CMPT 733 – Big Data Programming II

Causal Inference

Instructor   Steven Bergner

Course website   https://sfu-db.github.io/bigdata-cmpt733/

Slides by: Yi Xie, Nathan Yan, and Jiannan Wang

Outline

Why Should Data Scientists Care?

Basic Concepts

Causal Inference

Questions Data Scientists Can Answer

Is This A or B?	Classification
How much or How Many?	Regression
Is This Weird?	Anomaly Detection
How Is This Organized?	Clustering

From Prediction to Causation

Predicting user activity for Xbox

- Y = logins in next month
- X = logins in past month, number of friends,...

What if we increase the number of friends?
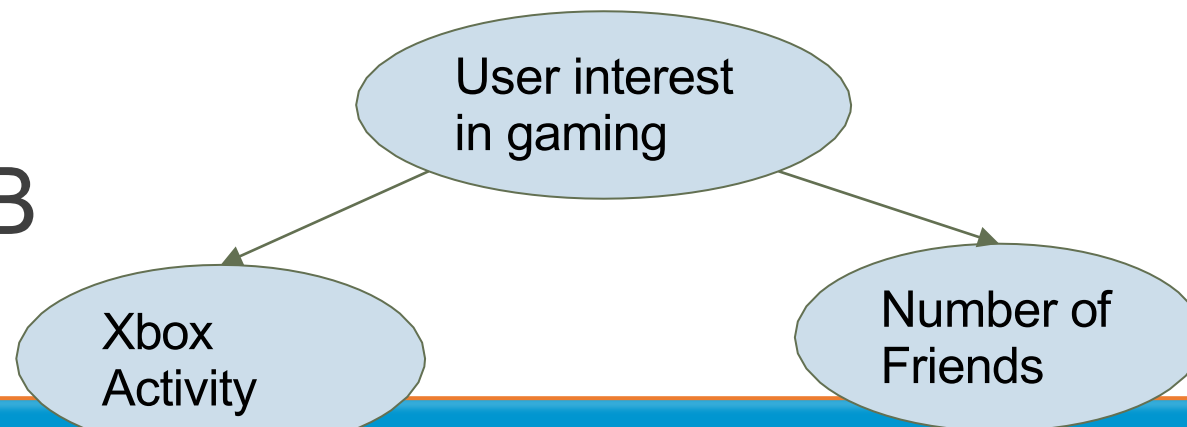
- Would it increase user activity?

Maybe or May be not (!)

A causes B

Number of Friends → Xbox Activity

B causes A

Xbox Activity → Number of Friends

C causes A and B

User interest in gaming → Xbox Activity

User interest in gaming → Number of Friends

A/B Testing Helps!

Treatment Group

- A random sample of users
- Launch a campaign to increase friends
- Average activity next month

Control Group

- A random sample of users
- **Not** Launch a campaign to increase friends
- Average activity next month

Hypothesis Testing

# A/B Testing Does Not Work

- **It is infeasible to do A/B testing**
  - What if you went to UBC rather than SFU, would it help you land a better job?

- **It is unethical to do A/B testing**
  - What if subscription price is set to $69 rather than $99, would it increase revenue?

Example 1: Causality ≠ Correlation

When a team was investigating the city death rate data, they found that reported cases of teenage drowning death increase while the sales of ice cream also increase.

# Example 1:Causality ≠ Correlation
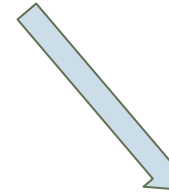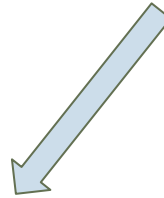
Example 2: Causality ≠ Correlation

Is Berkeley gender biased?

	Applicants	Admitted
Men	8442	**44%**
Women	4321	35%

Simpson's paradox

Example 2: Causality ≠ Correlation

Department	Men		Women	
	Applicants	Admitted	Applicants	Admitted
A	825	62%	108	**82%**
B	560	63%	25	**68%**
C	325	**37%**	593	34%
D	417	33%	375	**35%**
E	191	**28%**	393	24%
F	373	6%	341	**7%**

Department

Gender

Admit Rate
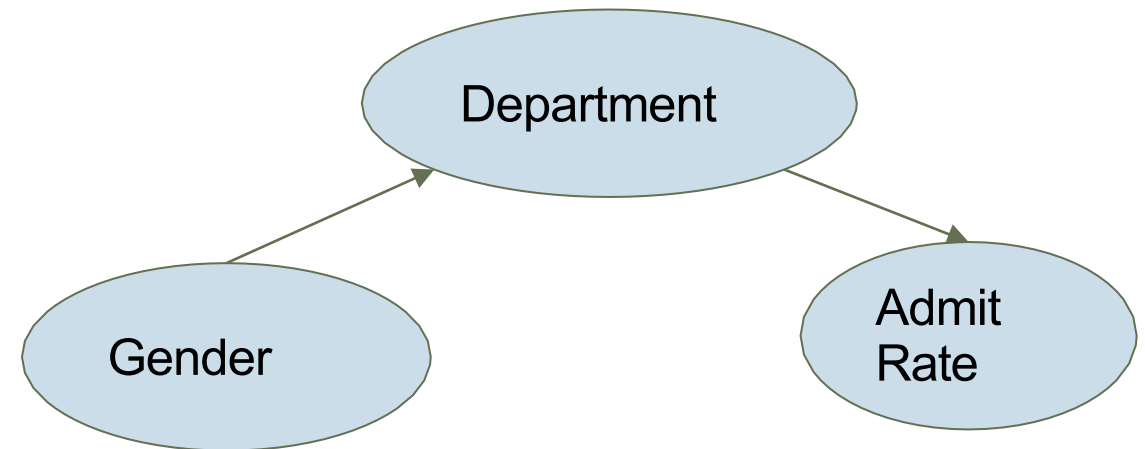
Outline

Why Should Data Scientists Care?

Basic Concepts

- Outcome / Treatment Variables
- Intervention and Do Operation
- Counterfactual
- Causal Graph

Causal Inference

Outcome / Treatment Variables

What if participating **Study Program**, would it improve **Grade**?

Treatment

Outcome

Student	Gender	Class	Study Program	Grade
Jacky	male	1	0	78
Terry	male	1	1	82
Mary	female	1	0	86
Sarah	female	2	1	83

Intervention and Do Operation

❖ Do operator will signal the experimental intervention (invented by Judea Pearl)

$$P(grade|do(enrolled\ in\ study\ program))$$

represents the distribution of grad if the person is enrolled in study program

Intervention and Do Operation

❖ P(A|B = b): probability of A being true given that B is observed as B = b

❖ P(A|do(B) = b): probability of A being true given an intervention that sets B to b

Counterfactual

❖ What would have happened if I had changed "Treatment Variable"

Student	Gender	Class	Study Program	Actual Grade	Counterfactual Grade (You cannot get it in reality)
Jacky	male	1	0	78	82
Terry	male	1	1	82	82
Mary	female	1	0	86	90
Sarah	female	2	1	83	85

Causality

Causality Definition

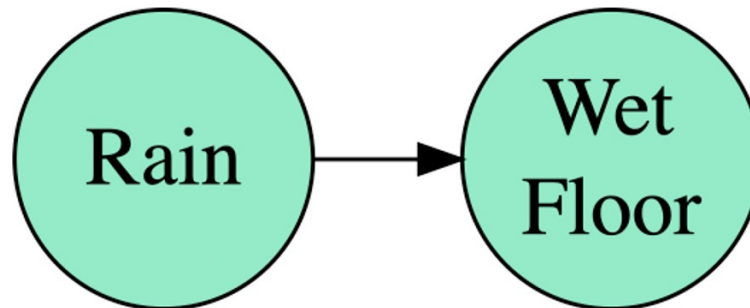- The difference between actual outcome and counterfactual outcome

The Fundamental Problem of Causal Inference

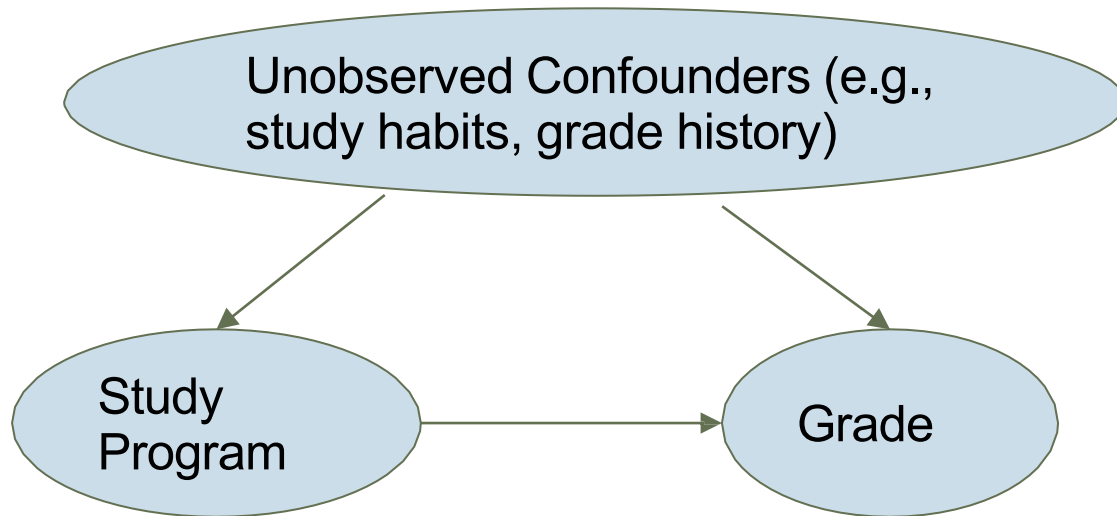- We cannot observe the counterfactual outcome

Causal Graph

Causal graph is a **directed** graph

- Nodes: variables
- Directed Edges: X affects Y

Causal Graph

❖ Confounding variables: common cause of treatment and outcome

Unobserved Confounders (e.g., study habits, grade history)

Study Program → Grade

Why Causal Graph?

● Helpful to identify which variables to control for

● Make assumptions explicit

Outline

Why Should Data Scientists Care?

Basic Concepts

Causal Inference

- Statistical Inference vs. Causal Inference
- Average Treatment Effect (ATE) Estimation
- Causal Inference in Practice

Statistical vs. Causal Inferences

Statistical inference

- Data is just a sample
- Your goal is to infer a population
- Think about how to go "backwards" from sample to population

Causal inference

- Derive a treatment group from Data
- Derive a control group (i.e., without treatment) from Data
- Think about how to infer the actual effect of treatment from the derived treatment and control groups

Outline

Why Should Data Scientists Care?

Basic Concepts

Causal Inference

- Statistical Inference vs. Causal Inference
- Average Treatment Effect (ATE) Estimation
- Causal Inference in Practice

Individual Treatment Effect

What is the grade difference between enrolling and not enrolling in the study program?

Student	Gender	Class	Enroll in Study Program	**Not** Enroll in Study Program	
Jacky	male	1	82	78	4
Terry	male	1	82	82	0
Mary	female	1	90	86	4
Sarah	female	2	83	85	-2

Average Treatment Effect (ATE)

The average of all values for individual treatment effects

$$ATE = (4 + 0 + 4 + -2) / 4 = 1.5$$

Student	Gender	Class	Enroll in Study Program	**Not** Enroll in Study Program	
Jacky	male	1	82	78	4
Terry	male	1	82	82	0
Mary	female	1	90	86	4
Sarah	female	2	83	85	-2

ATE Estimation

ATE estimation methods

❖ Matching based:
  ➢ Perfect matching
  ➢ Nearest neighbor matching
  ➢ Propensity score matching

❖ ML based:
  ➢ Regression method
  ➢ Representation learning

Perfect Matching

student	gender	class	Study program	grade
Terry	male	1	1	82
Sarah	female	2	1	83
Jacky	male	1	0	78
Mary	female	1	0	86

Find the "perfect matching in counterfactual world"

Perfect Matching

student	gender	class	Study program	grade
Terry	male	1	1	82
Sarah	female	2	1	83
Jacky	male	1	0	78
Mary	female	1	0	86

4

Find the "perfect matching in counterfactual world"

Perfect Matching

Student	Gender	Class	Study Program	Grade
Terry	male	1	1	82
Sarah	female	2	1	83
Jacky	male	1	0	78
Mary	female	1	0	86

By perfect matching, we can't find ATE over the population, because for tuple Sarah, we can't find the perfect match in control group

Nearest Neighbor Matching

Student	Gender	Class	Study Program	Grade
Terry	male	1	1	82
Sarah	female	2	1	83
Jacky	male	1	0	78
Mary	female	1	0	86

-3

Find the "nearest matching in counterfactual world"

Nearest Neighbor Matching

Student	Gender	Class	Study Program	Grade
Terry	male	1	1	82
Sarah	female	2	1	83
Jacky	male	1	0	78
Mary	female	1	0	86

4

-3

ATE = ½ [-3 + 4] = 0.5

Propensity Score Matching

Steps

❖ Using logistic regression to infer $e(x)=Pr[T=1|X=x]$

❖ Match users with treatment 0 with users with treatment 1 based on propensity score, using matching methods

Propensity Score Matching

Student	Gender	Class	Study Program	Grade	PSE
Terry	male	1	1	82	0.7
Sarah	female	2	1	83	0.6
Jacky	male	1	0	78	0.7
Mary	female	1	0	86	0.55

Based on PSM derived by logistic regression, the match we'll find for Terry is Jacky, and the match we'll find for Sarah is Mary

ATE = 0.5, computation is similar with nearest neighbor matching

Regression Method

Intuition: the distribution of Y given X is different when treatment is different

Train two separate regression models under Treatment = 0 or Treatment = 1, infer $p(Y | T = 0, X)$ and $p(Y | T = 1, X)$

Regression Method

Student	Gender	Class	Study Program	Grade
Terry	male	1	1	82
Sarah	female	2	1	83
Jacky	male	1	0	78
Mary	female	1	0	86

Treat one regression model 1on [Terry, Sarah] where Study Program = 1 treat another regression model 2 on [Jacky, Mary] where Study Program = 0

Using model 1to compute counterfactual outcome of [Jacky, Mary], same for model 2

Steven Bergner - CMPT 733

Other ML-based Methods

❖ Representation learning

❖ Intuition: transform dataset into a space where treatment assignment is more evenly distributed

❖ For other more techniques, please check latest publications

Outline

Why Should Data Scientists Care?

Basic Concepts

Causal Inference

- Statistical Inference vs. Causal Inference
- Average Treatment Effect (ATE) Estimation
- <mark>Causal Inference in Practice</mark>

DOWhy Python Library

❖ DoWhy is a Python library for causal inference that supports explicit modeling and testing of causal assumptions, developed by Microsoft.

❖ DoWhy is based on a unified language for causal inference, combining causal graphical models and potential outcomes frameworks.

https://github.com/microsoft/dowhy

Summary

Why Should Data Scientists Care?

- "What if" Question ?
- Why not A/B testing?
- Causality ≠ Correlation

Basic Concepts

- Outcome / Treatment Variables
- Intervention and Do Operation
- Counterfactual
- Causal Graph

Causal Inference

- Statistical Inference vs. Causal Inference
- Average Treatment Effect (ATE) Estimation
- Causal Inference in Practice