Lancaster University Management School

timberlake
statistics | econometrics | forecasting

Econ 614 Machine Learning 2023-2024

COURSE AIMS & OBJECTIVES, KEY SKILLS, AND LEARNING OUTCOMES

Course Aims & Objectives: The aim of this course is to introduce students to machine learning, which is a relatively new approach to data analytics at the intersection between statistics, computer science, and artificial intelligence. Students will be 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 will be based on the graphical language and intuition more than on algebra. The course will make use of instructional as well as real-world examples, and will balance evenly theory and practical sessions with the software Stata.

There are no prerequisite modules required before taking this course.
The course will make use of the software Stata.

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

- Implement factor-importance detection;
- Perform signal-from-noise extraction;
- Evaluate correct model specification;
- Understand model-free classification, both from a data-mining 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
- Work effectively both individually and within a team environment.

COURSE STRUCTURE

Econ 614 is a 10 credits course and therefore students are expected to input approximately 100 hours of study into the course. The total number of contact hours in the form of Lectures with 15 hours delivered over the first 3 weeks of the term (10 hours of lectures and 5 hours of tutorials). There will be optional clinics on the last day of the course.
During your private study you should strike a balance between reading the course material (which is the primary source of information) and the recommended textbooks, thinking critically about how these fit in to the body of knowledge on the subject and about how our level of knowledge can be improved, performing exercises, completing coursework and revising for examinations. You can expect to perform well on this course only if you work consistently through the year.
Econ 614 is 15 hours. This leaves 85 hours for private study. Course Delivery comes in the form of Lectures with 15 hours delivered over the first 3 weeks of the term (10 hours of lectures and 5 hours of tutorials). There will be optional clinics on the last day of the course.

During your private study you should strike a balance between reading the course material (which is the primary source of information) and the recommended textbooks, thinking critically about how these fit in to the body of knowledge on the subject and about how our level of knowledge can be improved, performing exercises, completing

coursework and revising for examinations. You can expect to perform well on this course only if you work consistently through the year.

COURSE CONVENOR

Dr Giovanni Cerulli

LECTURERS CONTACT INFORMATION (Including Office Hours)

Email: giovanni.cerulli@ircres.cnr.it
Available by appointment (please email to arrange a convenient time)

COURSEWORK ASSESSMENT

The CWA mark will be calculated as 100% coursework. The coursework will be assigned at the end of the course.
The coursework will be delivered to students at the end of week 6 (31st May) and is due for submission at the end of week 10 (28th June), allowing students 4 weeks for completion.
Coursework must be submitted electronically through the Moodle site for this course: https://mle.lancs.ac.uk/course. Login using your regular Lancaster University access details. This opens a page headed MLE: My home.

The format of the submission is as follows.

- The submitted file must be in pdf format with the following name
  stud#_studname_cw_**cw#.pdf**
  where: **stud#** is your student number, **studname** is your name in the format *surname_name*, **cw#** is either 1 or 2 according to the piece of coursework submitted. Eg a student with student number 111 would submit a file named *111_surname_firstname_cw_1.pdf*.

- Maximum file size is 2MB: figures resolution must be adjusted accordingly.
-

Note that your work will be screened using software designed to detect plagiarism.
Do not rely upon someone else to submit your coursework.
*Word counts are inclusive of all material submitted apart from the Bibliography.

FEEDBACK ON COURSEWORK

The coursework will be marked and returned to students within 4 weeks of the submission deadline. Feedback will consist of marker's notes appended to the pdf of your coursework.

MARKING CRITERIA AND PENALTIES

Marking criteria can be found in the Economics Undergraduate Handbook and the general course information paper. An electronic copy of this can be found via the Current Student page of the university website then follow the Academic Regulations link
https://gap.lancs.ac.uk/ASQ/QAE/MARP/Documents/UG-Assess-Regs.pdf

FINAL MARK INFORMATION

This course is assessed 100% by means of coursework. The final mark is the average of the marks obtained in the two pieces of coursework.

COURSE TEXT AND RECOMMENDED READING

Main texts

The main recommended textbook are

- Cerulli Giovanni (2023), *Fundamentals of Supervised Machine Learning: With*

Applications in Python, R, and Stata, Springer, 2023. Series: "Statistics and Computing".

- Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani (2013), *An Introduction to Statistical Learning with Applications in R*, Springer, New York, 2013. ISBN # 978-1-4614-7137-0.
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman (2008), The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second edition, Springer.

Note Copies of the lecture slides will be made available on the course web pages. You **MUST** print off the notes for each lecture **prior to** attending. Solutions to exercises, and some additional material associated with these lectures and course announcements will also be placed on this website.

COURSE OUTLINE/LECTURE SCHEDULE

Lecture 1: The Basics of Machine Learning I (1 hours):
· Machine Learning: definition, rational, usefulness
  o Supervised vs. unsupervised learning
  o Regression vs. classification problems
  o Inference vs. prediction
  o Sampling vs. specification error

Lecture 2: The Basics of Machine Learning II (1 hours):
· Coping with the fundamental non-identifiability of E(y|x):
  o Parametric vs. non-parametric models
  o The trade-off between prediction accuracy and model interpretability
· Goodness-of-fit measures
  o Measuring the quality of fit: in-sample vs. out-of-sample prediction power
  o Goodness-of-fit indices
  o The bias-variance trade-off and the Mean Square Error (MSE) minimization

Lecture 3: Simulation, Resampling and Validation Methods I (1 hours):
· Monte Carlo simulations
  o Logic and functioning of a Monte Carlo experiment
  o Implementing Monte Carlo experiments
· Bootstrap
  o The logic of the Bootstrap
  o Bootstrapping standard errors

Lecture 4: Simulation, Resampling and Validation Methods II (1 hours):
· Cross-Validation
· The validation set approach
· Leave-One-Out Cross-Validation
· K-fold cross-validation
· The Stata package crossfold

Lecture 5: Non-parametric regression: local methods I (1 hours):
· Beyond parametric models: the “why” and the “how”

Lecture 6: Non-parametric regression: local methods II (1 hours):
· Type of non-parametric regressions: local vs global approaches
  o Nearest-neighbor regression
  o Kernel-based regression
  o The Stata npregress command

Lecture 7: Non-parametric regression: global methods I (1 hours):
· Polynomial and series regression

· Spline regression

Lecture 8: Non-parametric regression: global methods II (1 hours):

· Generalized additive models

Lecture 9: Model Selection and Regularization I (1 hours):

· Subset Selection: Optimal, Foreword and Backward subset selection

Lecture 10: Model Selection and Regularization II (0.5 hours):

· Shrinkage Methods: Lasso, Ridge, and Elastic regression

Lecture 11: Tree-Based Regression I (0.5 hours):

· An introduction to Regression Trees

Lecture 12: Applications using Stata II (0.5 hours):

· Bagging, Random Forests, and Boosting