

## *Apprentissage en ligne et agrégation*

Pierre Alquier - ENSAE - CREST

Cours : 15 heures - TP : 6 heures

### *Objectif*

Ce cours constitue une introduction à l'apprentissage en-ligne, c'est-à-dire quand les données sont révélées au fur et à mesure du processus d'apprentissage plutôt que sous la forme d'un échantillon donné une fois pour toutes. Après une rapide introduction aux méthodes incontournables (halving, online gradient), on s'intéressera aux méthodes d'agrégation. L'idée de base est, étant donné plusieurs prédicteurs, de les faire voter en leur attribuant des poids spécifiques plutôt que d'en choisir un seul. Ces méthodes permettront des résultats optimaux dans des conditions extrêmement générales. Dans un second temps, on reviendra au cadre d'apprentissage 'batch' ou 'off-line' plus classique : on verra que les méthodes d'agrégation proposées précédemment peuvent également s'utiliser dans ce cas. Dans ce cas, on obtient des bornes précises sur l'erreur de généralisation, et des inégalités oracles précises, dans lesquelles la complexité de la famille de prédicteurs est contrôlée à l'aide d'une loi de probabilité qui joue un rôle similaire à la loi a priori en statistique bayésienne. Pour cette raison, ces résultats théoriques sont souvent appelés bornes 'PAC-Bayésiennes'. On discutera également les différents algorithmes possibles pour implémenter ces méthodes : MCMC et méthodes variationnelles. Trois séances seront dédiées à l'implémentation en R ou Python des algorithmes vus en cours et à leur test sur des jeux de données. Le cours sera évalué par un court projet (résumé d'un article de recherche, implémentation éventuelle des méthodes proposées dans l'article).

## Plan

LECTURE 1 : Introduction Learning theory : notations. Different settings of machine learning : online vs. batch. Realizable vs. non realizable. MS, C and L-type aggregation. Halving algorithm in the realizable case.

LECTURE 2 : Online gradient algorithm. A first example of online algorithm : online gradient descent. Examples. Online to batch bounds. REFERENCE FOR LECTURES 1 & 2 : S. Shalev-Schwartz, Online Learning and Online Convex Optimization, Foundations and Trends in Machine Learning vol. 4, 2011. (Chapters 1 and 2, but chapters 3, 4 and 5 are worth the effort if you have time !).

LECTURE 3 & 4 : Online PAC-Bayesian aggregation. PAC-Bayesian bounds for the Exponentially Weighted Aggregate (EWA) in the online setting. Slow rates, fast rates. Examples : classification, regression. Multiplicative weights algorithms for the MS-type aggregation. REFERENCES FOR LECTURES 3 & 4 : N. Cesa-Bianchi & G. Lugosi, Prediction, learning and games, Cambridge University Press, 2006. S. Gerchinovitz, Prediction of individual sequences and prediction in the statistical framework : some links around sparse regression and aggregation techniques, PhD Thesis, Univ. Paris 11, 2011. (Chapters 2 and 3).

LECTURE 5 & 6 : PAC-Bayesian bounds for batch learning. Hoeffding and Bernstein inequalities. PAC-Bayesian bounds for the EWA in the batch setting. Slow rates in the general case. Fast rates under Bernstein and margin assumptions. Examples : classification, regression, matrix factorization. REFERENCE FOR LECTURES 5 & 6 : O. Catoni, Pac-Bayesian supervised classification : the thermodynamics of statistical learning, IMS Lecture Notes, 2007. (This is an advanced reading !).

LECTURE 7 : Algorithms. REFERENCE FOR LECTURES 7 : B. Bishop, Pattern recognition and machine learning, Springer, 2006. (Chapters 10 for VB & 11 for Monte-Carlo).

+ 3 sessions with R or Python to implement the different algorithms

## Références

Sur les aspects algorithmiques :

B. Bishop, Pattern recognition and machine learning, Springer, 2006.

Sur les aspects plus théoriques :

S. Shalev-Schwartz, Online Learning and Online Convex Optimization, Foundations and Trends in Machine Learning vol. 4, 2011.

N. Cesa-Bianchi & G. Lugosi, Prediction, learning and games, Cambridge University Press, 2006.

O. Catoni, Statistical learning theory and stochastic optimization, Springer lecture notes in statistics, 2004.

C. Giraud, Introduction to high-dimensional statistics, CRC Press books, 2015.