GHENT UNIVERSITY

Course Specifications

Valid in the academic year 2022-2023

Causal Machine Learning (C004413)

Course size *(nominal values; actual values may depend on programme)*

Credits 5.0		**Study time 150 h**		**Contact hrs**	45.0h

Course offerings and teaching methods in academic year 2022-2023

A (semester 2)	English	Gent	lecture	22.5h
			seminar: coached exercises	22.5h
B (semester 2)	English	Gent	lecture	30.0h
			seminar: coached exercises	22.5h

Lecturers in academic year 2022-2023

Vansteelandt, Stijn	WE02	lecturer-in-charge
Dukes, Oliver	WE02	co-lecturer

Offered in the following programmes in 2022-2023	crdts	offering
Master of Science in Teaching in Science and Technology(main subject Mathematics)	6	B
Master of Science in Bioinformatics(main subject Systems Biology)	5	A
Master of Science in Computer Science	5	A
Master of Science in Mathematics	6	B
Master of Science in Statistical Data Analysis	5	A
Exchange Programme in Computer Science (master's level)	5	A
Exchange Programme in Mathematics (master's level)	6	B

Teaching languages

English

Keywords

Position of the course

This course offers a thorough investigation of statistical methods for causal inference from experimental and observational data. This methodology has wide applications in epidemiology, clinical studies, public health, agriculture, sociology, psychology, pedagogy, demography, economics...

Most scientific questions, such as those asked when evaluating policies or exposures, henceforth referred as treatments, are causal in nature, even if they are not specifically framed as such. Causal inference reasoning helps clarify the scientific question, and the assumptions necessary to express it in terms of the observed data. Once this is achieved, the focus shifts to estimation and inference. Estimating causal effects typically requires adjustment for confounding. This is the result of a lack of comparability between subjects due to possibly many factors that are related simultaneously to the outcome and the variable whose effect we aim to estimate. These adjustments can be achieved via parametric modelling. However, such traditional statistical tools are not entirely satisfactory as high-dimensional confounding is difficult to handle and model misspecification is likely. As even minor misspecifications can induce large bias in the treatment effect estimate, the task of learning functional relationships between variables in order to adjust for confounding is critical. Unsurprisingly, machine learning methods are increasingly being used to assist in this task. This is challenging because, while the prediction performance of a given machine learning algorithm can be measured by contrasting observed and predicted outcomes, performance evaluation becomes impossible for treatment effect estimation since the`ground truth', i.e. the true treatment effect, is unknown.

The aim of this course is to introduce machine learning-based methods for the evaluation of (causal) treatment effects. We will highlight that bias can be introduced if using standard machine learning methods that 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 course will include hands-on sessions in R where students can experience the problems with naive machine learning and understand how Targeted Learning works by implementing it in real-world settings. Students are free to use Python instead of R.

Contents

Part I: inferring the causal effect of an intervention

In this part, we will infer the effect of an intervention, treatment or exposure on an outcome at the population level, based on randomised experiments or observational studies (under an assumption of no unmeasured confounding). We will also discuss assumption-lean regression and the estimation of controlled direct effects, and the estimation of natural (in)direct effects to infer causal mechanism. Connections will be drawn to studies that evaluate the effect of a time-varying treatment, but will not be worked out.

- Confounding bias, selection bias
- Identification of the average causal effect, counterfactuals, exchangeability
- Plug-in estimators of the average causal effect and their limitations
- Debiased machine learning, targeted learning
- Implications for the analysis of randomised experiments, and comparison with unadjusted or adjusted, regression-based estimators
- Assumption-lean regression
- Causal diagrams (Causal Directed Acyclic Graphs)
- Debiased machine learning for controlled and natural direct effects

Part II: causal prediction

In this part, we will focus on individualised or conditional causal effects, as well as prediction of what an individual's outcome would be with versus without treatment (given specific covariate data available for that individual). We will moreover introduce conformal inference, a non-parametric strategy for obtaining prediction intervals.

- Shortcomings of standard prediction methods for causal prediction
- Causal random forests
- X-learner, R-learner, DR-learner, orthogonal statistical learning
- Conformal prediction (i.e., machine learning-based prediction intervals)

Initial competences

Having successfully completed a basic statistics course, as well as a regression course, or having acquired otherwise the corresponding competences. Machine learning is not a pre-requisite for this course, although some familiarity with machine learning techniques (e.g. random forests) is recommended.

Final competences

1 Make the fundamental distinction between association analysis and causal analysis.
2 Diagnose the possible presence of confounding and selection bias using causal diagrams.
3 Understand why standard data-adaptive (e.g., machine learning) approaches are fallible for inferring causal effects or causal predictions.
4 Understand the assumptions underlying different standard methods to correct for confounding bias in data analyses.
5 Apply causal machine learning methods for a variety of estimands that expresses the total, direct or indirect effect of an exposure, measured at a single time, on an outcome.
6 Apply orthogonal statistical learning algorithms to make causal predictions.
7 Express the uncertainty of results via confidence intervals or machine-learning based prediction intervals.

Conditions for credit contract

Access to this course unit via a credit contract is determined after successful competences assessment

Conditions for exam contract

This course unit cannot be taken via an exam contract

Teaching methods

Group work, Lecture, Seminar: coached exercises, Seminar: practical pc room classes

Extra information on the teaching methods

Exercises: written exercises and PC-labs using R (or Python, if the student prefers).

Learning materials and price

Scientific (review) papers and extended slides will be posted on Ufora. Cost: 5 EUR

References

- Pearl J (2000). Causality: Models, Reasoning, and Inference. Cambridge University Press.
- Pearl, J., & Mackenzie, D. (2018). *The book of why: the new science of cause and effect*. Basic books.
- Hernán, M.A. and Robins, J.M., 2010. Causal inference.

Course content-related study coaching

The students will frequently exercise the concepts and methods explained during the lectures, by analyzing realistic data sets during the practical sessions, where students will be closely supervised, and while making their project work. Besides the questions that students can ask before, during or after each lecture, there are several possibilities for asking questions: interactive support via Ufora (forum) and, in case of larger numbers of questions, personal coaching after electronic appointment.

Assessment moments

end-of-term and continuous assessment

Examination methods in case of periodic assessment during the first examination period

Oral examination, Open book examination, Written examination with open questions

Examination methods in case of periodic assessment during the second examination period

Oral examination, Open book examination, Written examination with open questions

Examination methods in case of permanent assessment

Assignment

Possibilities of retake in case of permanent assessment

examination during the second examination period is possible

Extra information on the examination methods

Theory: oral (open book)
Exercises: written (open book)
Project: written reporting
The entire exam assesses the student's insight into the basic principles of causal inference and causal machine learning and his/her ability to actively apply the statistical methods in the course. The exam will consist almost entirely of exercises; practical insight may also be tested via interpretation of given software-output (R).

Calculation of the examination mark

Theory: periodic
Exercises: periodic and permanent (project work)
One group project (written reporting) will be assigned.
Calculation of the total score: exam 80%, project 20%. A second examination chance for the project is possible. Non-participation to at least one of the project works implies a maximum score (exam + project) of at most 7/20, regardless of the score obtained on the final exam.