HCHE PhD Course

Supervised and Causal Machine Learning

Block course: from May 31st until June 4th 2021
(Online via Zoom)

Course Instructor: Prof. Michael C. Knaus, Assistant Professor of Econometrics at the University of St. Gallen

Course Value: 2 SWS or 5 LP

Course Objectives:
Participants of this course will learn and apply recent Supervised and Causal Machine Learning methods to analyse effects of either experimental or observational interventions. This course will focus on tools that are already mature in the sense that they are easy to implement for practitioners and covers three major topics:

1. Estimation of heterogeneous effects for experimental data
2. Estimation of average and heterogeneous effects for observational data

This course is aimed at PhD students in health economics, business administration, econometrics and computer science. Participants are expected to have a solid understanding of probability theory (conditional expectations) and regression analysis (OLS) as well as causal research designs. First experience with supervise ML methods is useful but not required.

Student evaluation: Students prepare a 5 to 10 page analysis of a real experiment of their choice or with simulated data.

Teaching language: English

Registration: Via e-mail to anmeldung@hche.de (According to availability)
Overview:

- Date: 31st May – 04th June 2021
- Time: 9:00-12:00
- Place: Online via Zoom [https://unisg.zoom.us/j/4608829567](https://unisg.zoom.us/j/4608829567)
- Instructor: Prof. Michael C. Knaus, Assistant Professor of Econometrics at the Swiss Institute for Empirical Economic Research of the University, St. Gallen
- Teaching language: English
- Credit Points: 2 SWS/5 LP (Universität Hamburg)
- Registration for PhD Course: anmeldung@hche.de
Introduction

Participants of this course will learn popular concepts and methods in Causal Machine Learning methods to analyse effects of either experimental or observational policy interventions. Causal Machine Learning combines two mature fields in data analytics. On the one hand, the field of Machine Learning (ML) advanced our ability to detect correlational pattern in data, which is important to form high-quality predictions. On the other hand, the field of Causal Inference advanced our knowledge about how to assess the effects of interventions, which is essential for high-quality decision making. The promise of Causal Machine Learning is to deliver the best of both worlds to draw (more) reliable and more informative causal inference.

This course will focus on tools that are already mature in the sense that they are available for implementation for practitioners in the software R and covers three major topics:

1. Introduction/recap of supervised ML with a focus on methods that are important ingredients of Causal ML
2. Estimation of average effects in the presence of confounding
3. Estimation of heterogeneous effects in experimental and observational settings

Course Prerequisites:

- Solid understanding of probability theory (conditional expectations) and regression analysis (OLS)
- Solid understanding of causal research designs, in particular randomized experiments and observational designs that control for confounding factors (selection-on-observables)
- (not required, but big advantage) First experience with supervise ML methods, in particular shrinkage methods (e.g., Lasso, Ridge) and tree-based methods (regression trees, random forest)

Target Group:

Participants will find the course useful if:

- They are familiar with causal inference and regression analysis and are curious how machine learning methods could enter their empirical toolbox.
- They work with experimental and/or observational data in economics or related fields.

Course and Learning Objectives:

By the end of the course participants will:

- Understand popular methods that are likely to appear in future studies they consume.
- Know in which settings and for which research questions the current state of Causal Machine Learning provides attractive alternatives to standard tools.
- Be able to apply Causal Machine Learning in basic settings.
- Have the background knowledge to learn about Causal Machine Learning methods for more complex settings that are not covered in the course.

**Organisational Structure of the Course:**
- About two hours of theory lecture.
- About one hour presentation of application in R notebooks.

**Software requirements:**
*There will be no programming in this course*

**Recommended Literature to look at in advance:**

**Biography**

**Michael Knaus** is Assistant Professor of Econometrics at the University of St. Gallen. He holds a PhD from the University St. Gallen and studied previously Economics at the University of Hohenheim. His research interests are at the intersection of causal inference and machine learning to answer questions in empirical, mostly labor, economics. In particular, he is interested in the estimation of average and heterogeneous effects of interventions in observational settings as well as in the estimation of policy recommendations.
Schedule

Day 1: Machine Learning and its use for Causal Inference

- Introduction/refresher of goal of supervised machine learning methods
- Introduction/refresher of shrinkage methods (Lasso, Ridge, ...)

Literature:

Day 2: Machine Learning and its use for Causal Inference

- Introduction/refresher of tree-based methods (regression trees, random forests, ...)
- High-level discussion why plain machine learning is not suited for Causal Inference

Literature:

Day 3: Average effect estimation under unconfoundedness (homogeneous effects)

- Double Selection (variable selection for OLS)
- Double ML with partially linear model (more flexible than OLS)

Literature:
Day 4: Average effect estimation under unconfoundedness (homogeneous effects)
- Double ML with Augmented IPW (even more flexible than OLS)
- Low dimensional effect heterogeneity based on Double ML

Literature:

Day 5: Heterogeneous effect estimation
- Estimating individualized treatment effects with Causal Tree, Causal Forest, R-learner and DR-learner

Literature:

Student Assessment Methods

Students prepare a 5 to 10 page analysis of a real experiment of their choice or with simulated data using the methods covered on day 5. Instructions that are more detailed will be provided.