Classification SVM pour données fonctionnelles

M2 ISIFAR
23/03/2020
Shon Amsalhem
François Le Gac

Introduction

Objectif :

Peut-on prendre en compte la structure spécifique des données fonctionnelles pour améliorer les performances de l'algorithme SVM ?

Définition : Variable fonctionnelle et donnée fonctionnelle

Une variable aléatoire est dite fonctionnelle si ses valeurs sont dans un espace de dimension infinie. Une observation d'une variable fonctionnelle est appelée donnée fonctionnelle. Ce qui veut dire que chaque individu correspond non plus à un vecteur mais à une courbe, à un continuum.

Exemples : *Données spectrométriques, reconnaissance vocale ...*

Application : le jeu de données octane

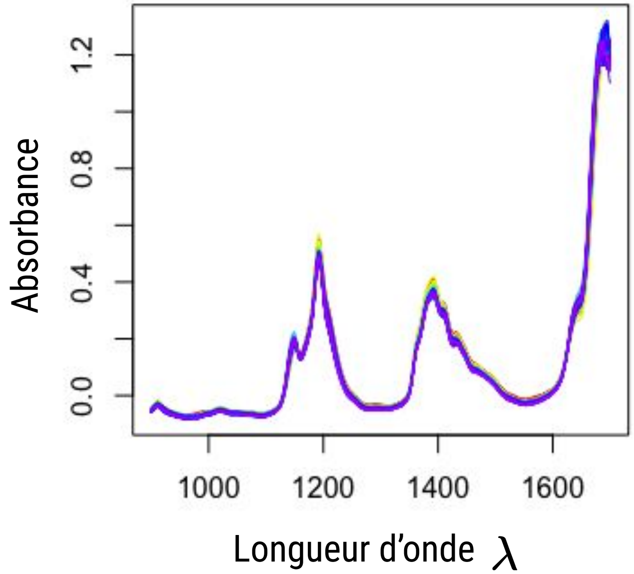


Fig 0. Spectres d'échantillons d'essence

	Col 1	Col 2		Col 400	Var. cible
Ech-1	$X_1(\lambda_{900})$	$X_1(\lambda_{902})$	...	$X_1(\lambda_{1700})$	0
...	...	...	...	...	...
Ech-60	$X_{60}(\lambda_{900})$	$X_{60}(\lambda_{902})$	...	$X_{60}(\lambda_{1700})$	1

[Octane data set : references]

Fig 1. jeu de données octane

Objectif : classer les 60 échantillons d'essence en deux catégories : “non résistante à l'auto-allumage (0)“ et “résistante à l'auto-allumage (1)”

Problème posé par les données fonctionnelles

Exemple 1 (Théorique) : Régression Linéaire

$X \in \mathbb{R}^d$ on souhaite prédire $Y \in \mathcal{Y}$

$$\mathbb{E}(Y|X) = X\beta$$

On estime aisément β par la méthode des moindres carrés, qui conduit à l'inversion de la matrice de covariance de X. Si d >> N (N = nombre d'obs.), alors il devient impossible d'inverser la matrice de covariance.

Exemple 2 (Pratique) : SVM

En pratique, les performances (précision) sont mauvaises : 62.9% (Linéaire), 63.2% (Gaussien)

Bref rappel sur SVM (1)

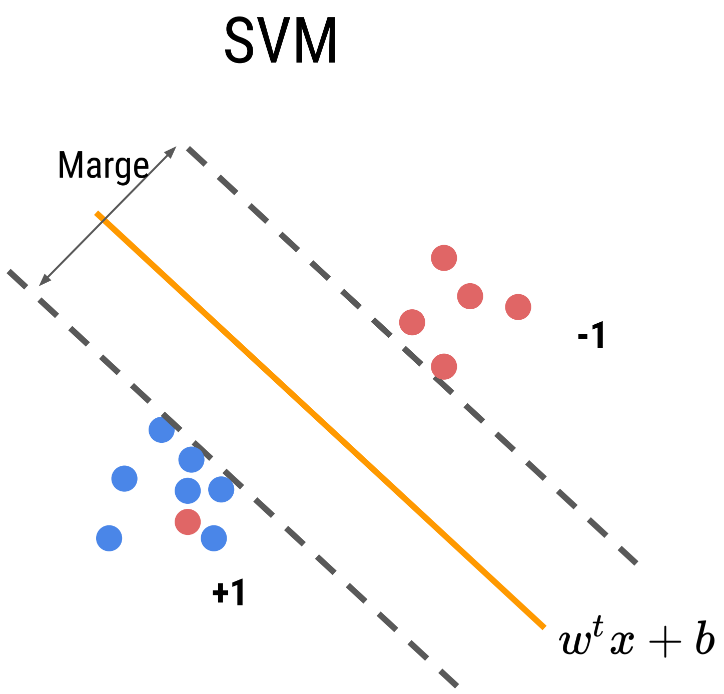


Fig 2. Support vector machine

Problème de “Soft Margin”

$$\begin{cases} min\ \frac{1}{2}||w||^2 + C\sum_{i=1}^{n} \xi_i \\ tel\ que\ \forall i, y_i(w^t x_i + b) \geq 1 - \xi_i \\ avec\ \xi_i \geq 0, i = 1, \ldots, n \end{cases}$$

SVM non linéaire :

Les données deviennent $(\phi(x_1), y_1), \ldots, (\phi(x_N), y_N)$

La règle de décision :

$$x \mapsto sign(\sum_{i=1}^{N} \alpha_i y_i \langle \phi(x_i), \phi(x) \rangle + b)$$

Bref rappel sur SVM (2)

Kernel

$$E \times E \to \mathbb{R}$$
$$(x_1, x_2) \mapsto K(x_1, x_2) = \phi(x_1)^t \phi(x_2)$$

(Th. Mercer : K doit être continue, symétrique, semi-définie positive)

Exemples :

- Linéaire : $K(x_1, x_2) = \langle x_1, x_2 \rangle + c$
- Gaussien : $K(x_1, x_2) = exp(-\gamma ||x_1 - x_2||^2)$

Transformation fonctionnelle (1)

Courbes des individus

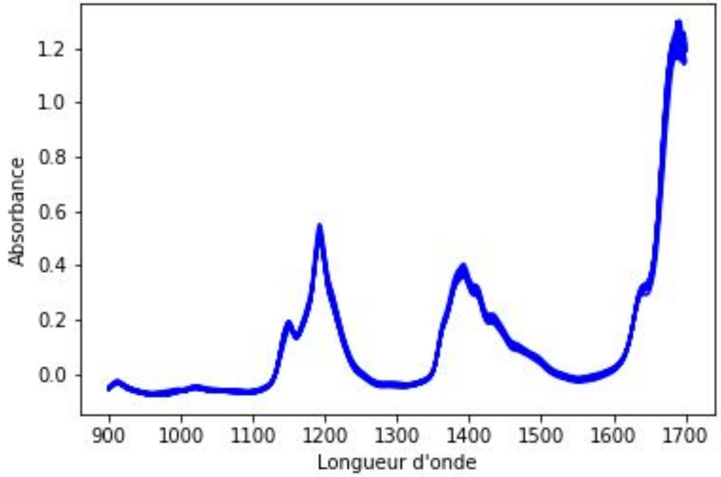


Fig 3. Courbe - non résistant à l'auto allumage (0)

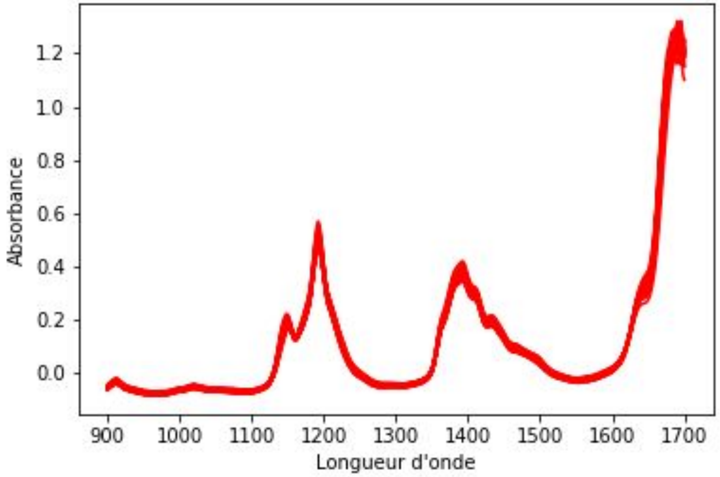


Fig 4. Courbe - résistant à l'auto allumage (1)

Transformation fonctionnelle (2)

Interpolation

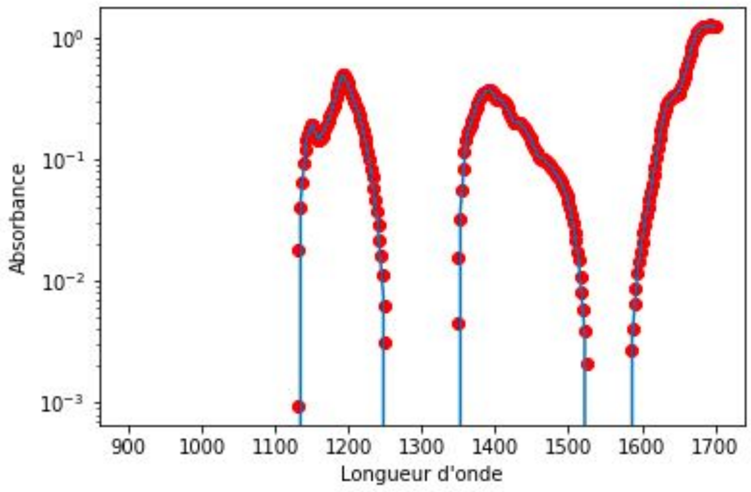


Fig 5. Interpolation

Dérivée seconde (1 obs.)

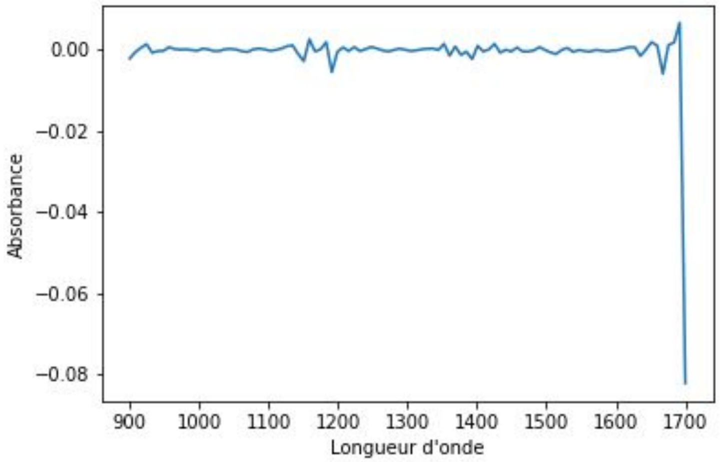


Fig 6. Dérivée seconde pour une obs.

Transformation fonctionnelle (3)

Dérivées secondes

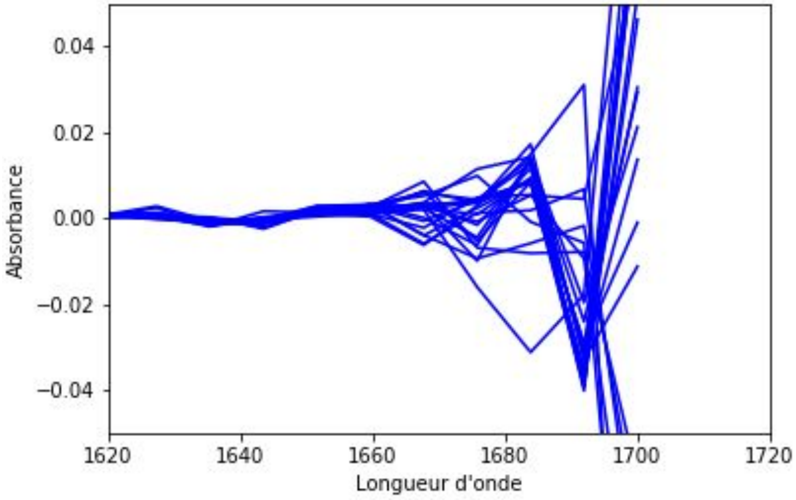


Fig 7. Dérivée seconde - non résistant à l'auto allumage (0)

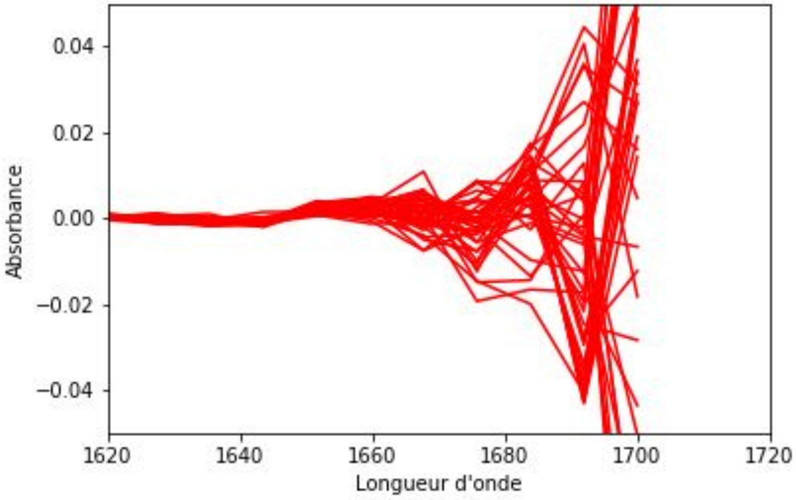
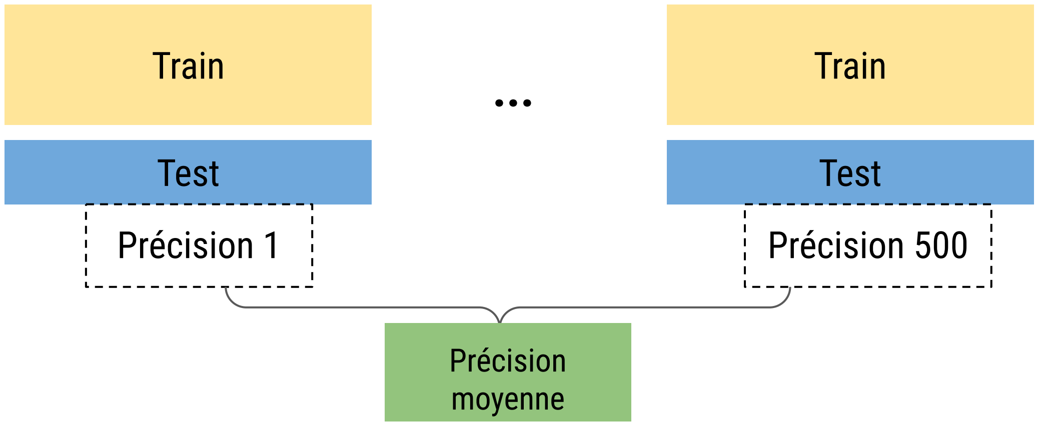


Fig 8. Dérivée seconde - résistant à l'auto allumage (1)

Résultats

Repeated CV (500x) :



	KNN	**SVM**			
	Méthode de base (données brut)	Linéaire - données brutes	Gaussien - données brutes	Linéaire - **dérivée seconde**	Gaussien - **dérivée seconde**
Accuracy	**0.81**	**0.63**	**0.63**	**0.65**	**0.60**

Fig 9. Performances des modèles

MERCI