Doctoral Course: “Causal Inference and Data Fusion in Management and Economics Research”

This version: (First official draft)

Course instructor

Name: Dr. Paul Hünermund
Mail: phu.si@cbs.dk

Location

Campus Munich. Room tba.

Application and registration procedure

Goal and target audience

This course will provide an introduction into the topic of causal inference in machine learning and AI for doctoral students, with a focus on applications to business decision-making and economic research.

Application/Registration process

Please register for the course directly with the instructor via email.

Course aims

What this course is

This course offers an introduction into causal inference with directed acyclic graphs (DAGs). DAGs combine mathematical graph theory with statistical probability concepts and provide a powerful approach for causal modeling. Originally developed in the computer science and artificial intelligence field, they recently gained increasing traction also in other scientific disciplines (such as economics, political science, sociology, health sciences, and philosophy). DAGs allow to check the validity of causal statements based on intuitive graphical criteria, that do not require algebra. In addition, they open the possibility to completely automatize the causal inference task with the help of special identification algorithms. As an encompassing framework for causal reasoning, DAGs are becoming an essential tool for everyone interested in data science and machine learning.

What this course is not

This course is mainly conceptual and aims to bridge the gap between machine learning and quantitative

research in econometrics and the social sciences. It is not a theoretical Ai course that covers all the mathematical and technical details of the introduced causal inference methods.

Course objectives

Knowledge Objectives

The course provides a good overview of the advances that have been made in the field of causal data science in the last thirty year. The focus lies on practical applications of the theory and students will be put into the position to apply the covered methodologies in their own research. In particular, common causal inference challenges such as backdoor adjustment, bad controls, instrumental variables, selection bias, mediation, and external validity will be discussed in one single framework.

Skills Objectives

Hands-on examples using dedicated libraries in R will guide through the presented material. There are no prerequisites for participating, but a good working knowledge in basic statistics and R are a plus.

Learning Objectives

By the end of the course participants will:

- Gain a better understanding of common causal inference problems
- Be able to draw better connections between a variety of quantitative methodologies
- Master a powerful formalism for causal modeling
- Have deeper insights into methodological approaches from the field of causal AI
- Acquire various practical tools for solving causal inference challenges in their own research

Preliminary schedule

The course consists of a total of ten sessions with two sessions daily and a lunch break in-between, as well as a coffee break during the periods that take place in the afternoon.

1. **Session: Introduction to the Course (10:00 – 12:00, 24th April)**
    - Causal inference vs. prediction problems
    - The importance of causal knowledge for data-driven decision-making
    - Ladder of causation

2. **Session: Graphical Causal Models I (13:00 – 16:00, 24th April)**
    - Directed acyclic graphs
    - D-separation & testable implications
    - Interventions in structural causal models

3. **Session: R Exercise (10:00 – 12:00, 25th April)**
    - Introduction to causal inference libraries in R

4. **Session: Graphical Causal Models II (13:00 – 16:00, 25th April)**
    - Backdoor criterion
    - Matching and Regression
    - Front-door criterion

5. **Session: A/B Testing & Surrogate Experiments (10:00 – 12:00, 26th April)**
    - Randomized control trials

- A/B testing in business
- Surrogate experiments
- Instrumental variables

6. **Session: Application of Causal Inference in Business (13:00 – 16:00, 26th April)**
   - Current state and future directions of causal ML in the business sector
   - Virtual guest lecture by Sean J. Taylor (Rideshare Labs, Lyft): *"In defense of curve fitting: how experimentation-driven and ML-enabled causal inference drives impact at Lyft"* (keynote address at CDSM20)

7. **Session: Causal Artificial Intelligence (10:00 – 12:00, 27th April)**
   - Do-calculus
   - Identification algorithms
   - Data fusion paradigm

8. **Session: Sample Selection Bias (13:00 – 16:00, 27th April)**
   - Selection diagrams
   - Recovering from selection bias in causal diagrams
   - Selection propensity score
   - Heckman selection model

9. **Session: Counterfactuals & Mediation (10:00 – 12:00, 28th April)**
   - Potential outcomes framework
   - Ignorability
   - Mediation and causal mechanisms
   - Fairness in algorithmic decision-making

10. **Session: External Validity (13:00 – 16:00, 28th April)**
    - External validity
    - The transportability problem
    - Meta-transportability

Core readings

- Hünermund, P., and E. Bareinboim (2021). Causal Inference and Data Fusion in Econometrics. https://arxiv.org/abs/1912.09104
- Pearl, J., and D. Mackenzie (2018). The Book of Why. Basic Books, New York.

Course procedures

Practice exercises will be taught in R. To follow the hands-on examples, it is recommended that participants have R (https://cran.r-project.org/) and RStudio (https://www.rstudio.com/) installed on their laptops. Both programs are free and open source.

Assessment

The course is not graded: pass y/n on the basis of class participation and exercises.

Workload

3 ECTS (min. 21 hours lectures, 90 hours total (attendance time, preparation & follow-up workload)