

SVM fonctionnels par interpolation spline

Nathalie Villa-Vialaneix Fabrice Rossi

Université Toulouse Le Mirail, Équipe GRIMM
villa@univ-tlse2.fr, <http://www.nathalievilla.org>

INRIA, Rocquencourt

Clamart, 2 juin 2006



SVM spline,
2 juin 06

Nathalie VILLA,
Fabrice ROSSI

SVM pour
données
fonctionnelles

Approche par
 L -Splines

Bibliographie

1 SVM pour données fonctionnelles

2 Approche par L -Splines

SVM spline,
2 juin 06

Nathalie VILLA,
Fabrice ROSSI

SVM pour
données
fonctionnelles

Approche par
L-Splines

Bibliographie

1 SVM pour données fonctionnelles

2 Approche par *L*-Splines

SVM spline,
2 juin 06

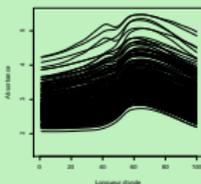
Nathalie VILLA,
Fabrice ROSSI

SVM pour
données
fonctionnelles

Approche par
L-Splines

Bibliographie

Le problème : Une discrimination binaire de courbes

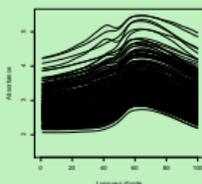


$X \in \mathcal{H}$

$\rightarrow \{-1; +1\}$

(espace de Hilbert de dim infinie)

Le problème : Une discrimination binaire de courbes



$X \in \mathcal{H}$

$\rightarrow \{-1; +1\}$

(espace de Hilbert de dim infinie)

Les données

On dispose de n réalisations indépendantes de (X, Y) :
 $(x_1, y_1), \dots, (x_n, y_n)$ ou x_i est connue aux points de
 discrétisation t_1, \dots, t_d .

SVM spline,
2 juin 06

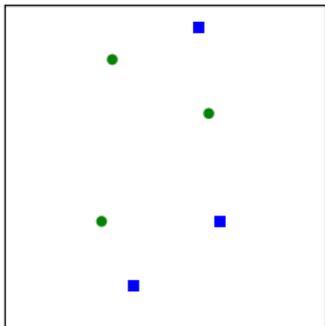
Nathalie VILLA,
Fabrice ROSSI

SVM pour
données
fonctionnelles

Approche par
L-Splines

Bibliographie

Espace initial \mathcal{H}



SVM spline,
2 juin 06

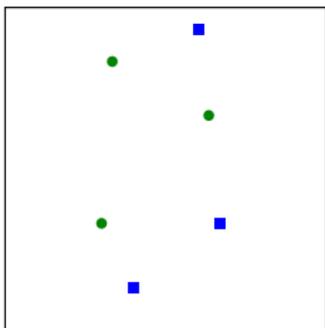
Nathalie VILLA,
Fabrice ROSSI

SVM pour
données
fonctionnelles

Approche par
L-Splines

Bibliographie

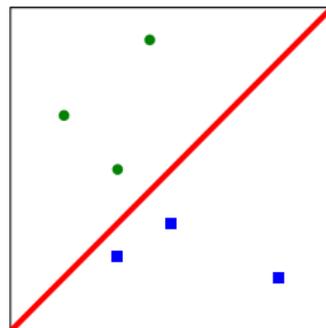
Espace initial \mathcal{H}



Φ (non linéaire)



Espace image \mathcal{X}



Rappel sur le principe des SVM

SVM spline,
2 juin 06

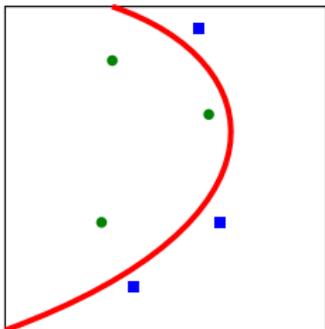
Nathalie VILLA,
Fabrice ROSSI

SVM pour
données
fonctionnelles

Approche par
L-Splines

Bibliographie

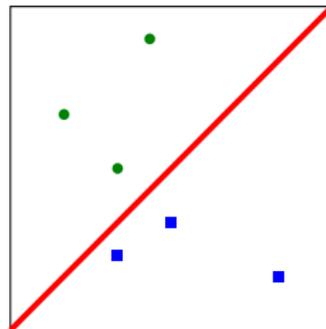
Espace initial \mathcal{H}



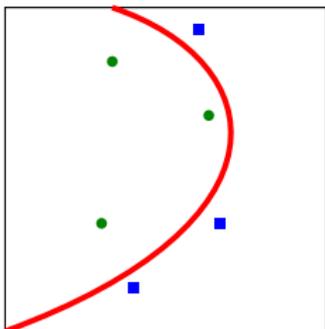
Φ (non linéaire)



Espace image \mathcal{X}



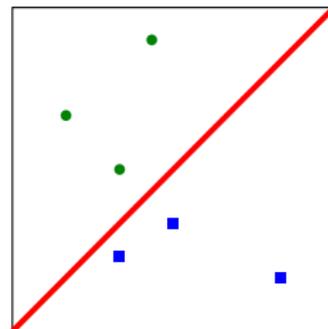
Espace initial \mathcal{H}



Φ (non linéaire)



Espace image \mathcal{X}



On cherche w tel que :

$$(P_{C,\mathcal{X}}) \quad \min_{w,b,\xi} \langle w, w \rangle + C \sum_{i=1}^n \xi_i,$$

sous les contraintes : $y_i(\langle w, \Phi(x_i) \rangle + b) \geq 1 - \xi_i, 1 \leq i \leq n,$
 $\xi_i \geq 0, 1 \leq i \leq n.$

SVM spline,
2 juin 06

Nathalie VILLA,
Fabrice Rossi

SVM pour
données
fonctionnelles

Approche par
L-Splines

Bibliographie

Formulation duale : $(P_{C,\mathcal{X}}) \Leftrightarrow$

$$\begin{aligned} (D_{C,\mathcal{X}}) \max_{\alpha} & \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle \Phi(x_i), \Phi(x_j) \rangle_{\mathcal{X}}, \\ \text{avec} & \sum_{i=1}^n \alpha_i y_i = 0, \\ & 0 \leq \alpha_i \leq C, \quad 1 \leq i \leq n. \end{aligned}$$

SVM spline,
2 juin 06

Nathalie VILLA,
Fabrice ROSSI

SVM pour
données
fonctionnelles

Approche par
L-Splines

Bibliographie

Formulation duale : $(P_{C,\mathcal{X}}) \Leftrightarrow$

$$(D_{C,\mathcal{X}}) \max_{\alpha} \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle \Phi(x_i), \Phi(x_j) \rangle_{\mathcal{X}},$$

avec $\sum_{i=1}^n \alpha_i y_i = 0,$
 $0 \leq \alpha_i \leq C, 1 \leq i \leq n.$

Produit scalaire dans \mathcal{X} :

$$\forall u, v \in \mathcal{X}, \quad K(u, v) = \langle \Phi(u), \Phi(v) \rangle_{\mathcal{X}}$$

SVM spline,
2 juin 06

Nathalie VILLA,
Fabrice ROSSI

SVM pour
données
fonctionnelles

Approche par
L-Splines

Bibliographie

Forme générale

Prétraitement : $P : \mathcal{H} \rightarrow \mathcal{D}$

$$\forall u, v \in \mathcal{H}, Q(u, v) = K(P(u), P(v)).$$

Forme générale

Prétraitement : $P : \mathcal{H} \rightarrow \mathcal{D}$

$$\forall u, v \in \mathcal{H}, Q(u, v) = K(P(u), P(v)).$$

① **Projections** : pour $V_D = \text{Vect} \{ \psi_1, \dots, \psi_D \}$,

$$P(x) = \sum_{j=1}^D \langle x, \psi_j \rangle \psi_j.$$

Forme générale

Prétraitement : $P : \mathcal{H} \rightarrow \mathcal{D}$

$$\forall u, v \in \mathcal{H}, Q(u, v) = K(P(u), P(v)).$$

- ① **Projections** : pour $V_D = \text{Vect} \{\psi_1, \dots, \psi_D\}$,

$$P(x) = \sum_{j=1}^D \langle x, \psi_j \rangle \psi_j.$$

- ② **Transformations fonctionnelles** : $P(x) = D^q x$.

Forme générale

Prétraitement : $P : \mathcal{H} \rightarrow \mathcal{D}$

$$\forall u, v \in \mathcal{H}, Q(u, v) = K(P(u), P(v)).$$

- 1 **Projections** : pour $V_D = \text{Vect} \{\psi_1, \dots, \psi_D\}$,

$$P(x) = \sum_{j=1}^D \langle x, \psi_j \rangle \psi_j.$$

- 2 **Transformations fonctionnelles** : $P(x) = D^q x$.

Problème : Approche en deux temps : calcul des dérivées par projection spline puis SVM ; pas de résultat de consistance.

SVM spline,
2 juin 06

Nathalie VILLA,
Fabrice ROSSI

SVM pour
données
fonctionnelles

Approche par
L-Splines

Bibliographie

1 SVM pour données fonctionnelles

2 Approche par *L*-Splines

SVM spline,
2 juin 06

Nathalie VILLA,
Fabrice ROSSI

SVM pour
données
fonctionnelles

Approche par
L-Splines

Bibliographie

Hypothèse de régularité

On suppose que X est **régulière** : $X \in \mathcal{H}^m = \{x : [0; 1] \rightarrow \mathbb{R} : \forall j = 1, \dots, m, D^j x \text{ existe et } D^m x \in L^2\}$.

Hypothèse de régularité

On suppose que X est **régulière** : $X \in \mathcal{H}^m = \{x : [0; 1] \rightarrow \mathbb{R} : \forall j = 1, \dots, m, D^j x \text{ existe et } D^m x \in L^2\}$.

Principe de l'interpolation L -spline

- on interpole **exactement** les observations x_1, \dots, x_n aux points de discrétisation t_1, \dots, t_d ;
- on minimise une pénalité de régularisation définie par un opérateur différentiel :

$$L = D^m + \sum_{j=1}^{m-1} a_j D^j.$$

Hypothèse de régularité

On suppose que X est **régulière** : $X \in \mathcal{H}^m = \{x : [0; 1] \rightarrow \mathbb{R} : \forall j = 1, \dots, m, D^j x \text{ existe et } D^m x \in L^2\}$.

Principe de l'interpolation L -spline

- on interpole **exactement** les observations x_1, \dots, x_n aux points de discrétisation t_1, \dots, t_d ;
- on minimise une pénalité de régularisation définie par un opérateur différentiel :

$$L = D^m + \sum_{j=1}^{m-1} a_j D^j.$$

- ces deux contraintes sont prises en compte **directement** dans le noyau.

SVM spline,
2 juin 06

Nathalie VILLA,
Fabrice ROSSI

SVM pour
données
fonctionnelles

Approche par
L-Splines

Bibliographie

$\mathcal{H}^m = \mathcal{H}_0 \oplus \mathcal{H}_1$ avec

- $\mathcal{H}_0 = \text{Ker}L$;
- \mathcal{H}_1 est défini par m contraintes aux bornes : espace de Hilbert de produit scalaire $\langle h_1, h_2 \rangle = \int_{[0,1]} Lh_1(t)Lh_2(t)dt$.
- \mathcal{H}_1 est un RKHS de noyau $K : \forall h \in \mathcal{H}_1, \forall t \in [0; 1], \langle h, K(t, \cdot) \rangle = h(t)$; K est déterminé par L et les contraintes définissant \mathcal{H}_1 (cf [Besse & Ramsay, 1986], [Berlinet & Thomas-Agnan, 2004]).

SVM spline,
2 juin 06

Nathalie VILLA,
Fabrice Rossi

SVM pour
données
fonctionnelles

Approche par
L-Splines

Bibliographie

$\mathcal{H}^m = \mathcal{H}_0 \oplus \mathcal{H}_1$ avec

- $\mathcal{H}_0 = \text{Ker}L$;
- \mathcal{H}_1 est défini par m contraintes aux bornes : espace de Hilbert de produit scalaire $\langle h_1, h_2 \rangle = \int_{[0,1]} Lh_1(t)Lh_2(t)dt$.
- \mathcal{H}_1 est un RKHS de noyau $K : \forall h \in \mathcal{H}_1, \forall t \in [0; 1], \langle h, K(t, \cdot) \rangle = h(t)$; K est déterminé par L et les contraintes définissant \mathcal{H}_1 (cf [Besse & Ramsay, 1986], [Berlinet & Thomas-Agnan, 2004]).

Exemple : \mathcal{H}^1 avec $L = I + D$ et $x(0) = 0$ ($Lx = 0 \Rightarrow x = ae^{-t}$ et $x(0) = a$) - dans ce cas, $K(s, t) = e^{-\max(s,t)} \sinh(\min(s, t))$.

SVM spline,
2 juin 06

Nathalie VILLA,
Fabrice ROSSI

SVM pour
données
fonctionnelles

Approche par
L-Splines

Bibliographie

On suppose $\mathbf{K}_d = (K(t_i, t_j))_{i,j=1,\dots,d}$ inversible.

Spline d'interpolation

Soit $x \in \mathcal{H}_1$ alors $h = \mathcal{P}_{\text{Vect}\{K(t_k, \cdot), k=1,\dots,d\}}(x)$.

SVM spline,
2 juin 06

Nathalie VILLA,
Fabrice ROSSI

SVM pour
données
fonctionnelles

Approche par
L-Splines

Bibliographie

On suppose $\mathbf{K}_d = (K(t_i, t_j))_{i,j=1,\dots,d}$ inversible.

Spline d'interpolation

Soit $x \in \mathcal{H}_1$ alors $h = \mathcal{P}_{\text{Vect}\{K(t_k, \cdot), k=1,\dots,d\}}(x)$.

Produit scalaire entre splines

Si $x_1, x_2 \in \mathcal{H}_1$,

$$\langle h_1, h_2 \rangle_{\mathcal{H}_1} = \langle \mathbf{x}_1, \mathbf{x}_2 \rangle_{(\mathbb{R}^d, \mathbf{K}_d^{-1/2})}.$$

où $\mathbf{x}_i = (x_i(t_1), \dots, x_i(t_d))$.

On note

- $G_\gamma^d(u, v) = e^{-\gamma \|u-v\|_{\mathbb{R}^d}^2}$ et $G_\gamma^\infty(u, v) = e^{-\gamma \|u-v\|_{L^2}^2}$;
- $\mathbf{K}_d = (K(t_i, t_j))_{i,j=1,\dots,d}$, supposée inversible ;
- $\forall i = 1, \dots, n$, h_i est la spline d'interpolation de x_i (supposée à valeurs dans \mathcal{H}_1) ;
- $\mathbf{x}_i = (x_i(t_1), \dots, x_i(t_d))$.

Théorème 1

SVM sur dérivées

SVM sur $(Lh_i)_i$ avec noyau G_γ^∞

→

$\phi_h^{n,d}$

↔

SVM sur $(\mathbf{x}_i)_i$ avec noyau $G_\gamma^d \circ \mathbf{K}_d^{-1/2}$

→

$\phi_{\mathbf{x}}^{n,d}$

SVM spline,
2 juin 06

Nathalie VILLA,
Fabrice ROSSI

SVM pour
données
fonctionnelles

Approche par
L-Splines

Bibliographie

Hypothèses

- **(H1)** : X est une variable aléatoire bornée à valeurs dans \mathcal{H}_1 ;
- **(H2)** : $(\tau_d)_d$ est une suite de points de discrétisation dans $[0, 1]$ telle que, pour tout $d \geq 1$, $\tau_d = \{t_k\}_{k=1, \dots, d}$, la matrice \mathbf{K}_d est définie positive et $\text{Span}\{K(t, \cdot), t \in \cup_{d \geq 1} \tau_d\}$ est dense dans \mathcal{H}_1 ;
- **(H3)** : $(C_n^d)_n$ est une suite de régularisation telle que $C_n^d = O(n^{1-\beta_d})$ pour $0 < \beta_d < 1/d$.

SVM spline,
2 juin 06

Nathalie VILLA,
Fabrice ROSSI

SVM pour
données
fonctionnelles

Approche par
L-Splines

Bibliographie

Théorème 2

Consistance universelle

Sous les hypothèses **(H1)**-**(H3)**, le SVM $\phi_h^{n,d}$, avec la suite de régularisation $(C_n^d)_n$, est universellement consistant :

$$\lim_{d \rightarrow +\infty} \lim_{n \rightarrow +\infty} \text{Err} \phi_h^{n,d} = \text{Err}^*$$

où Err^* est l'erreur de Bayes et $\text{Err} \phi = \mathbb{P}(\phi(X) \neq Y)$.

SVM spline,
2 juin 06

Nathalie VILLA,
Fabrice ROSSI

SVM pour
données
fonctionnelles

Approche par
L-Splines

Bibliographie

- Matrice \mathbf{K}_d mal conditionnée dès que d est grand \Rightarrow Difficilement inversible dans les problèmes réels ;
- **Perspectives** : Régularisation de la matrice par l'utilisation de splines de lissage.

SVM spline,
2 juin 06

Nathalie VILLA,
Fabrice ROSSI

SVM pour
données
fonctionnelles

Approche par
L-Splines

Bibliographie



Berlinet, A. & Thomas-Agnan, C. (2004).
Reproducing kernel Hilbert spaces in probability and statistics.
Kluwer Academic Publisher.



Besse, P. & Ramsay, J. (1986).
Principal component analysis of sampled curves.
Psychometrika, **51**, 285–311.



Rossi, F. & Villa, N. (2006).
Support vector machine for functional data classification.
Neurocomputing, **69**(7-9), 730–742.