COM-406

Foundations of Data Science

Gastpar Michael, Urbanke Rüdiger

Cursus	Sem.	Type
Computational science and Engineering	MA1, MA3	Opt.
Computer and Communication Sciences		Obl.
Cybersecurity	MA1, MA3	Opt.
Data Science	MA1, MA3	Obl.
Data science minor	H	Opt.
Digital Humanities	MA1, MA3	Opt.

Language of teaching	English
Credits	6
Session	Winter
Semester	Fall
Exam	Written
Workload	180h
Weeks	14
Hours	**6 weekly**
Courses	4 weekly
Exercises	2 weekly
Number of positions	

Summary

We discuss a set of topics that are important for the understanding of modern data science but that are typically not taught in an introductory ML course. In particular we discuss fundamental ideas and techniques that come from probability, information theory as well as signal processing.

Content

This class presents basic concepts of Information Theory and Signal Processing and their relevance to emerging problems in Data Science and Machine Learning.
A tentative list of topics covered is:

1. Information Measures
2. Signal Representations
3. Detection and Estimation
4. Multi-arm Bandits
5. Distribution Estimation, Property Testing, and Property Estimation
6. Exponential Families
7. Compression and Dimensionality Reduction
8. Information Measures and Generalization Error

Keywords

Information Theory, Signal Processing, Statistical Signal Processing, Machine Learning, Data Science.

Learning Prerequisites

Required courses

COM-300 Modèles stochastiques pour les communications

Recommended courses

Statistics

Important concepts to start the course

Solid understanding of linear algebra and probability as well as real and complex analysis.

Learning Outcomes

By the end of the course, the student must be able to:

• Formulate the fundamental concepts of signal processing such as basis representations and sampling

- Formulate the fundamental concepts of information theory such as entropy and mutual information
- Analyze problems in statistical settings using fundamental bounds from information theory
- Formulate problems using robust and universal techniques

Teaching methods

Ex cathedra lectures, exercises, and small projects.

Expected student activities

Follow lectures; independent work on problems (homework and small projects).

Assessment methods

Written final exam during the exam session.
Homework Problem Sets during the semester.
10% homework, 30% midterm, 60% final exam; (if for some reason the course has to be given over zoom then we will skip the midterm and the course will be evaluated by 10% homework and 90% final)

Resources

Bibliography

Cover and Thomas, Elements of Information Theory (Second Edition), Wiley, 2006.

Ressources en bibliothèque

- Elements of Information Theory / Cover

Notes/Handbook

Lectures notes will be available on the course web page.