

MODULE SPECIFICATION • ELECTIVE

Machine Learning

The purpose of this course is to introduce students to machine learning, a relatively new approach to data analytics at the intersection between statistics, computer science and artificial intelligence. Students are taught how to master the theory and the techniques that allow turning information into knowledge and value by “letting the data speak”. The teaching approach is based on graphical language and intuition more than on algebra. The course makes use of instructional as well as real-world examples, and balances theory evenly with practical sessions using the software Stata. There are no prerequisite modules required before taking this course.

10	15h	85h	100h
CREDITS	CONTACT HOURS	PRIVATE STUDY	TOTAL STUDY

COURSE CONVENOR

NAME	Prof Sébastien Laurent
OFFICE	By appointment — please email to arrange a convenient time

COURSE STRUCTURE

Econ 7914 is a 10-credit course and students are expected to input approximately 100 hours of study. The total number of contact hours is 15 hours, leaving 85 hours for private study. Delivery is by lectures — 15 hours across the first three weeks of term, comprising 10 hours of lectures and 5 hours of tutorials. Optional clinics run on the last day of the course.	During private study, balance the course material (the primary source) against the recommended textbooks: think critically about how these fit into the body of knowledge, perform the exercises, complete coursework and revise. You can expect to perform well only by working consistently through the year.
--	---

MODULE PROGRAMME

LECT.	TOPIC	SESSION CONTENT
01	The Basics of Machine Learning I	- Machine learning: definition, rationale, usefulness - Supervised vs. unsupervised learning - Regression vs. classification problems - Inference vs. prediction - Sampling vs. specification error - Coping with the fundamental non-identifiability of $E(y x)$
02	The Basics of Machine Learning II	- Measuring the quality of fit: in-sample vs. out-of-sample prediction power - Goodness-of-fit indices - The bias-variance trade-off and Mean Square Error (MSE) minimization
03	Simulation, Resampling and Validation Methods I	- Monte Carlo simulations - Logic and functioning of a Monte Carlo experiment - Implementing Monte Carlo experiments
04	Simulation, Resampling and Validation Methods II	- Bootstrap - The logic of the bootstrap - Bootstrapping standard errors - Cross-validation - The validation set approach - Leave-One-Out Cross-Validation - K-fold cross-validation - The Stata package crossfold
05	Non-parametric regression: local methods I	Beyond parametric models: the “why” and the “how”
06	Non-parametric regression: local methods II	Types of non-parametric regression: local vs global approaches
07	Non-parametric regression: local methods III	- Nearest-neighbor regression - Kernel-based regression - The Stata npregress command
08	Non-parametric regression: global methods I	- Polynomial and series regression - Spline regression
09	Non-parametric regression: global methods II	Generalized additive models
10	Model Selection and Regularization I	- Subset selection: optimal, forward and backward subset selection - Shrinkage methods: Lasso, Ridge and Elastic regression
11	Tree-Based Regression I	An introduction to regression trees
12	Applications using Stata II	Bagging, random forests and boosting

KEY SKILLS

By the end of this course, students should have some knowledge and understanding of:

- Implement factor-importance detection
- Perform signal-from-noise extraction
- Evaluate correct model specification
- Understand model-free classification, both from a datamining and a causal perspective

DESIRED OUTCOMES

By the end of this course, students should be able to:

- Engage in abstract thinking by extracting the essential features of complex systems to facilitate problem solving and decision-making
- Communicate and present complex arguments in oral and written form with clarity and succinctness
- Present, interpret and analyse information in numerical form
- Utilise effectively statistical and other packages
- Apply basic statistical techniques to analyse economic and financial datasets

ASSESSMENT

WEIGHTING	100% coursework
ISSUED	End of week 6
DUE	End of week 10 — four weeks for completion
SUBMISSION	Electronically via the Moodle site for this course
FEEDBACK	Returned within 4 weeks of the deadline, as marker's notes appended to the PDF of your coursework

Marking criteria and penalties are set out in the Economics Undergraduate Handbook and the general course information paper, available via the Current Student page of the university website under Academic Regulations.

COURSE TEXT & RECOMMENDED READING

Copies of the lecture slides are made available on the course web pages, together with additional material and course announcements.

- Cerulli, G. (2023). Fundamentals of Supervised Machine Learning: With Applications in Python, R, and Stata. Springer, series "Statistics and Computing"
- James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. Springer
- Hastie, T., Tibshirani, R. and Friedman, J. (2008). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edition. Springer