

High-dimension Statistics

Alexandre Tsybakov – ENSAE Paris

Course : 15 hours - TP : 9 hours

Objectives

Statistical science has undergone a profound transformation over the last decade with the development of large-scale statistical inference methods. This recent development has been driven by the need to deal with new data, such that for each individual there is a large number of observed variables, sometimes more than the number of individuals in the sample. Of course, not all variables are relevant and usually there are very few. The notion of parsimony (sparsity) is therefore fundamental to the statistical interpretation of large scale data. The aim of this course is to present some of the founding principles that emerge in this context. These principles are common to many problems that have emerged recently, such as large linear regression, estimation of large low-rank matrices, and network models, for example, stochastic block models. Emphasis will be placed on the construction of optimal methods in convergence speed and their oracle properties.

Outline

- Gaussian suite model. Sparsity and thresholding procedures.
- Large dimensional linear regression. BIC, Lasso, Dantzig selector, square root Lasso methods.
- Oracle properties and variable selection.
- Estimation of large low rank matrices. PCA Sparse.
- Network inference. Stochastic block model.

Bibliography

C.Giraud. Introduction to high-dimensional statistics. Chapman and Hall, 2015.

A.B.Tsybakov. Apprentissage statistique et estimation non-paramétrique. Polycopié de l'Ecole Polytechnique, 2014.

S.van de Geer. Estimation and testing under sparsity. Lecture Notes in Mathematics 2159. Springer, 2016.