

ECON3389 Machine Learning in Economics

Fall 2024

Instructor: Alberto Cappello

Contact Information: cappelae@bc.edu

Office Hours: Wednesday 3pm - 4pm. Send me an email if you want to meet at other times. Wednesdays are preferred.

Office Hours location: Room 330 Maloney Hall 3rd Floor (Economics department)

Grading TA: Trevor Petrin (petrint@bc.edu)

Course Overview

One of the defining features of the world around us today is the ever-increasing amounts of data that describe our daily lives. This "big data" phenomenon, as it became known, has led to developing of new methods, called "machine learning", that allow high-dimensional statistical analysis in ways that were either impossible or infeasible for classical statistical methods, such as regression analysis.

The goal of this course is to provide students with an introduction to modern data-driven learning in a framework that makes it applicable to causal economic analysis. While we will cover the necessary theoretical foundations, the emphasis will be placed on application and learning how and when to use these methods, as well as when and how these methods can fail.

The coursework will include homework assignments with simulated and real-world data, a group project in form of a case study and mid-term and final exams. We will use R as our main data analysis software. Prior coding experience is welcomed, but not required — everything you need to know about R will be taught in the video lectures.

Course Goals

- Understand common methods of modern data-driven learning in Economics.
- Apply these methods to real life data using R. Please spend considerable time atleast at the beginning of the course to learn the basics of R coding.
- This course will be intuitively rigorous about the methods focusing on when and why these methods will be applicable. Focus will not be on the mathematical aspect of the methods.
- Develop a mindset that helps with job interviews related to data science positions.

Prerequisites

For students majoring in Economics, Economic Statistics (ECON1151) is a required prerequisite and Econometric Methods (ECON2228) is a recommended prerequisite. Students coming from different departments should have similar command of statistical methods. A solid knowledge of differential calculus at the level of MATH1102 (the "preferred" co-requisite for ECON 2228) is highly recommended. Prior knowledge of programming (Stata, R, Python, Matlab) will be helpful, but is not a prerequisite, everything students need to know about coding in R will be taught in the course.

Textbooks

- **Main textbook (required):**

1. An Introduction to Statistical Learning, with Applications in R. James, G., Witten, D., Hastie, T., and Tibshirani, R Second Edition, **available for free** at: <https://www.statlearning.com/>

- Additional references (recommended)

1. Machine learning with R Cookbook. Chiu, Yu-Wei. (2015).
2. Data Mining and Business Analytics with R. Ledolter, J. (2013)
3. A Gentle Introduction to Effective Computing in Quantitative Research. Paarsch, H.J., Golyaev, K. (2016)
4. Fundamentals of Machine Learning for Predictive Data Analytics. Kelleher, J, Mac Namee, B. and D'Arcy, A. (2015)
5. Learning from Data: A Short Course. Abu-Mostafa, Y. , Magdon-Ismail, M. and Lin, H. (2012)

Required software

The primary software environment is the R statistical programming language, which can be downloaded for free from <http://www.r-project.org>. RStudio is the recommended interface for the R statistical programming language software, which can also be downloaded for free at <http://www.rstudio.org>.

Canvas

Canvas is the Learning Management System (LMS) at Boston College, designed to help faculty and students share ideas, collaborate on assignments, discuss course readings and materials, submit assignments, and much more - all online. As a Boston College student, you should familiarize yourself with this important tool. All class materials will be available online via Canvas.

Course components

Your course grade will be determined using the following components

- Homework 40%
There will be 5-6 assignments involving coding in R and theoretical exercises
- Mid-term Exam 15%
Will include material covered till Spring break
- Final Exam 30%
Will include all the material covered in the lectures. More weight on the materials covered post Spring break
- Case study 15%
This will be our opportunity to learn methods that are not covered in the course and their real life applications. Groups of 3 students will be assigned ML topics. You will need to submit a roughly 7-10 page write-up about the method and the main applications of the method and present a 12-15 minute presentation on the topic. Grading will be based on the writeup and questions answered during the presentation.

Course outline and schedule

Week	Topics	Text chapter
Module 1 (Aug 26)	Introduction	1 and 2
Module 2 (Aug 28 - Sept 11)	Regression Analysis	3
Module 3 (Sept 13 - Sept 25)	Classification methods	4
Module 4 (Sept 27 - Sept30, Oct 4 - Oct 11)	Resampling methods	5
Module 5 (Oct 16 - Oct 28)	Feature Selection methods	6
Module 6 (Oct 30 - Nov 13)	Decision Trees	8
Discussion (Nov 15 - Nov 18)	Revision/NLP	
Nov 20 - Nov 25, Dec 4 - Dec 6	Class Presentations	
Oct 2	MID-TERM EXAM	
Dec 9	FINAL EXAM	

Case Study Topics

- Methods
 1. Neural Networks
 2. Large Language Models and ChatGPT
 3. Recommender systems (The Netflix Prize competition)

4. Principal Component Analysis
 5. Elastic net regularization
 6. Clustering Methods: K means and hierarchical
- Applications
 1. Algorithmic Pricing
<https://www.brookings.edu/articles/are-online-prices-higher-because-of-pricing/>
<https://hbr.org/2021/09/the-pitfalls-of-pricing-algorithms#:~:text=The%20Problem,loyalty%2C%20and%20damage%20brand%20reputation> <https://www.wired.com/2011/04/amazon-flies-24-million/>
 2. Monte Carlo Simulation
<https://tech.instacart.com/no-order-left-behind-no-shopper-left-idle-24ba0600f>
 3. Quantile regression
<https://tech.instacart.com/how-instacart-delivers-on-time-using-quantile-regres>
 4. Text Analysis and Sentiment Analysis

Deadlines, Late work and make-up policy

All deadlines are strictly enforced. Late work is not accepted and no credit will be earned on late work unless the student has arranged an extension **ahead of time** with me (and that is quite possible, I am flexible with everyone's challenging circumstance and time constraints), with rare exceptions based on individual circumstances (e.g. inability to communicate with me ahead of time because of an emergency).

Please note that hardware ("WiFi was down") and software ("Canvas was glitching") issues are not considered valid excuses for late submissions.

Accommodation and Accessibility

Boston College is committed to providing accommodations to students, faculty, staff and visitors with disabilities. Specific documentation from the appropriate office is required for students seeking accommodation in this course. Advanced notice and formal registration with the appropriate office is required to facilitate this process. There are two separate offices at BC that coordinate services for students with disabilities

- The Connors Family Learning Center (CFLC) coordinates services for students with LD and ADHD
- The Disabilities Services Office (DSO) coordinates services for all other disabilities.

Find out more about BC's commitment to accessibility at www.bc.edu/sites/accessibility

Academic Integrity

Boston College values the academic integrity of its students and faculty. It is your responsibility to familiarize yourself with the university's policy on academic integrity. If you have any questions, always consult your professor. Violations of academic integrity will be reported to your class dean and judged by the academic integrity committee in your school. If you are found responsible for violating the policy, penalties may include a failing grade as well as possible probation, suspension, or expulsion, depending on the seriousness and circumstances of the violation. See [link](#) for a full discussion of the university's policies and procedures regarding academic integrity