

# 3MD4140 - Apprentissage statistique

Responsables : **Michel BARRET**

Langues d'enseignement : **FRANCAIS**

Campus où le cours est proposé : **CAMPUS DE METZ**

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

Nombre d'heures présentielles d'enseignement (HPE) : **23**

Année académique : **2024-2025**

Niveau avancé : **non**

---

## Présentation, objectifs généraux du cours :

L'apprentissage supervisé a pour objectif de proposer des méthodes qui, à partir d'une base d'exemples, permettent de prendre une décision portant sur un paramètre à partir d'observations, la décision devant être la meilleure possible en moyenne. Par exemple, classifier des images suivant leur contenu, c'est-à-dire décider si une image représente un chat, un chien, ou autre chose. Nous présenterons formellement le problème et étudierons les garanties de généralisation des algorithmes d'apprentissage supervisé, c'est-à-dire la qualité de prédiction de la sortie associée à une entrée non présente dans la base d'entraînement. Pour atteindre cet objectif, nous introduirons les concepts d'espace d'hypothèses ayant la capacité d'apprentissage PAC (probablement approximativement correcte), de dimension Vapnik-Chervonenkis d'un espace d'hypothèses. Nous énoncerons et démontrerons deux théorèmes fondamentaux de la théorie de l'apprentissage supervisé donnant une borne inférieure et une borne supérieure du risque réel au problème de classification binaire.

## Période(s) du cours (n° de séquence ou hors séquence) :

SM11

## Prérequis :

- cours de Probabilités de 1A (CIP-EDP, 1SL1000)
- cours de Statistique et apprentissage de ST4 (1CC5000)

## Plan détaillé du cours (contenu) :

- Formalisation du problème de l'apprentissage supervisé
- Capacité d'apprentissage PAC et convergence uniforme
- Le dilemme biais-complexité
- La dimension VC (Vapnik-Chervonenkis) d'un espace d'hypothèse
- Deux théorèmes fondamentaux de l'apprentissage supervisé

## Déroulement, organisation du cours :

10,5h de cours magistraux + 10,5h de travaux dirigés + examen écrit de 2h

## Organisation de l'évaluation :

examen écrit de 2h avec documents

## Moyens :

Les travaux dirigés, constitués d'exercices, permettront d'utiliser les concepts vus en cours.

## Acquis d'apprentissage visés dans le cours :

A l'issue de ce cours, les élèves devront être en mesure

- de comprendre et s'appropriier des éléments de la théorie de l'apprentissage supervisé ;
- de comprendre et s'appropriier le dilemme biais-complexité d'un espace d'hypothèses ;
- de comprendre et s'appropriier les bornes bayésiennes PAC de l'apprentissage supervisé (en particulier celles du problème de classification binaire).

## Description des compétences acquises à l'issue du cours :

C1.2: Select, use and develop modelling scales, allowing for appropriate simplifying hypotheses to be formulated and applied towards tackling a problem;

## Bibliographie :

S. Mallat, L'apprentissage face à la malédiction de la grande dimension, Cours du Collège de France, 2018.

S. Shalev-Shwartz et S. Ben-David, Understanding Machine Learning, from theory to algorithms, Cambridge University Press, 2014.

O. Catoni, Comment: Transductive PAC-Bayes Bounds Seen as a Generalization of Vapnik-Chervonenkis Bounds, Measures of Complexity, Festschrift for Alexey Chervonenkis, chap 10, Springer, 2015.