EECS 598-009: Causality and machine learning

Instructor: Maggie Makar (she/her) mmakar@umich.edu
Office hours: Wednesday, 1:40 - 2:40 P.M. Prior sign-up is required through this link

GSI: Trenton Chang (he/him) ctrenton@umich.edu
Office hours: Monday, 2:30 - 3:30 P.M. Prior sign-up is required through this link.

Course description

This course introduces the fundamental concepts of causality, and causal inference using machine learning models as well as causally-motivated predictive modeling. Topics will include: counterfactuals (potential outcomes and graphs), identification and estimation of conditional average treatment effects from randomized control trials and observational data, as well as causal inference under hidden confounding and limited overlap. On the predictive modeling side, we will study robust prediction using causal ideas, causally motivated regularization, fairness and causality among others.

Prerequisites:
Familiarity with statistics, probability and machine learning. Knowledge of Python.

Topics (see detailed lecture schedule below):

- Causal identification
- Double robustness
- Sensitivity analysis
- Causal discovery
- Causally-motivated predictive approaches

Course components

While class recordings will be available through canvas, it is highly recommended that you attend lectures in class since class participation is worth 10% of your final grade. Class attendance policies may change over the course of the term in case of spikes in COVID (or other emerging pathogen) cases. This course will follow all policies issued by the University, which are documented here: https://healthresponse.umich.edu/policies-guidance/. Such changes will be communicated through announcements on canvas so make sure you receive those notifications.

Lecture
Monday, Wednesdays 12-1:30, DOW 3150

Canvas
Information about the course including notes, assignments, schedule and course logistics will be posted on Canvas (https://canvas.umich.edu/). You are expected to check the site frequently. Please make sure that you are receiving notifications about announcements from Canvas.

Piazza
You are encouraged to use Piazza to answer each other's questions and ask clarifications on concepts, lecture topics, assignment problems and projects. You can access Piazza from the course Canvas or directly from https://piazza.com/umich/fall2023/eecs598009fa2023

Contacting course staff
You will get the fastest response if you send your emails to eecs598-cml-staff@umich.edu, which is monitored by the class staff. **Please do not contact the class staff individually unless you have sensitive/private information that you would rather communicate to only one member of the class staff.** If you must email the instructor/GSI separately please include [EECS 598 CML] in the subject line.

Course materials

Lecture notes and recordings are going to be posted within the 24 hours after the lecture. Announcements about class readings will be posted a week in advance. Not all readings will be required. We will notify you of required readings.

There is no required textbook, but here are some useful sources and we will follow some materials from them in the course:

- Causal Inference for Statistics, Social, and Biomedical Sciences, Imbens and Rubin
- Causality, Pearl
- Observational studies, Rosenbaum

Note that the first and third books are available as ebooks through UM library

Grading and late submission policy

- 4 homework assignments (evenly weighted): 39%
- Final project: 50%
    - This will be a group project of 2-3 students.
    - Includes a 2 page proposal, first draft, peer review of first draft, class presentation and an 8 page final draft.
    - The project can be a critical replication of an existing paper, an extension of an existing paper, or a novel application of an existing approach to an application.
- Class participation: 10%
- Course evaluations (midterm and final): 1%

The class will be graded on a curve

Three of the 4 HW assignments will be in the first half of the semester while the project components are all due (roughly) in the second half of the semester (see below for a tentative schedule).

We expect all HW/project components to be submitted on time (10pm on the due date). However, since there may be exceptional circumstances that occur, we will make **a total of 4 late days** available to you to use over the course of the semester for all homework assignments and project components. You can use:

- At most 3 of the 4 late days for a given HW assignment
- At most 2 of the 4 late days for the final project write up
- At most 1 late day for the project draft 1
- No late days can be used for the project presentation

Late days will be strictly enforced and are meant to *cover unexpected life events*. **Use these wisely!** We will count late days in increments of days starting immediately. For example, suppose that you submit 15 minutes late. This counts as a late day and will decrease your remaining late days by 1. Submissions made after the total number of late days has exceeded 4 or after the maximum number of allowed late days per submission has been used ***will not be accepted.***

The course evaluation is important to us and proof of completing the final course evaluation counts 1% towards your final grade: 0.5% for the midterm evaluations and 0.5% for the final evaluations. Students will receive full credit only if they submit the final course evaluations and upload a screenshot indicating completion (a corresponding assignment will be made available on Gradescope for this).

Students are allowed to submit regrade requests for HW assignments for up to 5 days after the due date (2 days after the solution is posted on Canvas).

Honor code and collaboration

Unless otherwise specified, all submitted work must be your own, original work. If you are referencing others' work, put it in quotes. If you are directly quoting, or building on others' writing, provide a citation. See the Rackham Graduate policy on Academic and Professional Integrity for the definition of plagiarism, and associated consequences. Violations of the Honor Code will be taken seriously; Please see details: https://elc.engin.umich.edu/honor-council/. Students are encouraged to collaborate on conceptual understanding. Please use Piazza to this effect and for other technical discussions. However, students are expected to write their solutions on their own and should not look at any other student's write-up.

Use of generative AI tools (e.g. ChatGPT, Dall-E, etc.) is permitted in this course in the following cases:

- If you are developing a method that builds upon these generative models as a part of your project (e.g., your project is on causally motivated fine-tuning of GPT4 embeddings)
- For checking grammar, writing style or making drafts of your original work more concise

It is not permitted in any other setting.

By enrolling in this class, you agree to not post the materials of the class publicly online.

Student mental health and wellbeing

We want you to be successful in and outside of this class. Students may experience stressors that can impact both their academic experience and their personal well-being. These may include academic pressure and challenges associated with relationships, mental health, alcohol or other drugs, identities, finances, etc.

If you are experiencing concerns, seeking help is a courageous thing to do for yourself and those who care about you. If the source of your stressors is academic, please contact Professor Makar so that we can find solutions together. For personal concerns, U-M offers many resources, some of which are listed at Resources for Student Well-being on the Well-being for U-M Students website (https://wellbeing.studentlife.umich.edu/resources-list).

Accommodations for students with disabilities

If you think you need an accommodation for a disability, please contact the Services for Students with Disabilities (SSD) office with your request. More information can be found here: https://ssd.umich.edu/students. Some aspects of this course may be modified to facilitate your

participation and progress. The course staff will treat any information you provide in this regard as private and confidential.

Detailed timeline

Note that this timeline is tentative and subject to change.

Date	Topic	Released	Due @ 10pm
Mon Aug 28	Lec 1: Introduction	HW0	
Wed Aug 30	Lec 2: Causal notation and language		
Mon Sep 4	LABOR DAY		
Wed Sep 6	Lec 3: ATE under randomization	HW1	HW0
Mon Sep 11	Lec 4: ATE with covariates -- plug-in estimators		
Wed Sep 13	Lec 5: ATE with covariates -- doubly robust estimators		HW1
Mon Sep 18	Lec 6: ATE under measured confounding		
Wed Sep 20	Lec 7: ATE under measured confounding -- doubly robust estimators	HW2	
Mon Sep 25	Lec 8: CATE under measured confounding		
Wed Sep 27	Lec 9: CATE under measured confounding		HW2
Mon Oct 2	Lec 10: Causally motivated predictive models -- robustness		
Wed Oct 4	Lec 11: CMPM -- robustness	Project prompt released	
Mon Oct 9	Lec 12: CMPM -- robustness		
Wed Oct 11	Lec 13: CMPM -- efficiency		
Mon Oct 16	FALL BREAK		
Wed Oct 18	Lec 14: CMPM -- fairness		Proposal due
Mon Oct 23	Lec 15: CMPM -- fairness		
Wed Oct 25	Project one-on-ones no lecture		
Mon Oct 30	Lec 16: ATE estimation with hidden confounding		
Wed Nov 1	Lec 17: CATE estimation with hidden confounding		Draft 1 due
Mon Nov 6	Lec 18: CATE estimation with hidden confounding		
Wed Nov 8	Lec 19: Proximal causal inference	HW3	Draft 1 reviews due
Mon Nov 13	Lec 20: Instrumental variables		
Wed Nov 15	Lec 21: Mediation		HW3
Mon Nov 20	Lec 22: Causal discovery		
Wed Nov 22	THANKSGIVING BREAK		
Mon Nov 27	Lec 23: Causal discovery		Project final drafts due
Wed Nov 29	Project presentations		
Mon Dec 4	Project presentations		