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

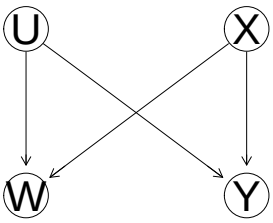
This assignment is going to cover several topics, including some that haven't been taught at the time this was assigned. We will cover those topics in subsequent classes.

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

Question 1: 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.

1.1.a

You are interested in estimating the causal effect of X on Y. What is the valid adjustment set? Valid adjustment set is the set of variables that if you adjust, you will get the unbiased results. (3 points)

There is nothing to adjust as there's no backdoor path. Valid adjustment set is {}

1.1.b

What would happen if you adjusted for W? Be specific. (2 points)

If W is conditioned, then the path U, W, X becomes active and hence there's a backdoor path that affects Y.

1.1.c

Suppose you want to assess the effect of X on Y for users who have a high amount of app usage. Fill in the blanks on the right-hand-side for the adjustment formula of interest:

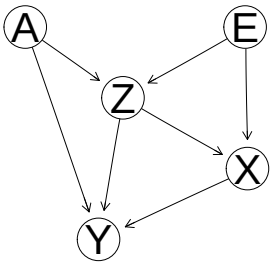
$$P(Y = y|do(X = x), W = high) = \sum_{?} P(Y = y|?)P(?|?) \tag{1}$$

(4 points)

$$P(Y = y|do(X = x), W = high) = \sum_{u \forall U} P(Y = y|X = x, W = high, U = u)P(U = u|W = High) \tag{2}$$

1.2

Consider the following DAG.



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

1.2.a

Is the set containing only Z a valid adjustment set? Why or why not? (2 points) No, If we adjust for z , then the triplet A, Z and E (path) becomes active. Hence, we need to condtion on A or E {A, Z} or {E, Z}

1.2.b

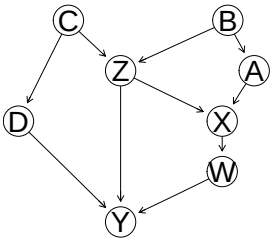
List all of the adjustment sets that blocks all the back doors(there are three) and write the adjustment formula for each adjustment set. (3 points)

{A, Z} {E, Z} {A, E, Z}

1.2.c

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 point)

{A, Z} ## 1.3 Consider the following DAG:



1.3.a

List all of the sets of variables that satisfy the backdoor criterion to determine the causal effect of X on Y. (3 points)

{ A, Z } { B, Z } { A, B, Z } { Z, C } { A, Z, C } { B, Z, C } { A, B, Z, C } { Z, D } { A, Z, D } { B, Z, D } { A, B, Z, D } { Z, C, D } { A, Z, C, D } { B, Z, C, D } { A, B, Z, C, D }

1.3.b

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 points) { D, Z } { C, Z } { B, Z } { A, Z }

1.3.c

List all the minimal sets of variables that need to be measured in order to identify the effect of D on Y. (3 points)

{ W, Z } { X, Z } { A, Z } { B, Z } { C }

1.3.d

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., $P(Y = y|do(D = d), do(W = w))$. (3 points)

{ Z } { C, X }