

# Introduction to Statistical Learning

Nicolas Vayatis (CMLA - ENS Cachan)



MVA 2016-2017

# Setup for the course

- About the mathematical theory of supervised learning and prediction
- High dimensional observations with label information
- Statistical model = classification data
- Goals of learning algorithms:
  - ▶ Binary classification
  - ▶ Multi-class classification
  - ▶ Preference learning
  - ▶ Bipartite scoring
  - ▶  $K$ -partite scoring
  - ▶ Ranking
  - ▶ Link prediction

with as many variants depending on various constraints...

# Learning with style!

- Performance - Do it better than parametrics!
- Approximation - Representation power
- Estimation - Robustness and stability
- Computation - Keep it simple!

# Course outline

- Chapter 1 - Modeling aspects:
  - ▶ Data, decision, risk, optimality
- Chapter 2 - Tools:
  - ▶ Concentration inequalities and complexity measures
- Chapter 3 - Theory:
  - ▶ Consistency and error bounds of mainstream learning algorithms
- Chapter 4 - Advanced topics:
  - ▶ Multiclass classification, ranking, link prediction, ...

- All sessions take place:
  - ▶ on Tuesdays morning 8:45am-10:45am,
  - ▶ at room Condorcet
  
- Practical sessions (TA: Cédric Malherbe):
  - ▶ October 25
  - ▶ November 8
  - ▶ November 29
  - ▶ December 13

# Important dates

- Mid-term exam (MTE): November 15 (mandatory)
- Sign-in for project validation: December 1st
- Final exam (FE) and project evaluation (PE): January 3

# Grading algorithm

- Presence in class required: attend at least 8 sessions  $\Leftrightarrow P = 1$
- Validation choice: if  $P=1$  then choose
  - ▶ Option 1: MTE+FE (recommended)
  - ▶ Option 2: MTE+PE
- Final grade (FG):

```
if Option=1 then
    FG = max(mean(MTE, FE), FE)
else
    FG = mean(MTE, PE)
end
```
- Re-take exam: under conditions  $P=1$  and  $FG > 0 \Rightarrow$  written exam or project

## (Most) related courses

- Convex optimization and applications in machine learning - A. D'ASPREMONT
- Reinforcement Learning - A. LAZARIC
- Graphs in machine learning - M. VALKO
- (Statistique en grande dimension et apprentissage - A. DALALYAN)
- Kernel methods for machine learning - J. MAIRAL, J.-P. VERT
- Prediction for individual sequences - V. PERCHET



# MLMDA@CMLA research group - Internships and more!

- MLMDA = Machine Learning and Massive Data Analysis
- Group leaders: Vianney Perchet and Nicolas Vayatis
- Research topics in Machine Learning Theory:
  - ▶ Sequential optimization and active learning
  - ▶ Problems with bandits, game theoretic approaches
  - ▶ Inference on graphs: epidemics, information cascades, control strategies
- Interdisciplinary research
  - ▶ Digitalized clinical evaluations in neurology, ENT, anesthesia, rehabilitation and training
  - ▶ Industrial data and signal analytics: energy, HPC, IoT, healthcare, ...
- Spin-off: ML Agora
- Join us!