EuroCIM Course: Causal machine learning

ABSTRACT

Evaluating treatment effects using observational data increasingly requires adjustment for high-dimensional set of covariates in order to control confounding. This is the result of a lack of comparability between treated and untreated subjects in possibly many (pre-treatment) factors that are also related to outcome. While such adjustment is routinely achieved via parametric modelling, it is not entirely satisfactory as model misspecification is likely, and even relatively minor misspecifications over the observed data range may induce large bias in the treatment effect estimate. Unsurprisingly, machine learning methods are increasingly being used to assist in this task.

This course will cover the use of machine learning in causal effects estimation. It should also be of interest to researchers who rely on variable selection procedures in their data analyses.

This course will consist of 2 parts. Participants can register to either parts or both.

The first part is an introduction to causal machine learning, accessible to applied statisticians, epidemiologists, and other quantitative researchers already familiar with the language of potential/counterfactual outcomes and with identification assumptions, for example for the Average treatment effect. While the first part of the course will explain how (and why) causal machine learning works when the so-called canonical gradient of the considered target estimand is given, the second, more advanced part will focus on how the canonical gradient of an estimand can be derived, and what conditions are needed for causal machine learning based on this gradient to work well. This second part of the course is unavoidably more technical, but will build on intuition to provide insight. For this second part, some familiarity with asymptotic statistics (e.g., the Delta method) is required. Diverse examples will be used to convey the very broad applicability of the theory.

Further details:

The first part of the course will start with a motivation, argue why the starting point of a data analysis should be the choice of an estimand rather than a statistical model, and then introduce machine learning-based methods for the evaluation of (causal) treatment effects. We will highlight that bias can be introduced if using standard variable selection or machine learning methods as they are tuned for prediction performance, as opposed to estimation of treatment effects. We will then introduce the framework of "Targeted Learning" and other causal machine learning approaches, as a principled solution with optimal statistical properties for the estimation of causal treatment effects. The part of the course will include a hands-on session in R (do it yourself offline + webinar demonstration) where participants can experience the problems with naive machine learning and understand how Targeted Learning works by implementing it in real-world settings.

The second part of the course will focus on how to calculate the so-called canonical gradient of a causal or statistical estimand, which is the key tool needed for a statistician to be able to incorporate machine learning in a data analysis. It will moreover study the asymptotic behaviour of causal machine learning estimators, and the conditions needed for them to achieve favourable asymptotic properties. Also this part of the course will include a hands-on session (do it yourself offline + webinar demonstration) with pen and paper exercises on how to calculate the canonical gradient of an estimand.

Draft timetable

Introductory course

10am CET/9am UK time Session 1 (45 min): INTRODUCTION TO CAUSAL MACHINE LEARNING

15 min Break

11 am CET/10 am UK time Session 2 (45 min): CAUSAL MACHINE LEARNING IN ACTION

(12 noon CET/11am UK end of course for those only attending introduction)

ADVANCED (beginning of course for those attending only advanced part)

14:00 CET /1pm UK time Session 3 (45 min): CALCULATING THE CANONICAL GRADIENT OF AN ESTIMAND

15 min Break

15:00 CET /2pm UK time Session 4 (45 min): HANDS-ON CALCULATING THE CANONICAL GRADIENT OF AN ESTIMAND

15 min Break

16:00 CET/3pm UK time Session 5 (45 min): THE VON MISES EXPANSION

16:45 Final Q&A

17hrs CET /4pm UK time end