

Big Data Analytics : Support Vector Machine

Nom : **PULL**

Prénom : **Yoann**

Année : **M2**

Semestre : **9**

Nature : **CM**

Volume horaire : **12**

ECTS / Coef : **2**

Prérequis	➤ Modèles de régression linéaire ➤ Méthodes d'optimisation
Résumé	Ce cours présente les principes de base des machines à vecteur de support (Support Vecteur Machine ou SVM) destinés à résoudre les problèmes de classification ou de régression. La première partie est consacrée à la présentation de l'intuition des SVM dans le cas d'un échantillon linéairement séparable et d'un problème de classification. Dans cette section, nous introduisons les concepts d'hyperplan séparateur optimal, de marge, de vecteur de support, etc. Dans une seconde partie, le problème d'optimisation des SVM est formalisé sous la forme primale et sous la forme duale. La troisième partie introduit la notion de « soft margin » et de variables ressorts. La quatrième partie est consacrée au cas d'un échantillon non séparable et au « kernel trick », ou astuce du kernel. Différentes applications sont proposées sur les logiciels R, Python et SAS. Ces applications visent à illustrer les concepts théoriques étudiés dans le cours, et notamment à mettre en évidence l'influence des hyperparamètres (paramètres des fonctions kernel, paramètre de pénalisation) sur les résultats de la classification ou de la régression.
Objectifs	L'objectif du cours est double. Le premier objectif vise à introduire la notion théorique de séparateur à vaste marge et sa formalisation mathématique. Dans ce cadre, nous insisterons tout particulièrement sur la notion de transformation de l'espace de représentation des inputs (feature) au travers de l'astuce du kernel. Le second objectif est de présenter les principales procédures d'implémentation des SVM (pour des problèmes de classification) ou des SVR (pour des problèmes de régression) sur les logiciels SAS, R et Python. Dans ce contexte, nous insisterons tout particulièrement sur l'influence des hyperparamètres (paramètres des fonctions kernel, paramètre de pénalisation) sur les résultats de la classification ou de la régression.
Bibliographie	➤ Cristianini, N., and Shawe-Taylor, J. (2000). An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. New York: Cambridge University Press. ➤ Hastie, T., Tibshirani R., and Friedman J. (2009). The elements of statistical learning. Data mining, inference, and prediction. 2nd ed. Springer ➤ Vapnik V.N. (1998). Statistical Learning Theory. John Wiley. ➤ Vapnik V.N. (1995). The Nature of Statistical Learning Theory. Springer, First Edition.

PLAN

Le plan de ce cours est le suivant :

1. Introduction
2. Intuition des SVM : le cas linéairement séparable
3. Formalisation du SVM
4. Soft Margin
5. Kernel trick
6. Applications du SVM sous SAS, R, Python et Matlab
7. Extensions du SVM
 - A. SVM et scores
 - B. SVM multi-classes
 - C. Régressions à vecteurs de support (SVR)
 - D. Least Square (LS) - SVM
8. Conclusion