

Comparaison des méthodes dites de « Machine learning »

Projet M2 MIGS 2016-17

Ce projet porte sur la comparaison de méthodes permettant de palier au problème de la grande dimension dans le cas du modèle de régression linéaire. La grande dimension correspond à la situation où le nombre de variables est nettement plus important que le nombre d'observations.

Avec l'avènement des nouvelles technologies de plus en plus sensibles, un grand nombre de méthodes ont été proposées dans la littérature. Dans ce projet on s'intéressera à la compréhension et à la comparaison de deux de ces méthodes, à savoir :

- Support Vector Machine (SVM) : Bennett, K. P. & Campbell, C. (2000). Support vector machines: Hype or hallelujah? *SIGKDD Explorations*,**2**(2).
<http://www.acm.org/sigs/sigkdd/explorations/issue2-2/bennett.pdf>
- Non-Negative Least Squares (NNLS): Slawski, M & Hein, M (2011). Sparse recovery by thresholded non-negative least squares. *Advances in Neural Information Processing Systems* 24. 1926—1934.
<http://papers.nips.cc/paper/4231-sparse-recovery-by-thresholded-non-negative-least-squares.pdf>

Remarque: les publications proposées sont des pistes pour commencer

Des données simulées, éventuellement réelles, seront utilisées pour comparer les 2 méthodes. Les packages *e1071* et *nnls* du logiciel R pourront être utilisés pour l'application de ces méthodes.

Projet encadré par C.Truntzer