

THE Mini Course

This course provides an introduction to causal machine learning with applications using the software “R”.

Causal machine learning aims at assessing the causal effect of some intervention or treatment, like a medical treatment or a training program, on an outcome of interest, like health or wage. The assessment of a causal effect requires that groups receiving and not receiving the treatment are comparable in background characteristics which also affect their outcome (e.g., pre-treatment health, education etc.). Causal machine learning can be used to generate such comparable groups in a data-driven way by estimating two separate models for how the characteristics affect the treatment and the outcome. Such approaches also permit detecting subgroups for whom the treatment effect is particularly large as a function of their observed characteristics (effect heterogeneity analysis). This is useful for optimally targeting specific subgroups by the treatment (optimal policy learning).

This course discusses the underlying assumptions, intuition, and usefulness of machine learning for causal analysis. It also introduces various causal machine learning algorithms, like double lasso regression, causal forests, double machine learning, and optimal policy trees. Using the statistical software “R” and its interface “R Studio”, these methods are applied to various real-world data sets.

Objectives

- To understand the ideas and goals of machine learning for causal analysis
- To understand the intuition, advantages, and disadvantages of alternative methods
- To be able to apply causal machine learning to real world data using the software “R” and its interface “R Studio”

Prerequisites

- Introductory statistics (probability theory, conditional means, linear regression), basic command of the statistical software “R” is desirable, but not strictly required.

If you have questions please contact jung.benjamin@uni-hohenheim.de.