

Mathematics for data science

ECTS : 5

Volume horaire : 36

Description du contenu de l'enseignement :

Volume horaire :

CM : 18h

TD : 18h

Data science relies heavily on mathematical concepts from analysis, linear algebra and statistics. In this course, we will investigate the theoretical foundations of data science, through two axes. The first part of the course will focus on (convex) optimization problems. Optimization is indeed at the heart of the key advances in machine learning, as it provides a framework in which data science tasks can be modeled and solved. The second part of the course will be concerned with statistical tools for data science, that are instrumental in studying the underlying distribution of data. We will cover statistical estimation in connection with regression tasks, as well as concentration inequalities for random vectors and random matrices.

Compétence à acquérir :

- Identify and exploit convexity for sets, functions, and optimization problems.
- Derive optimality and duality results for convex optimization formulations.
- Analyze properties of statistical estimators according to the data at hand.
- Apply concentration inequalities to random vectors and matrices.

Bibliographie, lectures recommandées :

References:

S. Boyd et L. Vandenberghe, Convex optimization (2004)

M. Mahoney, J. C. Duchi, A. C. Gilbert (eds), The mathematics of data (2018)

J. A. Tropp, An introduction to matrix concentration inequalities (2015)