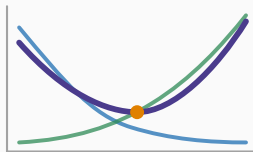


Apprentissage statistique

Chapitre 11 : Les machines à vecteurs de support



Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 27 août 2026

Séparer par une marge

La marge souple

L'astuce du noyau

Où les SVM se situent

Ce qu'il faut retenir

Séparer par une marge

L'intuition géométrique

Le problème posé autrement

Si deux classes sont séparables par une droite, il existe une **infinité** de droites qui les séparent.

Laquelle choisir ? Celle qui laisse la plus grande **marge** de sécurité de part et d'autre.

Hyperplan de marge maximale

La **marge** est la distance entre la frontière et les points les plus proches.

L'**hyperplan de marge maximale** est celui qui la rend la plus grande possible.

Vecteurs de support

Seuls les points situés **sur la marge** déterminent la frontière. Ce sont les **vecteurs de support**.

Déplacer un point éloigné ne change rien ; déplacer un vecteur de support change tout.

Une propriété remarquable, et ambivalente

La solution ne dépend que d'une poignée d'observations.

C'est une **force** — le modèle est compact, insensible aux points lointains. C'est aussi une **faiblesse** : une observation aberrante placée près de la frontière peut la déformer entièrement.

La marge souple

Pourquoi la marge dure ne suffit pas

Les données réelles ne sont presque jamais parfaitement séparables. Et même quand elles le sont, exiger une séparation parfaite conduit à une marge minuscule, dictée par quelques points douteux.

Le paramètre C

La **marge souple** autorise des violations, pénalisées par un coût C .

C	Marge	Violations	Comportement
petit	large	nombreuses tolérées	plus de biais, moins de variance
grand	étroite	peu tolérées	moins de biais, plus de variance

Intuition

C est un paramètre de **régularisation**, exactement comme λ au chapitre 4 et α au chapitre 8.

Un C petit contraint le modèle; un C grand le laisse coller aux données. Il se choisit par **validation croisée**, comme tous les autres.

Le vocabulaire change de chapitre en chapitre; **la question est toujours la même.**

L'astuce du noyau

Rendre linéaire ce qui ne l'est pas

L'idée

Des données non séparables dans leur espace d'origine peuvent le devenir dans un espace de **plus grande dimension**.

Un cas d'école

Des points d'une classe au centre, entourés d'un anneau de l'autre classe. Aucune droite ne les sépare dans le plan.

En ajoutant une troisième coordonnée $x_1^2 + x_2^2$, les points centraux descendent et les points périphériques montent : **un simple plan horizontal les sépare**.

L'astuce du noyau

L'algorithme ne dépend des données qu'à travers des **produits scalaires** entre observations.

Un **noyau** calcule directement ce produit scalaire dans l'espace transformé, **sans jamais construire les coordonnées**. On travaille donc en très grande dimension au prix d'un calcul en dimension d'origine.

Les noyaux usuels

Où les SVM se situent

Un bilan honnête

Forces et limites

Points forts

Efficaces en **grande dimension**

Solution compacte et stable

Le noyau donne beaucoup de souplesse

Théorie solide

Points faibles

coût de calcul **quadratique** en n

pas de probabilité directe

standardisation obligatoire

deux paramètres à régler conjointement

Quand y penser aujourd'hui

Les SVM ont dominé les années 1990 et 2000. Le boosting et les réseaux les ont largement dépassés depuis sur la plupart des tâches.

Ils restent excellents quand p est **grand** et n **modéré** — texte, données génomiques, spectres — c'est-à-dire précisément le cas où les méthodes d'ensemble manquent d'observations.

Ce n'est plus le premier choix, mais c'est parfois le bon.

Ce qu'il faut retenir

Cinq points

1. La SVM choisit l'hyperplan de **marge maximale**.
2. Seuls les **vecteurs de support** déterminent la frontière — force et fragilité à la fois.
3. La **marge souple** tolère des violations; C règle le compromis biais-variance.
4. L'**astuce du noyau** permet de travailler en très grande dimension sans construire les coordonnées.
5. Coût **quadratique en n** ; adaptées aux cas où p est grand et n modéré.

La suite

Cinq familles de modèles ont été construites, de plus en plus performantes et de moins en moins lisibles.

Le chapitre 12 pose la question qui reste : que peut-on dire d'un modèle qu'on ne peut plus lire?