Introduction to Machine Learning for Economists

Fall 2023

Course Organization

Jérémy Do Nascimento Miguel
Email: jdnmiguel@u-bordeaux.fr
Office: H-217.
Office Hours: by appointment (send an email).
Course website: see my Github page.

Course Description

This course aims to cover the fundamental concepts of machine learning in economics and their practical applications. It aims to provide an introduction to basic and key principles of machine learning, as well as potential utilization for economists. This course will not cover any details of algorithms and computational issues – there are many specialised courses from Computer Science that go into these details.

Assignment

There will be a problem set due at the end of the semester.

Objective

The primary goal of this class is that students will have a first glimpse and hands-on knowledge of basic machine learning methods used in economics.

Students can expect to gain familiarity with open source software used in data analysis:

- Anaconda (Python and Jupyter)
- Google Earth Engine (if we have time)

Software

- Python. Check my GitHub page to set up python on your device **before** the first class.

Prerequisites/Corequisites

Basic knowledge in statistics, econometrics, and data analysis (i.e., python). For those without much experience, I strongly suggest reviewing some fundamentals of quantitative analysis at these free resources:

- Khan Academy course on Python (video 1 to 6)

Materials

The main class material (slides, data, and code) will be posted on my Github page. I am relying heavily on the James, Witten, Hastie, Tibshirani, and Taylor (2023)'s textbook for explaining the key concepts. You can find the textbook directly on their website.

Class coverage

Introduction: Class overview and technical assistance

- How economists can use machine learning? (Athey and Imbens, 2019; Mullainathan and Spiess, 2017)
- Big data (Varian, 2014)

Part 1: Introduction to Machine Learning

- Statistical learning, model fit and bias-variance trade-off (Chapter 2)
- Cross-validation (Chapter 5.1)
- Regularized regression (Chapter 6.2) and tree-based method (Chapter 8)

Part 2: Average treatment effects with *too many* control variables

- Double selection estimator (Belloni, Chernozhukov, and Hansen, 2014b)
- Lasso model (Belloni, Chernozhukov, and Hansen, 2014a)
- Double debiased estimators (Chernozhukov et al., 2017)

Part 3: Estimating heterogeneous treatment effect

- Honest estimation
- Causal trees and forests (Athey and Imbens, 2016; Wager and Athey, 2018). See Davis and Heller (2017) for a friendly and non-technical application.
- Generic ML inference (Chernozhukov et al., 2018). [Presentation here].

Part 4: Other applications

- Remote sensing (Donaldson and Storeygard, 2016)
- Text analysis (Grimmer, Roberts, and Stewart, 2022)

Some references

Main reference

Gareth, James, Witten Daniela, Robert Tibshirani Trevor Hastie, and Taylor J. (2023). *An introduction to statistical learning: with applications in Python*. Spinger.

Causal inference and machine learning

Athey, Susan and Guido Imbens (2016). "Recursive partitioning for heterogeneous causal effects". In: *Proceedings of the National Academy of Sciences* 113.27, pp. 7353–7360.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2014a). "High-dimensional methods and inference on structural and treatment effects". In: *Journal of Economic Perspectives* 28.2, pp. 29–50.

— (2014b). "Inference on treatment effects after selection among high-dimensional controls". In: *Review of Economic Studies* 81.2, pp. 608–650.

Brand, Jennie E., Xiang Zhou, and Yu Xie (2023). "Recent Developments in Causal Inference and Machine Learning". In: *Annual Review of Sociology* 49.1, null.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, and Whitney Newey (2017). "Double/debiased/neyman machine learning of treatment effects". In: *American Economic Review* 107.5, pp. 261–265.

Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Ivan Fernandez-Val (2018). *Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India*. Tech. rep. National Bureau of Economic Research.

Davis, Jonathan MV and Sara B Heller (2017). "Using causal forests to predict treatment heterogeneity: An application to summer jobs". In: *American Economic Review* 107.5, pp. 546–550.

Wager, Stefan and Susan Athey (2018). "Estimation and inference of heterogeneous treatment effects using random forests". In: *Journal of the American Statistical Association* 113.523, pp. 1228–1242.

To go further

Athey, Susan (2018). "The impact of machine learning on economics". In: *The economics of artificial intelligence: An agenda*. University of Chicago Press, pp. 507–547.

Athey, Susan and Guido W Imbens (2019). "Machine learning methods that economists should know about". In: *Annual Review of Economics* 11, pp. 685–725.

Hastie, Trevor, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer.

Mullainathan, Sendhil and Jann Spiess (2017). "Machine learning: an applied econometric approach". In: *Journal of Economic Perspectives* 31.2, pp. 87–106.

Varian, Hal R (2014). "Big data: New tricks for econometrics". In: *Journal of economic perspectives* 28.2, pp. 3–28.

If you want to go further, check Susan Athey's lecture about machine learning and causal inference.

Empirical applications

Aiken, Emily, Suzanne Bellue, Dean Karlan, Chris Udry, and Joshua E Blumenstock (2022). "Machine learning and phone data can improve targeting of humanitarian aid". In: *Nature* 603.7903, pp. 864–870.

Becker, Sascha O, Thiemo Fetzer, and Dennis Novy (2017). "Who voted for Brexit? A comprehensive district-level analysis". In: *Economic Policy* 32.92, pp. 601–650.

Blumenstock, Joshua, Gabriel Cadamuro, and Robert On (2015). "Predicting poverty and wealth from mobile phone metadata". In: *Science* 350.6264, pp. 1073–1076.

Ghoddusi, Hamed, Germán G Creamer, and Nima Rafizadeh (2019). "Machine learning in energy economics and finance: A review". In: *Energy Economics* 81, pp. 709–727.

Longuet Marx, Nicolas (2023). "Party Lines or Voter Preferences? Explaining Political Realignment". In: *Working Paper*.

Storm, Hugo, Kathy Baylis, and Thomas Heckelei (2020). "Machine learning in agricultural and applied economics". In: *European Review of Agricultural Economics* 47.3, pp. 849–892.

Remote sensing

Donaldson, Dave and Adam Storeygard (2016). "The view from above: Applications of satellite data in economics". In: *Journal of Economic Perspectives* 30.4, pp. 171–198.

Harari, Mariaflavia (2020). "Cities in bad shape: Urban geometry in India". In: *American Economic Review* 110.8, pp. 2377–2421.

Text analysis

Grimmer, Justin, Margaret E Roberts, and Brandon M Stewart (2022). *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press.