Fundamentals of Machine Learning for Economists

July 26 – July 30, 2021

Program:

Day 1:

9.00—10:30:	Introduction & Key Principles of Machine/Statistical Learning
11:00—12:30	Learning vs. Fitting, Over-fitting, and regularization
14:00—15:30	Working with the data & datasets, the importance of feature engineering
16:00—17:30	office hours

Day 2:

9.00—10:30:	Cross-validation, model selection and combination
11:00—12:30	Lasso/Ridge/Elastic Net
14:00—15:30	Examples & hands-on, discussions
16:00—17:30	office hours

Day 3:

9.00—10:30:	Classification issues, Unbalanced Data
11:00—12:30	Regression and Classification Trees & Random Forests
14:00—15:30	Explainable Machine Learning, Examples & hands-on
16:00—17:30	office hours

Day 4:

9.00—10:30:	Neural Networks & Deep Learning
11:00—12:30	ML techniques in Economics – function approximation & solving models
14:00—15:30	Examples, hands-on, and the discussion
16:00—17:30	office hours

Day 5:

9.00—10:30:	Machine Learning and [Causal] Inference, Inference vs. Prediction
11:00—12:30	Unsupervised learning, kNN/k-Means
14:00—15:30	Reinforcement learning, Discussions

Course Goals

The course is an **introduction** to key ideas and principles in machine/statistical learning. Emphasis is on understanding principles and complementarity to an applied economist toolbox.

The concepts like over-fitting, regularization, importance of data transformations, and sample splitting will be explored and applied. The course will be lead primarily from a "frequentist statistics" point of view, with only some Bayesian concept mentioned. The material discussed is primarily focused on prediction, but recent advances in causal inference using ML techniques will be introduced too (e.g. "double machine learning", causal trees, …).

Course requirements:

Basic understanding of statistics, familiarity with computation software. Examples & demos will be provided in Matlab and Python. **Participants should have access to at least one of these platforms**.

Example & test dataset will be provided, to use in a platform of choice. References to procedures available in Stata will be provided. People should be able to process and manipulate data in the software, no need for advanced coding skills [but super helpful]. In applied work, 90 % of your time is data cleaning and manipulation.

Useful references:

Monographies:

Hastie, Tibshirani, Friedman: The Elements of Statistical Learning

Elements of Statistical Learning: data mining, inference, and prediction. 2nd Edition. (stanford.edu)

Murphy, K.: Machine Learning: A Probabilistic Perspective, pml-book | "Probabilistic Machine Learning" - a book series by Kevin Murphy (probml.github.io)

Efron, B. and T. Hastie: Computer Age Statistical Inference, Computer Age Statistical Inference: Algorithms, Evidence and Data Science (stanford.edu)

Hastie, Tibshirani, Wainwright: Statistical Learning with Sparsity – The Lasso and Generalizations

Statistical Learning with Sparsity: the Lasso and Generalizations (stanford.edu)