Causal Inference and Machine Learning (IFT 6251)

Winter 2022

January 11, 2022

Causal Inference and Machine Learning (~~IFT 6251~~)

IFT 6168

Contact Celine Begin if you have issues with enrollment!

Day 1 Plan

- Class logistics
- Brief introductions
- Bird eye’s overview of causality
- Summary of topics covered in the class

Course information

- **Most important:** Check course webpage (https://www.dsridhar.com/courses/787) regularly. I'll update it frequently with readings for every class, and other announcements (e.g., Covid-related changes).

- Best email for me is dhanya.sridhar@mila.quebec.

Course information

- **Course time:** Tuesdays, 12:30 — 2:30 PM **and** Fridays, 11:30 — 1:30 PM

- 5 min break halfway through, and we'll try to end 5-10 mins before allotted end time.

- **Course location:** Zoom for now, and when allowed, hybrid mode with in-person at **Mila Agora[1].**

- **Office hours:** Tues, 2:30 — 3:30 PM, on this Zoom meeting for now and in-person at **Mila** when allowed.

[1] 6650 Rue Saint-Urbain

Course prerequisites

- I expect students to have taken a course that covers the fundamentals of supervised and unsupervised machine learning.
- I expect knowledge of basic probability and statistics (i.e., not measure theory).
- Background on probabilistic graphical models will help.

Course workload

- There will be an assigned reading for every class after today. You'll turn in a reader report each class, and I'll mark whether they were turned in or not.

- You'll complete a self-directed course project. I'll grade the project reports.

Reader Report (QCRs)

- The format is: Question, Comment, Research idea (QCR).
- On average, they'll be 1 page and should take 15 minutes.
- Bullet points are fine.
- Feel free to use LaTeX if you want to include math. Plain text is also fine.
- **Remember,** the reader reports are for your benefit — they'll help you get more out of the class!

Course format

- Until approximately Jan. 28, we'll have lectures. I'll introduce background on causality.
- After that, students will take turns leading discussions on the assigned reading.

Discussion format

- Students upload reader reports by 8 AM morning of class.
- Leader clusters questions and comments from reports into some key discussion points.
- Leader presents for **10 mins.**
    - **Mainly:** situate the paper — what's the significance of reading this paper?
    - Summarize the key technical insights or results in the paper.
- We discuss the leader's clustered discussion points.

Course project

- Around March 11, you'll turn in a 0.5 to 1 page project proposal.
- Around April 1, you'll turn in an "aspirational abstract" for your project report.
- I'll give you as long as I can to turn in your project report.

Course project

What are examples of strong projects?

- A systematic comparison of methods for a causal inference problem we discussed.
- Application of causal inference techniques to interesting causal question and dataset.
- New method or theory for a causal inference problem.

Course grading

- **30%: Course participation**
  - 15%: Turning in QCR reports
  - 15%: Leading a paper discussion
- **70%: Course project**
  - 5%: Turning in project proposal
  - 5%: Turning in aspirational abstract
  - 60% Project report

Brief introductions (< *1 min*)

- Name
- Program
- Year in program,
- **Optional:** area of interest *(e.g., bioinformatics, NLP, economics, digital humanities, social science)*

5 Minute Break

Examples of ML tasks

- I have high-dimensional medical images of tissue for some patients and physician-provided labels whether the issue is normal or cancerous. For an unseen future patient, I want to predict their cancer status from imaging data.

- I have 1 million articles from PubMed. I want to automatically group similar articles together, and group similar words together.

Examples of ML tasks

- I have high-dimensional medical images of tissue for some patients and physician-provided labels whether the issue is normal or cancerous. For an unseen future patient, I want to predict their cancer status from imaging data.

$$P(Y = 1 \mid X_1, \ldots, X_N)$$

- I have 1 million articles from PubMed. I want to automatically group similar articles together, and group similar words together.

$$P(X_1, \ldots, X_N); \quad X_i = (X_{i1}, \ldots, X_{iM})$$

Modern ML aspects

$$P(Y = 1 \mid X_1, \ldots, X_N) \quad P(X_1, \ldots, X_N)$$

- Data like images or words in every document are high-dimensional. Because of this, learning low-dimensional representations is a key task.
- Inference involves observed distributions.
- Inference requires assumptions. The most common one is iid samples.

Examples of causal questions

- Does a new drug therapy cause adverse reactions in patients?

- I took Ibuprofen and my headache didn’t go away. Had I taken acetaminophen, would my headache have gone away?

Examples of causal questions

- Does a new drug therapy cause adverse reactions in patients?

$$P(Y; \mathrm{do}(X = 1)) \quad \textbf{vs} \quad P(Y; \mathrm{do}(X = 0))$$

Has adverse reaction

Takes new drug

Examples of causal questions

- Does a new drug therapy cause adverse reactions in patients?

$$P(Y; \mathrm{do}(X = 1)) \quad \textbf{vs} \quad P(Y; \mathrm{do}(X = 0))$$

- I took Ibuprofen and my headache didn't go away. Had I taken acetaminophen, would my headache have gone away?

$$Y_{X \leftarrow x_0} \mid Y = 1, X = x_1$$

Has headache

Takes acetaminophen

Takes ibuprofen

Causal inference aspects

$$P(Y; \mathrm{do}(X = 1))$$

$$Y_{X \leftarrow x_0} \mid Y = 1, X = x_1$$

- At face value, seems to be about one-dimensional variables.
- Inferences involve parameters of interventional distributions, or counterfactual distributions.
- We need assumptions to connect an intervention distribution to a distribution over observed variables.

How does this connect to ML?

	High-dimensional regime?	Inference	Assumptions
Modern ML	Yes	Observed distributions	• Samples are iid. • Independence assumptions about latent variables.
Classical causal inference	No	Interventional or counterfactual distributions	• Causal model of the system. • … • Samples are iid.

How does this connect to ML?

	High-dimensional regime?	Inference	Assumptions
Modern ML	Yes	Observed distributions	• Samples are iid. • Independence assumptions about [illegible]nt vari[illegible]
Classical causal inference	No	Interventional or counterfactual distributions	• Causal model of the system. • … • Samples are iid.

This class

Topics covered

- Introduction to causality
  - Causal graphical models
  - Defining causal quantities: interventions and counterfactuals
  - Identifying causal quantities: graphical criteria, and instrumental variables
  - Estimating causal quantities

- ML helps causality
  - Adapting neural networks for estimation
  - Learning representations for causal inference
  - Sensitivity analysis
  - Causal discovery

- Causality helps ML
  - Defining disentanglement
  - Criteria for better out-of-distribution generalization
  - Criteria for fair prediction

For next class

Read: Chapters 18, 19 and 20 of *Advanced Data Analysis from an Elementary Point of View,* Cosma Shalizi.

Key concepts: The “do” operator, directed acyclic graphical model, d-separation, causal graphical model, structural causal model, causal effect