

RTG 2484 Mini-Course

“Machine Learning for Economists and Business Analysts”

Standard machine learning methods are powerful prediction tools, but they cannot be deployed for causal inference without putting additional structure on the estimation problem. This course provides a practical introduction to predictive and causal machine learning. In the first part of the course, we focus on supervised machine learning methods. We discuss penalized regression and tree-based methods. In the second part of the course, we discuss how machine learning can be combined with causal inference. In particular, we focus on methods that allow to control for high-dimensional confounders in a data-driven way (double selection procedure, T-learning, double machine learning). We solve real-world economic and business problems in practical R session, which are integral part of the course.

Learning objectives/competences: 1) Participants can distinguish between questions that can be answered with predictive and causal methods. 2) Participants will be familiar with the principles of prediction. 3) Participants can deploy predictive machine learning methods to economic and business problems. 4) Participant can deploy causal machine learning methods to account for confounders.

The entire course material will be provided on Github:

<https://github.com/AStrittmatter/RTG-2484-Mini-Course>

Prerequisites:

Basic knowledge of econometrics of statistics (keywords: multivariate regression, R-squared, MSE, IPW)

Schedule:

Wednesday, June 23:

9:15-10:45 Introduction: Principals of Prediction

11:00-12:30 Penalized Regression Methods

14:15-15:45 PC Lab: Penalized Regression Methods

Thursday, June 24:

9:15-10:45 Trees and Random Forests

11:00-12:30 PC Lab: Trees and Random Forests

14:15-15:45 Post Double Selection Procedure

Friday, June 25:

9:15-10:45 PC Lab: Post Double Selection Procedure

11:00-12:30 Double Machine Learning

14:15-15:45 PC Lab: Double Machine Learning

References:

- Athey (2017): “Beyond Prediction: Using Big Data for Policy Problems”, Science, 355 (6324), pp. 483-485, [download](#).
- Belloni, Chernozhukov, and Hansen (2014): ”High-Dimensional Methods and Inference on Structural and Treatment Effects”, Journal of Economic Perspectives, 28 (2), pp. 29-50, [download](#).
- Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, and Newey (2017): ”Double/Debiased/Neyman Machine Learning of Treatment Effects”, American Economic Review, 107 (5), pp. 261-265, [download](#).
- James, Witten, Hastie, and Tibshirani (2013): ”An Introduction to Statistical Learning”,

- Springer, Chapter 6.2, [download](#).
- Mullainathan and Spiess (2017): “Machine Learning: An Applied Econometric Approach”, Journal of Economic Perspectives, 31 (2), pp. 87-106, [download](#).

Software:

This is a BYOD-course (bring-your-own-device). We will use the open-source software R during the PC lab sessions. We will work on a server with interactive Jupyter Notebooks. Usually it is not necessary to install any software to participate in the PC labs. However, for the case of bad internet connectivity it would be good if all participants install R and RStudio on their own device before the course (free-versions are sufficient).

Participants can use the following link to join the PC lab sessions:

<https://mybinder.org/v2/gh/AStrittmatter/RTG-2484-Mini-Course/HEAD>