

CS7290 Causal Modeling in Machine Learning: Homework 3

Submission guidelines

Use a Jupyter notebook and/or R Markdown file to combine code and text answers. Compile your solution to a static PDF document(s). Submit both the compiled PDF and source files. The TA's will recompile your solutions, and a failing grade will be assigned if the document fails to recompile due to bugs in the code. If you use Google Collab, send the link as well as downloaded PDF and source files.

Background

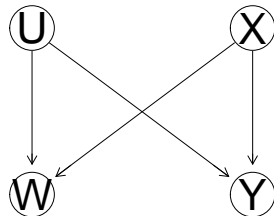
This assignment is going to cover several topics, including some that haven't been taught at the time this was assigned. We will cover those topics in subsequent classes.

- Recognizing valid adjustment sets
- Covariate adjustment with parent and back-door criterion
- Front-door criterion
- Propensity matching
- Intro to structural causal models

1. Recognizing valid adjustment sets

1.1

The following DAG represents a causal model of user behavior in an app.



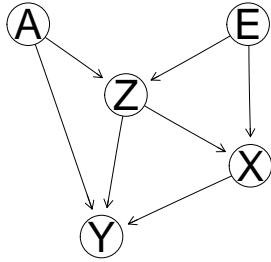
U represents the user specific preferences. X represents the introduction of a feature designed to make users make certain in-app purchases, Y was whether or not the user made the purchase, W represents app usage after the feature is introduced.

1. You are interested in estimating the causal effect of X on Y. What is the valid adjustment set?
2. What would happen if you adjusted for W? Be specific.
3. Suppose you want to assess the effect of X on Y for users who go on to have a high amount of app usage. You wanted to compute the causal effect for each level of W. Fill in the blanks on the right-hand-side and left-hand-side for the adjustment formula of interest:

$$P(Y = y|?) = \sum_{?} P(Y = y|?)P(?|?) \quad (1)$$

1.2

Consider the following DAG.

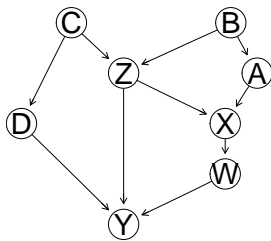


You are interested in estimating the causal effect of X on Y.

1. Is the set containing only Z a valid adjustment set? Why or why not?
2. List all of the valid adjustment sets (there are three) and write the adjustment formula for each adjustment set.
3. Suppose that E and A are both observable, but observing E costs \$10 per data point and observing A costs \$5 per data point. Which conditioning set do you go with?

1.3

Consider the following DAG:



1. List all of the sets of variables that satisfy the backdoor criterion to determine the causal effect of X on Y.
2. List all of the minimal sets of variables that satisfy the backdoor criterion to determine the causal effect of X on Y (i.e., any set of variables such that, if you removed any one of the variables from the set, it would no longer meet the criterion).
3. List all the minimal sets of variables that need to be measured in order to identify the effect of D on Y.
4. Now suppose we want to know the causal effect of intervening on 2 variables. List all the minimal sets of variables that need to be measured in order to identify the effect of set $\{D, W\}$ on Y, i.e., $P(Y = y | do(D = d), do(W = w))$.

2. Covariate adjustment

2.1

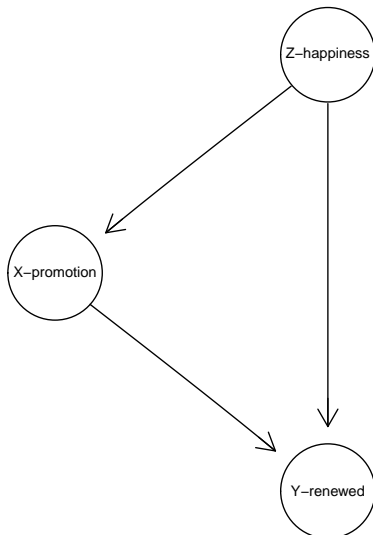
You are a data scientist at a prominent tech company with paid subscription entertainment media streaming service. You come across some data on a promotional campaign. The campaign targeted 70K subscribers users who were coming to a subscription renewal time and were at high risk of not renewing. They were targeted with two types of promotions, call them promotion 0 and promotion 1.

	Overall
Promotion 0	77.9% (27272/35000)
Promotion 1	82.6% (28902/35000)

You do some digging and find out the promotions the users were offered depended on how happy the users were (quantified from user behavior and customer service interactions).

	Overall	Unhappy	Happy
Promotion 0	77.9% (27272/35000)	93.2% (8173/8769)	73.3% (19228/26231)
Promotion 1	82.6% (28902/35000)	86.9% (23339 / 26872)	68.7% (5582/8128)

You assume the following causal DAG:



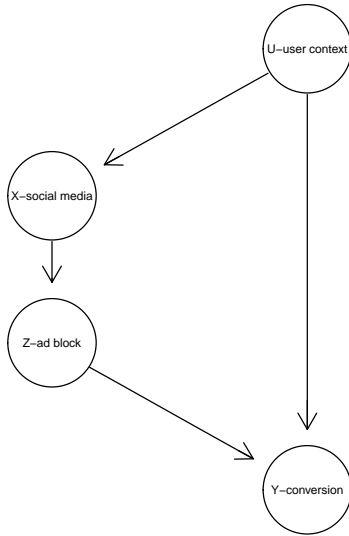
You are interested in the average causal effect $P(Y = 1|\text{do}(X = 0)) - P(Y = 1|\text{do}(X = 1))$

1. Build the model with Pyro using the values in the table. Use `pyro.condition` to calculate the causal effect by adjusting for happiness.
2. Suppose you could not observe happiness. Use `pyro.do` to calculate the causal effect with do-calculus.

unit = 1K	No adblock (50%)		Adblock (50%)		All subjects (800)	
	social	no social	social	no social	social	no social
Total	380	20	20	380	400	400
No conversion	323 (85%)	1 (5%)	18 (90%)	38 (10%)	341 (85.25%)	39 (9.75%)
Conversion	57 (15%)	19 (95%)	2 (10%)	342 (90%)	59 (14.75%)	361 (90.25%)
	No adblock	Adblock	No adblock	Adblock	No adblock	Adblock
No conversion	323 (85%)	18 (90%)	1 (5%)	38 (10%)	324 (81%)	56 (14%)
Conversion	57 (15%)	2 (10%)	19 (95%)	342 (90%)	76 (19%)	344 (86%)

2.2

Consider the table and the corresponding causal model.



1. User context is unobserved. Use `pyro.condition` to calculate the causal effect of social media on conversions using front-door adjustment.
2. Verify your result from number 1 using do-calculus with `pyro.do`.

3. Propensity scores (WIP)

This is a work in progress.

4. Structural causal models

4.1

Consider the SCM \mathbb{M} :

$$\begin{aligned}X &:= N_X \\Y &:= X^2 + N_Y \\N_X, N_Y &\stackrel{\text{i.i.d}}{\sim} N(0, 1)\end{aligned}$$

Write this model in Pyro and generate 10 samples of X and Y .

4.2

Consider the SCM \mathbb{M} :

$$\begin{aligned}X &:= N_X \\Y &:= 4X + N_Y \\N_X, N_Y &\stackrel{\text{i.i.d}}{\sim} N(0, 1)\end{aligned}$$

1. Draw a picture of the model's DAG.
2. $P_Y^{\mathbb{M}}$ is a normal distribution with what mean and variance?
3. $P_Y^{\mathbb{M}:do(X=2)}$ is a normal distribution with what mean and variance?
4. How and why does $P_Y^{\mathbb{M}:X=2}$ differ or not differ from $P_Y^{\mathbb{M}:do(X=2)}$?
5. $P_X^{\mathbb{M}:Y=2}$ is a normal distribution with what mean and variance?
6. $P_X^{\mathbb{M}:do(Y=2)}$ is a normal distribution with what mean and variance?
7. Write this model in code and generate 10 samples from $P_{X,Y}^{\mathbb{M}}$.
8. Use the `do` operator to generate 100 samples from $P_Y^{\mathbb{M}:do(X=2)}$ and visualize the results in a histogram.
9. Use the `condition` operator and a Pyro inference algorithm to generate 10 samples from $P_X^{\mathbb{M}:Y=2}$. Use one of the Bayesian inference procedures described in the lecture notes.

4.3 (WIP)

This problem is a work in progress. It will help you practice counterfactual reasoning using structural causal models.