



Statistical Inference with Ensemble of Clustered Desparsified Lasso

Jérôme Alexis Chevalier – Inria

DataSense: Thèse HiDimStat - Statistical control of sparse models in high dimension – 2017

Abstract:

Medical imaging involves high-dimensional data, yet their acquisition is obtained for limited samples. Multivariate predictive models have become popular in the last decades to fit some external variables from imaging data, and standard algorithms yield point estimates of the model parameters. It is however challenging to attribute confidence to these parameter estimates, which makes solutions hardly trustworthy. We present a new algorithm "Ensemble of Clustered Desparsified Lasso (ECDL)" that assesses parameters statistical significance and that can scale even when the number of predictors $p \geq 10^5$ is much higher than the number of samples $n \leq 10^3$, by leveraging structure among features. Our algorithm combines three main ingredients: a powerful inference procedure for linear models –the so-called Desparsified Lasso–, feature clustering and an ensembling step.

Mini-Bio :

Jérôme Alexis Chevalier est en deuxième année de thèse dirigé par Bertrand Thirion directeur de recherche de l'équipe pariétal de l'INRIA et Joseph Salmon professeur à l'Université de Montpellier (anciennement maître de conférence à Telecom ParisTech). Sa thèse s'intéresse au problème des statistiques en grande dimension avec applications en neuro-imagerie.