



## STATISTIQUE APPLIQUÉE AUX SCIENCES DE L'INGÉNIEUR

### STATISTICAL ENGINEERING

Responsable(s): Céline HARTWEG-HELBERT, Alexandre SAIDI

| Cours : 16.0 | TD : 0.0 | TP : 0.0 | Autonomie : 0.0 | BE : 12.0 | Projet : 0.0 | Langue du cours : FR

#### Objectifs de la formation

L'objectif de ce cours est de fournir les outils classiques de la statistique mathématique qui permettent d'aborder le choix du modèle probabiliste, son estimation et son évaluation. Nous nous intéresserons tout particulièrement au modèle linéaire et à ses extensions dans le contexte d'apprentissage statistique en grande dimension (LASSO, RIDGE, PCR PLS), au modèle logistique et aux modèles à base d'arbres (CART, RF, Boosting etc.). Le but de ce cours est aussi d'assurer une formation à la manipulation de données et à la mise en oeuvre pratique des modèles étudiés. Pour cela, une partie conséquente du cours est orienté vers la mise en oeuvre des différents modèles à l'aide du logiciel R à travers l'étude d'un grand nombre d'exemples.

**Mots-clés :** Régression linéaire et logistique. Sélection de modèles. Plans d'expériences. Régression pénalisée L1, L2. Modèles à base d'arbres.

#### Programme

- 1) Rappel sur la régression linéaire. Validités et limites de la méthode. Sélection de modèles.
- 2) Plans d'expériences : criblage et surface de réponses
- 3) Régression logistique
- 4) Apprentissage en grande dimension (régression LASSO, Ridge etc, modèles à base d'arbres)
- 5) Classification

#### ACTIVITES PRATIQUES

#### Compétences

- Savoir reconnaître différentes classes de problèmes de l'apprentissage statistique.
- Savoir mettre en oeuvre des modèles de base de l'apprentissage statistique et valider leur pertinence.
- Savoir proposer des méthodes d'apprentissage adaptées à la grande dimension
- Savoir utiliser le logiciel R.

#### Travail en autonomie

**Objectifs :** Cette activité n'est pas concernée par des activités d'autonomie cadrées en dehors du travail personnel.

**Méthodes :** Cette activité n'est pas concernée par des activités d'autonomie cadrées en dehors du travail personnel.

#### Bibliographie

G. Saporta, *PROBABILITÉS, ANALYSE DES DONNÉES ET STATISTIQUE*, Technip, 2020  
T Hastie, R Tibshirani, J Friedman *THE ELEMENTS OF STATISTICAL LEARNING: DATA MINING, INFERENCE, AND PREDICTION*, Springer, 2009

#### Contrôle des connaissances

Note finale = 60% Note de savoir + 40% Note de savoir-faire  
Note de savoir-faire = 100% contrôle continu  
Note de savoir = 100% examen terminal