\|_w^2} \\
 &= \frac{2^{\alpha+\beta+1} \Gamma(\alpha+1) \Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)} \left[1 + \frac{(\alpha+1)(\beta+1)}{(\alpha+\beta+3)} P_1^{(\alpha, \beta)}(x) P_1^{(\alpha, \beta)}(y) \right. \\
 &\quad \left. + \frac{(\alpha+1)(\alpha+2)(\beta+1)(\beta+2)}{2(5+\alpha+\beta)(\alpha+\beta+2)} P_2^{(\alpha, \beta)}(x) P_2^{(\alpha, \beta)}(y) \right].
 \end{aligned}$$

3.2 Noyau à base de polynômes orthogonaux pour les SVMs

L'intérêt pour nous est d'avoir un noyau sur $\mathbb{R}^d$, pour se faire, on considère $x = (x_1, \dots, x_d)^T \in \mathbb{R}^d$ et $z = (z_1, \dots, z_d)^T \in \mathbb{R}^d$, on se basant sur le 2^{ième} chapitre, en particulier, le lemme(2.2), on peut définir la fonction noyau k par

$$k(x, z) = \prod_{j=1}^d K(x_j, z_j) = \prod_{j=1}^d \sum_{n=0}^N \frac{\Phi_n(x_j) \Phi_n(z_j)}{\|\Phi_n\|_w^2}.$$

avec $\{\Phi_n\}_{n \in \mathbb{N}}$ peut être une famille des polynôme d'Hermite, Laguerre ou Jacobi. Dans le chapitre qui suit, une étude numérique sera réalisée afin de faire une comparaison entre les divers noyaux proposés ici dans le cadre des SVMs.

Chapitre 4

Tests numériques

Le choix du noyau dans le cadre des problèmes de classification non linéaire, fait toujours l'objet de recherches récentes [8], [7], [5], pas uniquement en mathématiques, mais aussi dans d'autres disciplines (Informatique, Robotique ext...). En pratique, le noyau sélectionné est celui qui donne de bonnes performances numériques ainsi qu'une bonne généralisation. Trouver un noyau qui fonctionne bien avec plusieurs types de données est un problème très difficile. Dans notre travail, une comparaison sera réalisée sur un exemple académique de données non linéairement séparables entre les noyaux basés sur les polynômes orthogonaux d'Hermite, Laguerre et Jacobi. L'exemple considéré est nommé *twospirals Data* [4]; il s'agit d'un ensemble de points dans $\mathbb{R}^2$ étiquetés en deux classes et ayant une forme spirale de telle sorte qu'il est quasiment impossible de les séparer linéairement (Par un hyperplan).

1 Méthode d'évaluation

L'exemple *twospirals Data* contient 200 points dans $\mathbb{R}^2$, 100 pour chaque classe (voir la figure 4.1). Les données seront d'abord normalisées en fonction des domaines de définition de chaque type de polynôme, par la formule (3.7). Un ensemble d'apprentissage qui contient 90% des données totales sera sélectionné aléatoirement, et séparé des 10% données restantes, qui seront considérées comme ensemble test. Plusieurs SVMs ont été réalisés sur l'ensemble d'apprentissage afin de sélectionner les paramètres optimaux en utilisant la méthode statistique 10 folds validation croisée (En anglais 10 folds cross validation) [1], en se basant sur le taux d'exactitude (*accuracy*) qui est donnée par la formule suivante :

$$accuracy = \frac{|\{prédictions\ correctes\}|}{|\{tests\}|} \times 100.$$

Le paramètre c prend les valeurs 2^i lorsque $i = -5, \dots, 0, \dots, 15$. Le paramètre degré N pour les divers polynômes varie de 1, ..., 13, pour le polynôme de Laguerre, le paramètre α varie de 0.5, ..., 2, et pour Jacobi, les deux paramètres α et β varient de 0.5, ..., 2. Les programmes utilisés sont implémentés avec le logiciel Matlab, version 2014, sur un PC

1. MÉTHODE D'ÉVALUATION

de 2.53 GHZ avec 4 GO de RAM. Il est important de noter qu'on a aussi considéré le nombre de vecteurs de support (vs) pour évaluer la fonction noyau utilisée. La dispersion des données *twospirals* est présentée par la figure 4.1 :

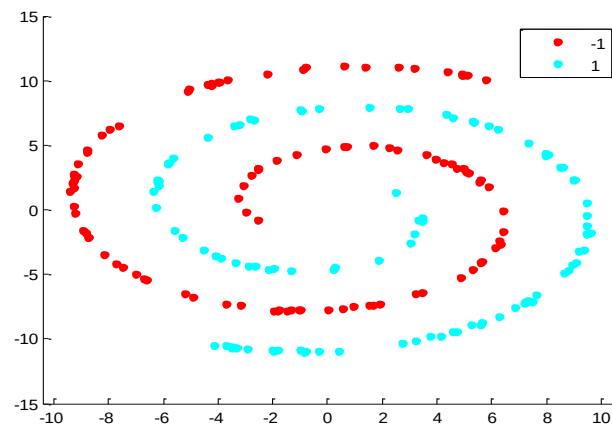


FIGURE 4.1 – Les données *twospirals*.

La méthode d'évaluation peut être schématisé par l'organigramme dans la figure 4.2

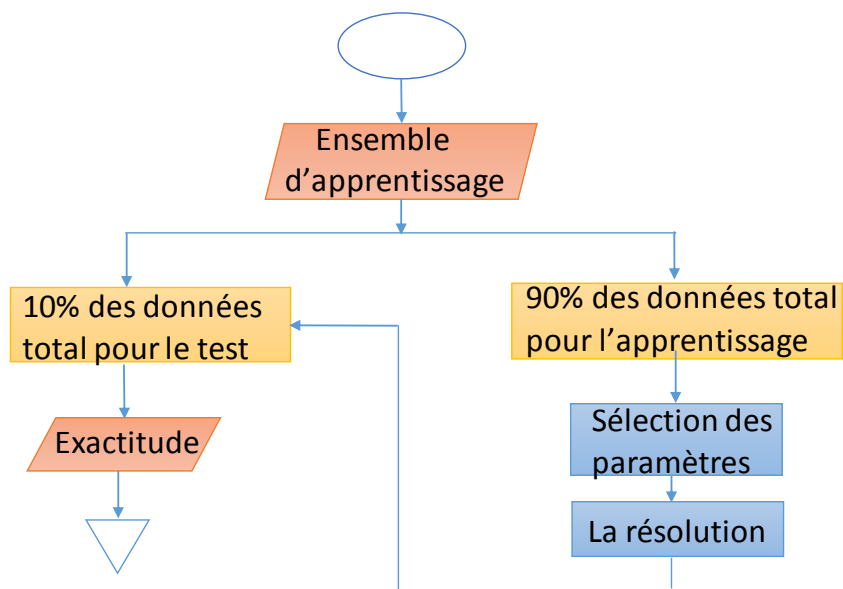


FIGURE 4.2 – Organigramme de la méthode d'évaluation

La section suivante résume les résultats obtenus.

2 Résultats

Les résultats sont donnés par le tableau 4.3

Noyau	Meilleur c	Meilleur N	Meilleur α	Meilleur β	Nombre de VS	Accuracy	Temps d'exécution
Noyau d'Hermite	8192	15	-	-	26	100%	134,4313 s
Noyau de Laguerre	0,0313	5	1,5	-	29	100%	115,0355 s
Noyau de Jacobi	32	8	-0,75	-0,75	37	100%	289,2329 s

FIGURE 4.3 – Les résultats obtenus.

Pour une meilleure lisibilité de nos résultats, on a choisi leur représentation graphique (les figures 4.4, 4.5, 4.6).

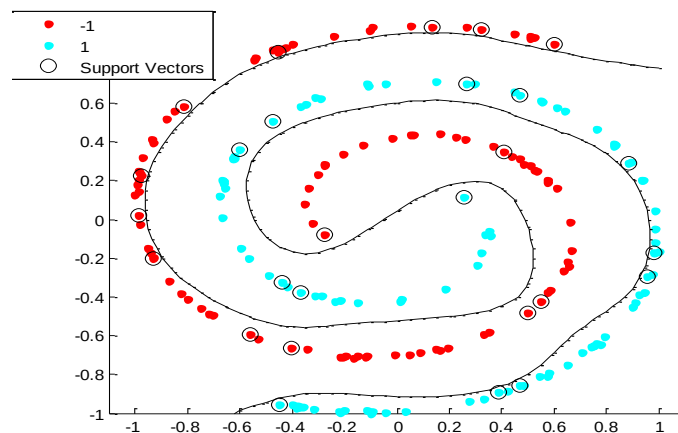


FIGURE 4.4 – La courbe de décision du noyau d'Hermite dans l'espace des entrées

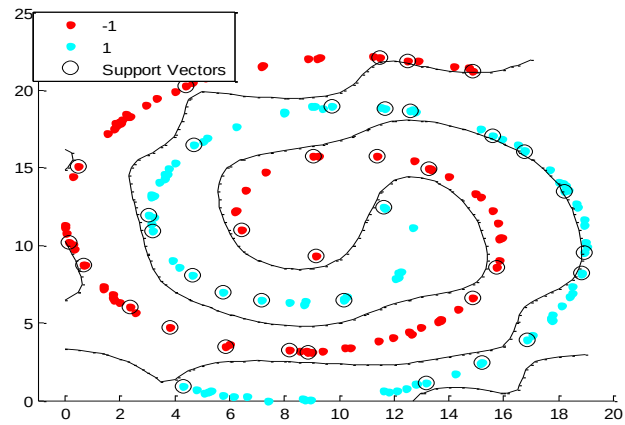


FIGURE 4.5 – La courbe de décision du noyau de Laguerre dans l'espace des entrées

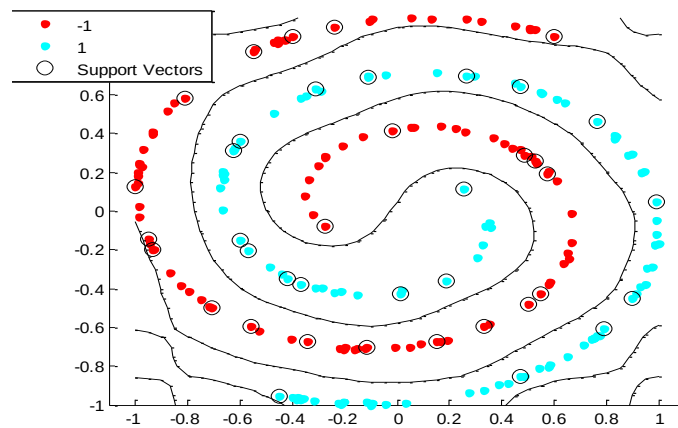


FIGURE 4.6 – La courbe de décision du noyau de Jacobi dans l'espace des entrées

Conclusion et perspectives

Dans ce mémoire, on a donné une vision purement mathématique des SVMs, cette méthode de classification est basée sur la recherche d'un hyperplan qui permet de séparer au mieux les ensembles des données. On a exposé le cas linéairement séparable et non-linéairement séparables qui nécessite l'utilisation des fonctions noyaux. Parmi les fonction noyaux, nous sommes particulièrement intéressés par les noyaux à base de polynômes orthogonaux de Jacobi, Hermite et Laguerre. Ce multiple choix de noyaux, rend les SVMs plus intéressants et surtout plus riches, puisqu'on peut toujours chercher de nouveaux noyaux, qui peuvent être mieux adaptés à la tâche qu'on veut accomplir.

Ce modeste travail, nous a permis d'envisager des perspectives dans les deux volés, analyse et méthode. D'une part, la possibilité de dégager les propriétés mathématiques que les noyaux à base de polynômes orthogonaux possèdent par rapport aux noyaux classiques tel que RBF et Polynômial. La possibilité de réduire le temps dans le calcul de la matrice de Gram, rendant ainsi ces noyaux plus compétitifs.

Bibliographie

- [1] **P. A. Devijver, J. Kittler**, A Statistical Approach, Pattern Recognition, 1982. [23](#)
- [2] **D. JACKSON**, Fourier Series and Orthogonal Polynomials, The Mathematical Association of America, 170-190, 1948. [3](#), [14](#), [18](#)
- [3] **D. A. Kopriva**, Implementing Spectral Methods for Partial Differential Equations. Algorithms for Scientists and Engineers, 23-26, 2009. [2](#), [13](#), [20](#)
- [4] **K. J. Lang and M. J. Witbrock**,). Learning to tell two spirals apart, Proceedings of the Connectionist Models Summer School, 52-59, 1988. [3](#), [23](#)
- [5] **V. H. Moghaddam and J. Hamidzadeh**, New Hermite orthogonal polynomial kernel and combined kernels in Support Vector Machine classifier, Pattern Recognition, 921-935, 1982. 2016. [3](#), [23](#)
- [6] **J. Nocedal and S. J. Wright**, Numerical Optimization, Springer Science and Business Media, 2006. [6](#)
- [7] **S. Ozer and C. Chen and H. A. Cirpan**, A set of new Chebyshev kernel functions for support vector machine pattern classification, Pattern Recognition, 1435-1447, 2011. [3](#), [23](#)
- [8] **Z. B. PAN and H. CHEN and X. YOU**, Support Vector Machine With Orthogonal Legendre Kernel, Proceedings of the 2012 International Conference on Wavelet Analysis and Pattern Recognition, 125-129, 2012. [3](#), [23](#)
- [9] **G. SZEGÖ**, Orthogonal Polynomials, American Mathematical Society, 25-30, 1939. [3](#), [14](#)
- [10] **I. Steinwart and A. Christmann**, Support Vector Machines, Information Science and Statistics, 111-118, 2008. [2](#), [10](#), [11](#), [12](#)
- [11] **V. Vapnik**, Statistical learning theory, New York : WileyInterscience, 1998. [2](#), [6](#)
- [12] **F. Zhou and Z. Fang and J. Xu**, Constructing Support Vector Machine Kernels from Orthogonal Polynomials for Face and Speaker Verification, Fourth International Conference on Image and Graphics, 327-631, 2007. [3](#), [20](#)