MACHINE LEARNING : INTRODUCTION À L'APPRENTISSAGE STATISTIQUE

Cours de Mastère Spécialisé 2026-2027

Claude PETIT, Université de Rennes - Insee



École nationale
de la statistique
et de l'analyse
de l'information



Université
de Rennes

1 Déroulement du cours

2 Objectifs

3 Évaluation

4 Programme

5 Matériel pédagogique

Déroulement du cours

- 30h de cours (dont 12h de travaux pratiques).
- Octobre 2026 à janvier 2027.
- Un mélange de cours, TD & TP : 10 × 3h.
- Intervenant : claude.petit@univ-rennes.fr

Objectif du cours

- Une introduction aux grands principes de l'apprentissage supervisé et quelques aspects de l'apprentissage non-supervisé et semi-supervisé.
- Maîtriser les bases théoriques et savoir utiliser les algorithmes classiques de l'apprentissage supervisé (et quelques techniques d'apprentissage non supervisé).

- Prérequis:
  - Un bon cours de statistique mathématique (vous l'avez déjà eu).
  - Des connaissances de base en Python.

Évaluation

- Par projet informatique à effectuer par groupes de 3 ou 4.
- Sur une ou deux bases de données réelles.
- Plus de précisions d'ici quelques semaines.

Programme des 10 séances - 1/2

- 1. Introduction : statistique en grande dimension, malédiction de la dimension, dilemne biais variance, théorie de la décision: oracle, classification binaire, régression, minimisation du risque, surapprentissage et sous-apprentissage.
- 2. Méthodes à base de partitionnement : algorithme des plus proches voisins, arbres de décisions et algorithme CART.
- 3. TP1 : classification, régression, plus proches voisins, arbres de décision, validation croisée.
- 4. Méthodes à base de convexification: machines à vecteurs de support (SVM,SVR), méthodes à noyaux.
- 5. TP2 : SVM, SVR, méthodes à noyaux.

Programme des 10 séances 2/2

- 6. Méthodes à base d'agrégation: Boosting, Bagging, Stacking, forêts aléatoires.
- 7. TP3 : méthodes d'agrégation, Bootstrap.
- 8. Apprentissage non et semi-supervisé: CAH, clustering, réseau neuronal sur graphe (GNN).
- 9. Au choix : réseaux de neurones d'un point de vue statistique ou bien Machine Learning et économétrie.
- 10. TP4 : synthèse de toutes les méthodes. GNN. Algorithmes de descente de gradient.

Ce que l'on ne fera pas

- Les réseaux de neurones et le Deep Learning (vus dans un autre cours).
- Les régressions pénalisées LASSO, Ridge, etc. (vues dans un autre cours).
- Les méthodes semi-supervisées (sauf une brève introduction aux GNN).

Matériel pédagogique - 1/2

- Polycopié de cours, diapos, exercices, TP, journal de cours et matériel pédagogique sur cpmath.fr/ML.
- Des questions sur le cours : par mail (claude.petit@univ-rennes.fr).
- Bibliographie sommaire.

Matériel pédagogique - 2/2

- Bibliographie sommaire :

  - The Elements of Statistical Learning, de Hastie, Tibshirani, Friedmann, éditions Springer, 2008: https://www.sas.upenn.edu/~fdiebold/NoHesitations/BookAdvanced.pdf.
  - An introduction to Statistical Learning with applications in Python, de James, Witten, Hastie, Tibshirani, Taylor, chez Springer : https://www.statlearning.com/.
  - Introduction à l'apprentissage statistique, de Frédéric Sur : https://members.loria.fr/FSur/enseignement/apprauto/poly_apprauto_FSur.pdf.
  - Cours d'apprentissage et Data Mining d'Arnak Dalalyan (ENSAE) : https://adalalyan.github.io/cours.html.