Machine Learning II

Introduction

Souhaib Ben Taieb

University of Mons

Teaching staff

Souhaib BEN TAIEB (Instructor)

Instructor
Big Data and Machine Learning Lab
De Vinci Building, second floor, room 2.15
souhaib.bentaieb@umons.ac.be

Sukanya PATRA (TA)

PhD candidate
De Vinci Building, ground floor
sukanya.patra@umons.ac.be

Victor DHEUR (TA)

PhD candidate
De Vinci Building, ground floor
victor.dheur@umons.ac.be

S-INFO-075: Machine Learning II

- This course will be taught in English (lectures, labs, communications, emails, etc)
- Prerequisites
  - Machine learning I (S-INFO-256)
  - Probability and Statistics
  - Multivariate calculus
  - Linear algebra
  - Optimization
- **Course Webpage**
  - `https://github.com/bsouhaib/ML2-2023`
  - Lecture notes, project details, etc.
- **Moodle**
  - `https://moodle.umons.ac.be/course/view.php?id=2786`
  - Forum for asking questions, project submission, etc.

About the course

- **Objectives**
  - Learn advanced topics in machine learning
  - Learn how to do research/development in machine learning
- **Content**
  - (First few weeks) Standard lectures and labs
  - (Following weeks) Journal club with seminars given by **researchers** and **students**
    - Papers, online recorded lectures, book chapters, etc.
- **Seminar preparation and presentation**
  - Everyone read a selected machine learning paper
  - One person presents the paper
  - Everyone participate to the critical discussion

Project

- Read a selected machine learning paper
- Write a report (including experiments, and necessary proofs)
- Prepare a lecture, covering the necessary background and discussing the paper
- **More details to be announced later**

Assessment

- Oral exam (E) (*open book*): **60%**
- Project (P): **40%**
- Final mark:
  - If $E \geq 50\%$ and $P \geq 50\%$
    - Final mark $= E \times 0.6 + P \times 0.4$
  - Otherwise:
    - Final mark $= \min(E, P)$

Topics covered in Machine Learning I

- Introduction to machine learning (supervised, unsupervised, semi-supervised, ...)
- Supervised learning framework (components of learning, KNN, training and testing errors, model selection, cross-validation, optimal predictions, bias and variance tradeoff, ...)
- Linear regression (least squares, MLE, variable selection, nonlinear effects, ...)
- Linear classification (logistic regression, discriminant analysis)
- The bootstrap
- Tree-based methods (regression and classification trees, bagging, random forests, boosting)
- Dimension reduction and principal component analysis
- High-dimensional regression (ridge, lasso, ...)
- (Python: Pandas and Scikit-learn)