Intitulé du cours

Course title – Intitulé du cours	Foundations of Machine Learning
Level / Semester – Niveau /semestre	2
School – Composante	Ecole d'Economie de Toulouse
Teacher – Enseignant responsable	A. Daouia
Other teacher(s) – Autre(s) enseignant(s)	S Gadat
Other teacher(s) – Autre(s) enseignant(s)	
Other teacher(s) – Autre(s) enseignant(s)	
Other teacher(s) – Autre(s) enseignant(s)	
Other teacher(s) – Autre(s) enseignant(s)	
Lecture Hours – Volume Horaire CM	30
TA Hours – Volume horaire TD	15
TP Hours – Volume horaire TP	
Course Language – Langue du cours	English
TA and/or TP Language – Langue des TD et/ou TP	

Teaching staff contacts – Coordonnées de l'équipe pédagogique :

A. Daouia, S. Gadat

Course Objectives – Objectifs du cours :

The course is splitted into three parts.

- The first part of the course is dedicated to the theory of point estimation of a parameter in a parametric statistical model with an introduction to the Fisher information theory.
- The second part of the course is about the theory of interval estimation and testing theory.
- The third part is then related to the introduction of statistical learning theory.

The aim of the course is to give students theoretical tools to compare estimators and tests and have arguments to defend their choices. The course outline is the following:

1) Elements of information theory.
2) Point estimation (maximum likelihood, method of moments, asymptotic behavior, optimality, efficiency).
3) Interval estimation.
4) Testing theory: rejection region, first and second kind risk, level, power, empirical significance level.
5) Classical tests (about means, variances, proportions, independence)
6) Neyman theory: uniformly most powerful tests for a simple hypothesis.
7) Likelihood ratio tests, tests of a one-sided hypothesis in a monotone likelihood ratio model.
8) Introduction to statistical learning theory (binary classification, Bayes classifier, plug-in classifier: logistic regression, linear discriminant analysis, nearest neighbor)
9) Empirical Risk minimization: excess risk and concentration inequalities

Prerequisites – Pré requis :

Probability and Statistics for Data Science

Practical information about the sessions – Modalités pratiques de gestion du cours :

Personal laptops and tablets are accepted in the class

Grading system – Modalités d'évaluation :

Mid-term (Part 1) and final exam (Part 2). The final grade is the average mark of the two parts.

Bibliography/references – Bibliographie/références :

Mathematical Statistics, Jun Shao, Springer texts in Statistics, 1999.
Theory of Statistics, Mark Schervish, Springer series in Statistics,1995.
A course in mathematical statistics, G. Roussas, Academic Press, second edition, 1997.
Learning theory from first principles, F. Bach,
https://www.di.ens.fr/%7Efbach/ltfp_book.pdf
Introduction to Mathematical Statistics: Hogg, McKean, Craig, 8 edition, 2019

Session planning – Planification des séances

Distance learning – *Enseignement à distance :*

Distance learning can be provided when necessary by implementing:

- *Interactive virtual classrooms*

- *Recorded lectures (videos)*

- *MCQ tests and other online exercises / assignments*

- *Remote (online) tutorials (classes)*

- *Chatrooms*

En cas de nécessité, un enseignement à distance sera assuré en mobilisant:

- *Classe en ligne interactive*

- *Vidéo enregistrée de la présentation du matériel pédagogique*

- *QCM et exercices en ligne*

- *TP/TD à distance*

- *Forum...*