SVM (Séparateur à Vaste Marge)

Schölkopf et Smola (2002) → livre de référence

Classe d’algorithmes d’apprentissage initialement définis pour la discrimination c’est-à-dire la prévision d’une **variable qualitative binaire** / quantitative.

Le principe est donc de trouver un classifieur, ou une fonction de discrimination, dont la capacité de généralisation (qualité de prévision) est la plus grande possible.

Elle s’est focalisée sur les propriétés de généralisation (ou prévision) d’un modèle en contrôlant sa complexité.

Le principe fondateur des SVM est justement d’intégrer à l’estimation le contrôle de la complexité c’est-à-dire le nombre de paramètres qui est associé dans ce cas au nombre de vecteurs supports.

2e idée : éviter de substituer à l’objectif initial : la discrimination, un ou des problèmes qui s’avèrent finalement plus complexes à résoudre comme par exemple l’estimation non-paramétrique de la densité d’une loi multidimensionnelle en analyse discriminante.

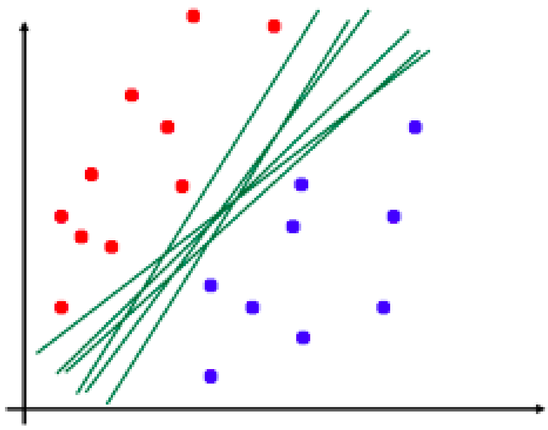
La minimisation du risque dépend :

- du risque empirique
- un risque structurel lié au terme dVC qui dépend de la complexité du modèle h choisi (VC-dimension[a])

Ainsi, si pour construire un bon modèle d’apprentissage, il est nécessaire de : minimiser les erreurs sur la base d’apprentissage c’est le principe d’induction naïf utilisé dans les réseaux de neurones ; de construire un système capable de généraliser correctement

Il existe de nombreux choix possibles pour **w** et b :

$$y(x) = signe(\mathbf{w} \cdot \mathbf{x} + b) = signe(k\mathbf{w} \cdot \mathbf{x} + k \cdot b)$$



Nous avons deux classes de points, et nous voulons chercher l’hyperplan dans lequel la marge entre les deux classes est la plus grande possible → la maximiser

$r = | w . x + b | / \| w \| = 2 / \| w \|$ (trouve w et b pour que cela soit maximal) $\rightarrow$ $1/2 \| w \|$

Les points qui vérifient $y_i(w \cdot x_i + b) = 1$, sont appelés “vecteurs supports”. Ce sont les points les plus près de la marge. Ils sont sensés être peu nombreux par rapport `a l’ensemble des exemples.

SMO / SimpleSVM / LASVM

L’idée est d’ajouter des variables d’ajustement ξ_i dans la formulation pour prendre en compte les erreurs de classification ou le bruit

Le principe de base des SVM consiste de ramener le problème de la discrimination à celui, linéaire, de la recherche d’un hyperplan optimal :

- La première consiste à définir **l’hyperplan** comme solution d’un problème d’optimisation sous contraintes dont la fonction objectif ne s’exprime qu’à l’aide de produits scalaires entre vecteurs et dans lequel le nombre de contraintes “actives” ou vecteurs supports contrôle la complexité du modèle.
- Le passage à la recherche de surfaces séparatrices non linéaires est obtenu par l’introduction d’une fonction noyau (kernel) dans le produit scalaire induisant implicitement une transformation non linéaire des données vers un espace intermédiaire (feature space) de plus grande dimension. D’où l’appellation couramment rencontrée de machine à noyau ou kernel machine. Sur le plan théorique, la fonction noyau définit un espace hilbertien, dit auto-reproduisant et isométrique par la transformation non linéaire de l’espace initial et dans lequel est résolu le problème linéaire
    - *espace hilbertien : un espace vectoriel réel (resp. complexe) muni d'un produit scalaire euclidien (resp. hermitien)*
    - *kernel machine : utiliser un classifieur linéaire pour résoudre un problème non linéaire*

Principaux principes des SVM : marge, programmation quadratique, vecteur support, formulation duale et matrice de gram

Si f (.) est discrète on parle de classification
Si f (.) est une fonction continue on parle alors de régression

Noyau de Mercer

- On appel noyau de Mercer une fonction continue, symétrique, semi-définie positive $K(x, y)$
- $\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j k(x_i, x_j), \forall \alpha \in \mathbb{R}$

Matrice de Gram

Matrice des termes $< x_i, x_j >$. Elle est symétrique et semi-définie positive pour une noyau de Mercer

Théorème de Moore-Aronszajn (1950)

- Toute fonction semi-définie positive $k(x, y)$ est un noyau, et réciproquement. Elle peut s'exprimer comme un produit scalaire dans un espace de grande dimension.
- $k(x_i, x_j) =< \phi(x_i), \phi(x_j) >$

Applications de SVM

- Les avantages théoriques (minimisation de l'erreur empirique et structurelle) et pratiques (algorithmes optimisés) des SVM en ont fait un outil très prisé dans de nombreux problèmes pratiques de classification.
- Dans bien des cas, il s'agit de construire un noyau (donc une mesure de similarité) adapté aux données à traiter.

NN (Neuron Networks)

Simulation du cerveau humain par des procédés informatique.

Un réseau de neurones artificiels, ou réseau neuronal artificiel, est un système dont la conception est à l'origine schématiquement inspirée du fonctionnement des neurones biologiques, et qui par la suite s'est rapproché des méthodes statistiques.

L'apprentissage profond (en anglais **deep learning**) est un ensemble de méthodes d'apprentissage automatique tentant de modéliser avec un haut niveau d'abstraction des données grâce à des architectures articulées de différentes transformations non linéaires.
→ prédire une sortie Y (une caractéristique) à travers un ensemble de données Xi en entrée,

Les techniques d'apprentissage profond constituent une classe d'algorithmes d'apprentissage automatique qui :

- utilisent différentes couches d'unité de traitement non linéaire pour l'extraction et la transformation des caractéristiques ; chaque couche prend en entrée la sortie de la précédente ; les algorithmes peuvent être supervisés ou non supervisés, et leurs applications comprennent la reconnaissance de modèles et la classification statistique ;
- fonctionnent avec un apprentissage à plusieurs niveaux de détail ou de représentation des données ; à travers les différentes couches, on passe de paramètres de bas niveau à des paramètres de plus haut niveau, où les différents niveaux correspondent à différents

niveaux d’abstraction des données.

Bibliographie SVM

https://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-app-svm.pdf

http://helios.mi.parisdescartes.fr/~lomn/Cours/DM/Material/ComplementsCours/SVM.pdf

http://pageperso.univ-lr.fr/arnaud.revel/MesPolys/SVM.pdf

Bibliographie NN

https://fr.wikipedia.org/wiki/Réseau_de_neurones_artificiels

https://meritis.fr/ia/deep-learning/