

3MD3015 - Modélisation en Grande Dimension

Responsables : **Sarah LEMLER , Fragkiskos MALLIAROS**

Langues d'enseignement : **FRANCAIS**

Campus où le cours est proposé : **CAMPUS DE PARIS - SACLAY**

Nombre d'heures d'études élèves (HEE) : **40**

Nombre d'heures présentielles d'enseignement (HPE) : **24**

Année académique : **2024-2025**

Niveau avancé : **non**

Présentation, objectifs généraux du cours :

In this course, we will look at the issue of high dimensionality when the number of covariates (or explanatory variables) exceeds the number of observations.

We will show the limitations of standard procedures in this context. We present variable selection methods, highlighting their advantages and disadvantages from both a theoretical and a practical point of view. We will present regularization methods adapted to different problems. Finally, we will introduce screening methods to handle the case of ultra-high dimensions when regularization methods are insufficient. There will be practical exercises in R to put into practice the various concepts covered in the course.

Prérequis :

Good knowledge of statistics

Plan détaillé du cours (contenu) :

- Motivations
- Variable selection (Cp-Mallows, AIC, BIC...)
- Regularization methods (Ridge regression, Lasso and Lasso variants for linear and logistic regressions)
- Screening methods (in the case of ultra-high dimension)

Déroulement, organisation du cours :

Lectures, practical exercises in R, exam

Organisation de l'évaluation :

Project, written exam