

GIMAS9AD STATISTIQUE EN GRANDE DIMENSION

GIMAS9AD1 - Master IMSD - Mines Nancy

Statistique en grande dimension

Crédits : 2 ECTS

Durée : 21 heures

Semestre : S9

Responsable(s) :

Anne Gegout-Petit, professeur anne.gegout-petit@univ-lorraine.fr

Mots clés :

Data Mining, data science

Pré requis :

Statistical test theory, standard tests, regression

Objectif général :

Principales méthodes d'analyse de données et du Data Mining

Programmes et contenus :

Multiple testing issue, False Discovery Rate (FDR), usual method (Bonferroni, local FDR, Benjamini-Hochberg,...), case of correlated data

Penalised regression: LASSO, RIDGE, ELASTICNET

Decision trees and random forest, variable importance

Criteria of model selection: AIC, BIC, ...

Criteria of goodness of fit: RMSE, confusion table ROC curve

Variable selection: Cross validation, knockoffs, stability selection

Learning outcomes: Understand the need for a correction procedure in multiple testing, know how to choose and apply the usual methods in this case. Understand the need for penalization in the context of regression with a large number of variables and the associated optimization problem.

Targeted competencies: To be able to recognize a high dimensional statistical problem and to choose and/or adapt the usual methods of inference to this framework.

Compétences :

| Niveaux | Description et verbes opérationnels |
|-------------|-------------------------------------|
| Connaître | |
| Comprendre | |
| Appliquer | |
| Analyser | |
| Synthétiser | |
| Évaluer | |

Évaluations :

| | | | | |
|--|--|---|---------------------------------|---|
| <input checked="" type="checkbox"/> Test écrit | <input checked="" type="checkbox"/> Contrôle continu | <input type="checkbox"/> Oral, soutenance | <input type="checkbox"/> Projet | <input checked="" type="checkbox"/> Rapport |
|--|--|---|---------------------------------|---|