Machine Learning and Statistical Learning

Ewen Gallic
ewen.gallic@gmail.com

MASTER in Economics - Track EBDS - 2nd Year

Objectives

This course provides a broad introduction to **statistical learning** and **machine learning**.

The main objective is to provide students with the knowledge necessary to understand the basics of machine learning methods.

At the end of the course, students must be able to implement statistical analyses using machine learning techniques and understand their theoretical foundations.

Course outline (1/2)

- Week 1
  - Introduction
  - Optimization problem: gradient descent
- Week 2
  - Linear regression + Hands-on (overfitting) ♻
  - Quantile regression + Hands-on
  - Logistic regression + Hands-on (goodness-of fit for logistic regression) ♻

Course outline (2/2)

- Week 3
  - KNN + Hands-on
  - SVM + Hands-on
  - Dealing with images ♻
- Week 4
  - Working with imbalanced datasets
  - Explainability: Shap Values + Hands-on
  - Machine learning and ethics

References (books)

Berk, R. A. (2016). *Statistical Learning from a Regression Perspective*. Springer Texts in Statistics. doi:10.1007/978-3-319-44048-4

Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep learning*. Cambridge: MIT press.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2021). *An introduction to statistical learning, 2nd edition*. New York: springer. doi:10.1007/978-1-0716-1418-1_1

Murphy, K. (2012). *Machine Learning: A Probabilistic Perspective*. The MIT Press. ISBN: ISBN: 9780262018029

References (other)

- Dror, R. and Ng, A. (2018). *CS229: Machine Learning*. http://cs229.stanford.edu/
  - See also Shervine Amidi's cheatsheets: https://stanford.edu/~shervine/teaching/cs-229/cheatsheet-supervised-learning

Need help with R or Python?

For R:

- R for Data Science, Hadley Wickham and Garrett Grolemund, 2016
- Introduction à l'analyse d'enquêtes avec R et RStudio, Joseph Larmarange
- Notes de cours de R, Ewen Gallic, 2020
- R Notebooks, Ewen Gallic, 2022

For Python:

- Python Data Science Handbook, Jake VanderPlas, 2016
- Python for economists, Ewen Gallic, 2019 (French version also available)

Organization

- Sessions alternating theoretical presentations and applications.
- The applications will be carried out on computers, using Python or R (it is up to you).

This course is closely related to those of:

- **Pierre Michel**: more applied sessions as well as supplementary materials (Unsupervised learning, Tree-based methods, neural networks, ...).
- **Sullivan Hué**: dimension reduction, model selection, regularization.
- **Emmanuel Flachaire**: resampling methods, non-parametric methods

Grading

The final grade will be based on:

- Homework (1/3)
- Final project (2/3).

Regarding the final project, you will:

- work by **groups of two or three** (hence, not alone nor in groups of 4 our more)
- include an analysis based on **real data of your choice** (taken from Kaggle, for example) and will present the **theoretical elements** necessary to understand the analysis of the results.
- return a single project for this course and those of Sullivan Hué and Pierre Michel
  - each of us will evaluate the content in a different way

To contact me

- By e-mail:
  - ewen.gallic@univ-amu.fr
- Office:
  - Office 1-04
    Campus Marseille-Centre
    Site Ilot Bernard Du Bois
    5-9 boulevard Maurice Bourdet
    13205 Marseille cedex 1