

MASTER 1

MENTION ECONOMIE PARCOURS EXPERTISE ECONOMIQUE

BIG DATA AND MACHINE LEARNING / S2

ENSEIGNANTS :

Nom : Eric Guerci

Mail : Eric.GUERCI@univ-cotedazur.fr

OBJECTIVE OF THE COURSE:

Big data analysis is most commonly associated with "machine learning" techniques and algorithms: This course adopts an introductory but broad perspective on machine learning opportunities and solutions to successfully operate in the field of big data analysis for the economic and finance profession.

In this course students will learn about the core concepts in machine learning, as well as train the skills necessary to apply these methods widely and develop their programming abilities in the R language. Finally, they will familiarize themselves with the applied literature in the topic.

This is an introductory course, so the lectures and problem sets will be focused on the intuition and the mechanics behind machine learning concepts rather than the mathematical fundamentals. This course covers the mechanics underlying machine learning methods and discusses how these techniques can be leveraged by social scientists to gain new insight from their data. Specifically, the course will cover: decision trees, random forests, boosting, k-means clustering and nearest neighbours, support vector machines, kernels, neural networks, and ensemble learning. We will also discuss topics related to best practices, including error rates, cross-validation, and the use of bootstrapping methods to develop uncertainty estimates.

PRE-REQUISITE :

Attendance to introductory courses on calculus, probability theory and statistics. Students are expected to build R coding, thus R programming skills are also strongly recommended.

COURSE PLAN

This schedule is subject to change:

- Introduction. What is Machine Learning? A conceptual approach
 - Intro to machine learning.
 - Intro to R and the R software packages for machine learning
- Supervised Learning
 - Intro to Classification and Regression Analysis
 - Comparing classification Methods
 - Intro to regularization
 - Tree models
 - Bagging / Random forests
 - Neural networks
 - Support Vector Machine (L1/L2 regularization)
- Unsupervised Learning
 - Intro to Unsupervised Learning

- Principal Components
- Clustering

REFERENCES :

Reference books (some adopting R and some available for free online) :

- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, second edition. New York: Springer.
- Kuhn, Max and Kjell Johnson. 2013. *Applied Predictive Modeling*. New York: Springer.
https://vuquangnguyen2016.files.wordpress.com/2018/03/applied-predictive-modeling-max-kuhn-kjell-johnson_1518.pdf
- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning with Applications in R*. New York: Springer. <http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Seventh%20Printing.pdf>
- Laurent Gatto's [An Introduction to Machine Learning with R](#) online textbook
- Murphy, Kevin P. 2012. *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press.
- Berk, Richard A. 2016. *Statistical Learning from a Regression Perspective*, second edition. New York: Springer.

Some articles :

- Mullainathan, Susan and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31 (2): 87-106.
- Künzel, Sören R., Jasjeet S. Sekhon, Peter J. Bickel, and Bin Yu. 2018. "Meta-learners for Estimating Heterogeneous Treatment Effects using Machine Learning."
<https://arxiv.org/abs/1706.03461>

- Grimmer, Justin, Solomon Messing, and Sean J. Westwood. 2017. "*Estimating Heterogeneous Treatment Effects and the Effects of Heterogeneous Treatments with Ensemble Methods.*" *Political Analysis* 25 (4): 413-434.
- Sechidis, Konstantinos and Gavin Brown. 2018. "*Simple Strategies for Semi-supervised Feature Selection.*" *Machine Learning* 107 (2): 357395.
- Szegedy, Christian et al. 2013. "*Intriguing Properties of Neural Networks.*" <https://arxiv.org/abs/1312.6199>
- Varian, Hal R. 2014. "*Big Data: New Tricks for Econometrics.*" *Journal of Economic Perspectives* 28 (2):3–28.