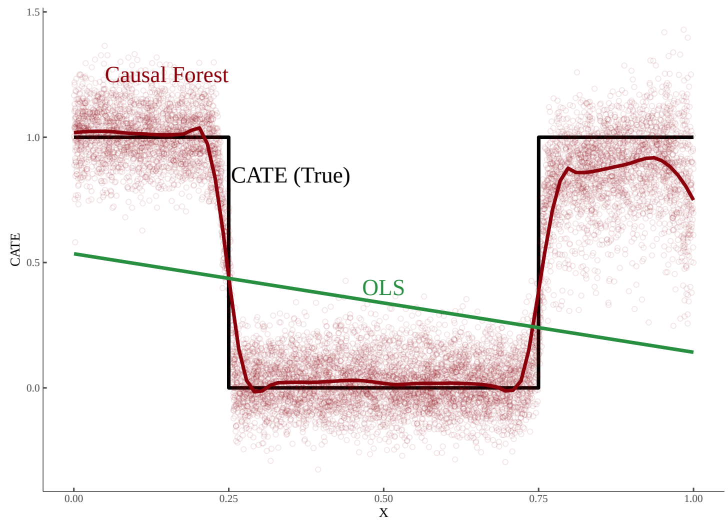
Causal Machine Learning for Observational and Experimental Research

ICPSR Summer Program in Quantitative Methods of Social Research
6–10 July, 2026 (9am–6pm EST)
Zoom



Instructors

Sam Fuller
Postdoctoral Fellow
Center for American Political Studies (CAPS)
Department of Government
Harvard University
sfuller@fas.harvard.edu

Jack T. Rametta
Postdoctoral Fellow
Ronald Reagan Presidential Foundation & Institute
jtrametta@gmail.com

Course Description

Machine learning is increasingly being applied at various stages in the social science workflow, and particularly for causal inference. This course provides a comprehensive overview for social scientists looking to incorporate machine learning methods into design-oriented, causal analyses, a broad set of methods that we call *Causal Machine Learning* (*CML*). Specifically, the course walks students through a flexible, data-driven approach to research design and analysis that represents a marked improvement over standard methods for applications such as treatment effect estimation (e.g., ATEs, CATEs, etc.), balance testing, and attrition detection. The topics covered include doubly-robust machine learning (DRML); interpretable ML; variable importance analysis; applications to observational data, experiments, and design-based cases (e.g., difference-in-differences, regression-discontinuity, etc.); and various CML methods such as Causal Forest.

Throughout, we compare the utility of causal machine learning approaches to traditional methods and provide practical guidance for evaluating validity, enhancing performance, and presenting results. Importantly, we provide detailed guidance on how to apply CML methods in a robust, principled way. Overall, this course provides an approachable guide for social scientists who wish to apply cutting-edge machine learning methods to enhance the quality and resolution of their causal research. All of our analyses will be conducted and demonstrated in R.

Required Software

`R` is available for download from CRAN (Comprehensive R Archive Network): https://cran.r-project.org/. Whatever flavor of integrated development environment (IDE) you prefer is fine for this course but you will need one. We generally recommend `RStudio` for beginners and intermediate users. For advanced users, you might also consider `VSCode`; `Positron`, a new IDE from the folks who brought you `RStudio`; or the more recent, AI-focused IDE `Cursor`. There are many required `R` packages that will be provided in an installation script separately.

While `Python` is not strictly required, it is recommended for some additional/bonus content we may cover in PyTorch (time permitting both EconML and TabPFN). We will also provide a basic demonstration for Hugging Face, a large repository for machine learning models typically written in `Python`, as well as how students can deploy models remotely through Google Colab, which can be quite powerful and provide access to very expensive and performant hardware.

Course Materials

Course materials (slides & code) will be available on the course Canvas page and updated daily.

Readings

Given the compact schedule for the course, we do not expect readings to be completed for class; instead, we include these readings as references.

Recommended Course Preparation

While not required, we recommend students come to the course with a research question and/or dataset of interest. This can be a well formulated project or a general research idea the student is interested in investigating. This will enable you to better engage with the course material and understand how these methods can be applied to their own research.

Tentative Schedule

This schedule is subject to change:

- Day One: Introduction, Causal Inference, & Classical Machine Learning
  - **Intro to the Class**
    Overview, expectations, (tentative) schedule
    ⇒ *Module 0*
  - **Causal Inference Refresher**
    Estimands, design-based inference, classical inference methods
    ⇒ *Module 1*
    ↪*Cunningham Chapter 4.0–4.1, Wager 1, & Chernozhukov et al. 2.1*
  - **Classical Machine Learning Crash-Course**
    Decision trees, random forests, gradient boosting, etc.
    ⇒ *Module 1*
    ↪*Hastie et al. 9.2, 10.1, 15*

- Day Two: Classical ML to Doubly Robust ML
  - **Transitioning from Classical to Causal ML**
    Interpretable machine learning and causal applications of classical ML
    ⇒ *Module 2*
    ↪ *Molnar 12, 13, 17, 18, 19–24*
  - **Doubly-Robust Machine Learning & Causal Forest for ATEs**
    The DRML framework, causal forest, and ATE estimation
    ⇒ *Module 3*
    ↪ *Wager 3, & Chernozhukov et al. 4, 10*
- Day Three: Heterogeneous Treatment Effects with Causal Forest
  - **Estimation and inference of heterogeneous treatment effects: "Soup to Nuts"**
    ⇒ *Module 4*
    ↪ *Wager Ch. 4, Chernozhukov et al. 15*
- Day Four: Experimental Research Design & Selection on Observables
  - **Experimental Research Design for CML**
    Designing and executing experiments with causal forest
    ⇒ *Module 5*
    ↪ *Wager Ch. 4, Chernozhukov et al. 15*
  - **Selection on Observables (CML for Observational Data)**
    Executing observational analyses with causal forest or double machine learning
    ⇒ *Module 6*
    ↪ *Wager 2, 7 & Chernozhukov et al. 5, Hubert & Copus (2021)*
- Day Five: Design-Based CML & Grab-Bag
  - **Design-Based Models with CML**
    Difference-in-differences, RDD, etc. with CML
    ⇒ *Module 7*
    ↪ *Wager Ch. 8, 13, Chernozhukov et al. 12, 16, & 17*
  - **Grab-bag, or: "If We Have Time"**
    Bonus `Python` implementations (if time permits)
    ⇒ *Module 8*
    ↪ *tabPFN, Samii's Conformal Tutorial, & econML*

Texts

In addition to our own materials, we will rely heavily on selections from Wager (2024) and Chernozhukov et al. (2024) for this course as well as a smattering of other published works. These are excellent resources and, remarkably, all are available for free online.[1]

1. Wager, Stefan. 2024. *Causal Inference: A Statistical Learning Approach.*
   https://web.stanford.edu/~swager/causal_inf_book.pdf
2. Chernozhukov, Hansen, Kallus, Spindler, and Syrgkanis. 2024. *Applied Causal Inference Powered by ML and AI.*
   https://causalml-book.org/
3. Cunningham, Scott. 2021. *Causal Inference: The Mixtape.*
   https://mixtape.scunning.com/
4. Hastie, Trevor, Tibshirani, Robert and Friedman, Jerome. 2017. *The Elements of Statistical Learning Data Mining, Inference, and Prediction.*
   https://hastie.su.domains/ElemStatLearn/
5. Molnar, Christoph. 2024. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.*
   https://christophm.github.io/interpretable-ml-book/
6. Alves, Matheus Facure. 2022. *Causal Inference for the Brave and True.*
   https://matheusfacure.github.io/python-causality-handbook/landing-page.html
7. Ashworth, Scott, Christopher R. Berry, and Ethan Bueno de Mesquita. 2021. *Theory and Credibility: Integrating Theoretical and Empirical Social Science.* Princeton University Press.
   https://press.princeton.edu/books/paperback/9780691213828/theory-and-credibility?srsltid=AfmBOopWRAIROr_E0hWBnqgNRb5ej_F0z2xejYebxPo1KH4MJDFs5Xe3

Papers

This paper provides a selection of papers that fruitfully apply causal machine learning methods in social science journals. Our bias is towards political science in this list, there are of course many others.

1. Fuller, Sam; de la Cerda, Nicolas; and Rametta, Jack T. 2025. Affect, Not Ideology: The Heterogeneous Effects of Partisan Cues on Policy Support. Political Behavior.
2. Jeres, Jake and Malhotra, Neil. 2024. Policy Impact and Voter Mobilization: Evidence from Farmers' Trade War Experiences. American Political Science Review.
3. Ratkovic, Marc. 2023. Relaxing assumptions, improving inference: integrating machine learning and the linear regression. American Political Science Review.
4. Grimmer, Justin; Roberts, Margaret E.; and Stewart, Brandon M. 2021. Machine Learning for Social Science: An Agnostic Approach. Annual Review of Political Science.

[1] With the exception of the Ashworth et al. book which can be purchased, the PDF copy is also available through some online libraries.