

## INTRODUCTION AUX MACHINES A VECTEURS DE SUPPORT (SVM)

### 1. Historique

L'aptitude à généraliser des résultats obtenus à partir d'un échantillon limité constitue l'enjeu essentiel de l'apprentissage artificiel (*machine learning*). Il n'est plus à démontrer que la seule minimisation du risque empirique (l'erreur d'apprentissage) ne garantit pas une faible erreur sur un corpus de test. Ainsi les techniques de régularisation, utilisées depuis les années 1960, permettent-elles de réaliser un compromis entre la capacité du modèle à apprendre (liée à sa complexité) et son aptitude à généraliser. Du point de vue conceptuel, la notion de risque structural introduite par Vladimir Vapnik dans les années 1990 [1] donne une borne de l'erreur de test en fonction de l'erreur d'apprentissage et de la complexité du modèle. En pratique, les machines à vecteurs de support (*Support Vector Machines* ou *SVM*), présentées en 1992 [2], offrent un moyen opérationnel pour minimiser le risque structural, ce qui explique l'engouement suscité dans la communauté scientifique.

### 2. Principe

A l'origine conçues pour les tâches de classification ou reconnaissance de formes, les SVM permettent également de traiter les problèmes de régression non linéaire. Nous privilégions ici leur intérêt pour la reconnaissance de formes (i.e. la découverte de caractéristiques communes dans un ensemble de données sans connaissance *a priori* du domaine). Dans ce contexte, l'acronyme SVM peut être judicieusement traduit par Séparateurs à Vaste Marge [3].

Le principe théorique des SVM comporte deux points fondamentaux :

- la transformation non linéaire ( $\Phi$ ) des exemples de l'espace d'entrée vers un espace dit de redescription de grande dimension muni d'un produit scalaire (espace de Hilbert),

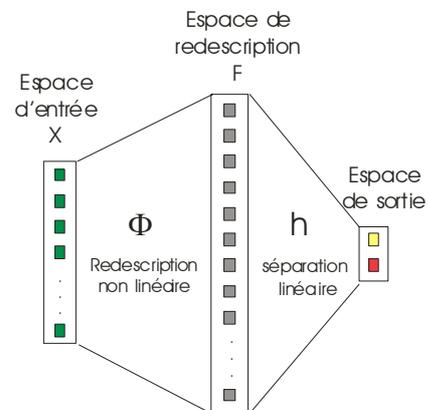


Figure 1. La transformation non linéaire des données peut permettre une séparation linéaire des exemples dans un nouvel espace. Adapté de [3].

- la détermination d'un hyperplan permettant une séparation linéaire optimale dans cet espace de grande dimension.

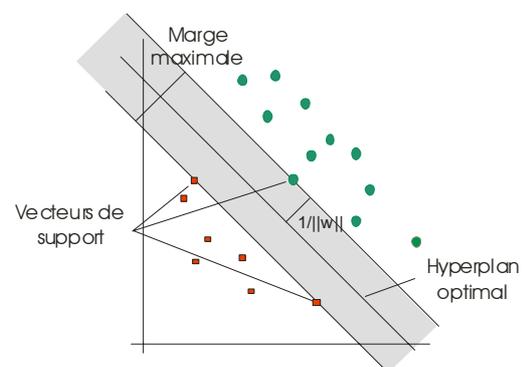


Figure 2. L'hyperplan séparateur optimal est celui qui maximise la marge dans l'espace de redescription.

L'intérêt est que dans l'espace de redescription la reconnaissance de formes peut s'avérer tâche aisée : en effet, intuitivement, plus la dimension de l'espace de redescription est grande, plus la probabilité de pouvoir trouver un hyperplan séparateur entre les exemples est élevée. Ce qu'illustre le schéma suivant.

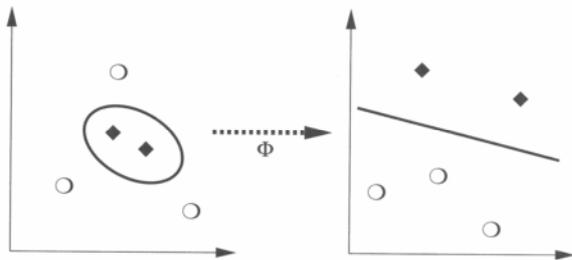


Figure 3. L'idée des SVM : transformer un problème de séparation non linéaire dans l'espace de représentation en un problème de séparation linéaire dans un espace de re-description de grande dimension. Adapté de [5].

Du point de vue mathématique, la transformation non linéaire ( $\Phi$ ) est réalisée *via* une fonction noyau (ou noyau de Hilbert-Schmidt) facile à calculer. Ainsi l'espace de redescription reste virtuel, jamais explicité.

En pratique, quelques familles de fonctions noyau paramétrables sont connues et il revient à l'utilisateur de SVM d'effectuer des tests pour déterminer celle qui convient le mieux pour son application (il s'agit de traduire le maximum de connaissances préalables dont on dispose sur le problème étudié et sur les données).

### 3. Optimisation

Le problème de recherche de l'hyperplan séparateur optimal possède une formulation duale. Ceci est particulièrement intéressant car, sous cette formulation duale, le problème peut être résolu au moyen de méthodes d'optimisation quadratique standard.

Différents programmes d'optimisation sont disponibles gratuitement sur Internet [4] : CPLEX, LOQO, MINOS, SMO, etc.

### 4. Applications

Les méthodes à noyaux, inspirées de la théorie statistique de l'apprentissage de Vapnik, et dont les SVM constituent la forme la plus connue, ont été adaptées à divers contextes : classification avec apprentissage, estimation de support de densité, régression, etc.

Les SVM sont appliqués avec une efficacité remarquable à la reconnaissance de caractères manuscrits, au traitement d'images, à la prédiction de séries temporelles, au diagnostic médical, au contrôle qualité, etc. Des exemples d'application réussie des SVM peuvent être consultés sur Internet [6].

### Références

- [1] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, 1995.
- [2] B.E. Boser, I.M. Guyon, V. Vapnik, *A Training Algorithm for Optimal Margin Classifiers*, Proc. Fifth Ann. Workshop Computational Learning Theory, ACM Press, 1992, pp. 144-152.
- [3] A. Cornuéjols, L. Miclet, Y. Kodratoff, *Apprentissage artificiel*, Eyrolles, 2002. Chapitres 2 et 9.
- [4] <http://www.kernel-machines.org/>
- [5] B. Schölkopf, A.J. Smola, *Learning with kernels*, MIT Press, 2002.
- [6] <http://www.clopinet.com/isabelle/Projects/SVM/appllist.html>.