

# Foundation of Data Science



COUILLET  
Romain



NGUYEN  
Kim Thang

# Foundation of Data Science

- High dimension/massive data
- Data with distances (Similarity Search, Nearest Neighbor, Dimension Reduction)
- Generalization and Regularization
- Understanding Principal Components Analysis
- Sampling and Estimation
- Online Learning with Multiplicative Weights
- Advanced topics (if time permitted)
  - Interpretability and explainability
  - Fairness of machine learning algorithms

# Organization

- 7 x 3h courses + 4 x 3h lab (Python)
- Grading: project (40%) + final exam (60%)
- Pre-requisites:
  - Basics of Python (incl. numpy), basics of algebra, basics of probability