Causal Discovery

INFO/STSCI/ILRST 3900: Causal Inference

21 Nov 2023

Logistics

- Final project write-up due today at 5pm
- One person submits for each group
- Rubric on canvas
- Presentations on Nov 29

Learning goals for today

At the end of class, you will be able to:

1. Understand how conditional independence statements can help identify a causal graph
2. Understand the limits of using conditional independence to identify the causal graph

Causal Discovery

- We use DAGs to decide what needs to be "conditioned on" to identify causal effects

Causal Discovery

- We use DAGs to decide what needs to be "conditioned on" to identify causal effects
- Rules about open/blocked paths...

Causal Discovery

- We use DAGs to decide what needs to be "conditioned on" to identify causal effects
- Rules about open/blocked paths...
- In this class, we have mostly assumed the DAG is known from expert knowledge
- DAG tells us about conditional independence we would observe in data

$$\text{DAG} \Rightarrow \text{Conditional independence in data}$$

Causal Discovery

- We use DAGs to decide what needs to be "conditioned on" to identify causal effects
- Rules about open/blocked paths...
- In this class, we have mostly assumed the DAG is known from expert knowledge
- DAG tells us about conditional independence we would observe in data

$$\text{DAG} \Rightarrow \text{Conditional independence in data}$$

Can we estimate the DAG from data?

Causal Discovery

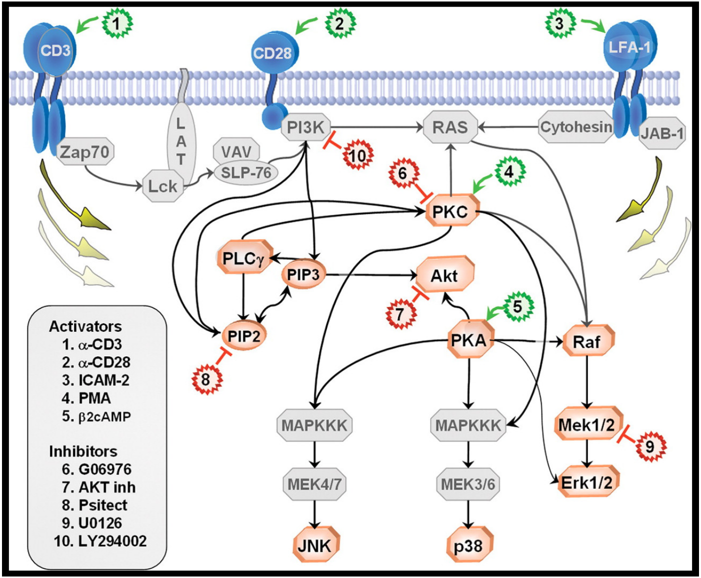


Figure: Regulatory network from Sachs et al. (2005)

Causal Discovery

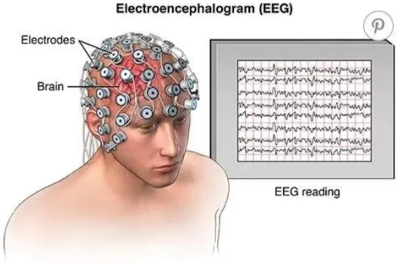


Figure: electroencephalogram (EEG)

Causal Discovery

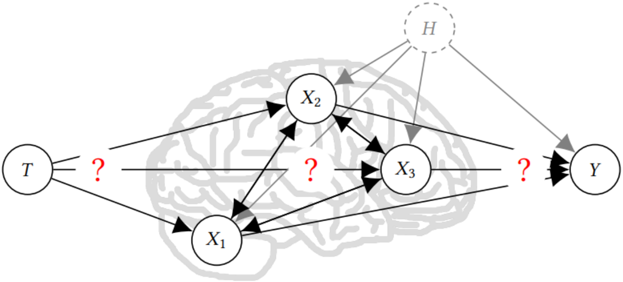


Figure: Neuroscience example from Weichwald and Peters (2020)

Causal Discovery

- Conditional independence is a observational quantity (i.e., not causal)
- Can be tested in observed data
- Can we go in the opposite direction?

$$\text{Conditional independence in data} \overset{?}{\Rightarrow} \text{DAG}$$

Open or blocked?

Two variables X and Y are conditionally independent given L if there exists an open path between X and Y given L

Open or blocked?

Two variables X and Y are conditionally independent given L if there exists an open path between X and Y given L

How to check if a path is open or blocked:

1. Traverse the path node by node
2. If any node is blocked, the entire path is blocked
3. If all nodes are open, then entire path is open

Open or blocked?

Two variables X and Y are conditionally independent given L if there exists an open path between X and Y given L

How to check if a path is open or blocked:

1. Traverse the path node by node
2. If any node is blocked, the entire path is blocked
3. If all nodes are open, then entire path is open

How to check if a node is open or blocked:

- If non-collider ($X \rightarrow L \rightarrow Y$ or $X \leftarrow L \rightarrow Y$) :

Open or blocked?

Two variables X and Y are conditionally independent given L if there exists an open path between X and Y given L

How to check if a path is open or blocked:

1. Traverse the path node by node
2. If any node is blocked, the entire path is blocked
3. If all nodes are open, then entire path is open

How to check if a node is open or blocked:

- If non-collider ($X \rightarrow L \rightarrow Y$ or $X \leftarrow L \rightarrow Y$) :
  - Open if it is not in the conditioning set
  - Blocked if it is in the conditioning set
- If collider ($X \rightarrow L \leftarrow Y$):

Open or blocked?

Two variables X and Y are conditionally independent given L if there exists an open path between X and Y given L

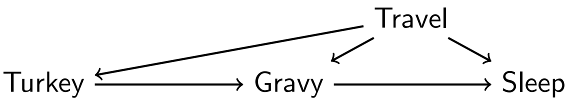
How to check if a path is open or blocked:

1. Traverse the path node by node
2. If any node is blocked, the entire path is blocked
3. If all nodes are open, then entire path is open

How to check if a node is open or blocked:

- If non-collider ($X \to L \to Y$ or $X \leftarrow L \to Y$) :
  - Open if it is not in the conditioning set
  - Blocked if it is in the conditioning set
- If collider ($X \to L \leftarrow Y$):
  - Open if it or any of its descendants are in the conditioning set
  - Otherwise it is blocked

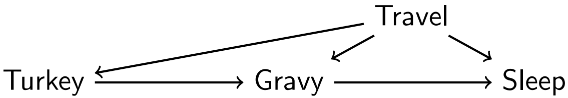
Practice



If we condition on $L = \{Gravy\}$, which paths are open? Which paths are blocked?

- $Turkey \rightarrow \underbrace{Gravy}_{\text{NC}} \rightarrow Sleep$
- $Turkey \leftarrow \underbrace{Travel}_{\text{NC}} \rightarrow Sleep$
- $Turkey \rightarrow \underbrace{Gravy}_{\text{Col}} \leftarrow \underbrace{Travel}_{\text{NC}} \rightarrow Sleep$
- $Turkey \leftarrow \underbrace{Travel}_{\text{NC}} \rightarrow \underbrace{Gravy}_{\text{NC}} \rightarrow Sleep$

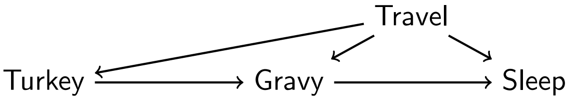
Practice



If we condition on $L = \{Gravy\}$, which paths are open? Which paths are blocked?

- $Turkey \rightarrow \underbrace{Gravy}_{\text{NC}} \rightarrow Sleep$ Blocked
- $Turkey \leftarrow \underbrace{Travel}_{\text{NC}} \rightarrow Sleep$
- $Turkey \rightarrow \underbrace{Gravy}_{\text{Col}} \leftarrow \underbrace{Travel}_{\text{NC}} \rightarrow Sleep$
- $Turkey \leftarrow \underbrace{Travel}_{\text{NC}} \rightarrow \underbrace{Gravy}_{\text{NC}} \rightarrow Sleep$

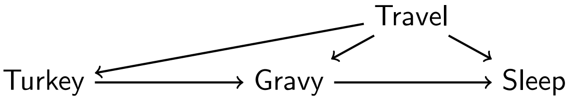
Practice



If we condition on $L = \{Gravy\}$, which paths are open? Which paths are blocked?

- $Turkey \rightarrow \underbrace{Gravy}_{\text{NC}} \rightarrow Sleep$ Blocked
- $Turkey \leftarrow \underbrace{Travel}_{\text{NC}} \rightarrow Sleep$ Open
- $Turkey \rightarrow \underbrace{Gravy}_{\text{Col}} \leftarrow \underbrace{Travel}_{\text{NC}} \rightarrow Sleep$
- $Turkey \leftarrow \underbrace{Travel}_{\text{NC}} \rightarrow \underbrace{Gravy}_{\text{NC}} \rightarrow Sleep$

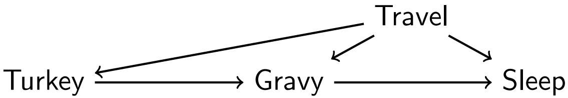
Practice



If we condition on $L = \{Gravy\}$, which paths are open? Which paths are blocked?

- $Turkey \rightarrow \underbrace{Gravy}_{\text{NC}} \rightarrow Sleep$ Blocked
- $Turkey \leftarrow \underbrace{Travel}_{\text{NC}} \rightarrow Sleep$ Open
- $Turkey \rightarrow \underbrace{Gravy}_{\text{Col}} \leftarrow \underbrace{Travel}_{\text{NC}} \rightarrow Sleep$ Blocked
- $Turkey \leftarrow \underbrace{Travel}_{\text{NC}} \rightarrow \underbrace{Gravy}_{\text{NC}} \rightarrow Sleep$

Practice



If we condition on $L = \{Gravy\}$, which paths are open? Which paths are blocked?

- $Turkey \rightarrow \underbrace{Gravy}_{\text{NC}} \rightarrow Sleep$ Blocked
- $Turkey \leftarrow \underbrace{Travel}_{\text{NC}} \rightarrow Sleep$ Open
- $Turkey \rightarrow \underbrace{Gravy}_{\text{Col}} \leftarrow \underbrace{Travel}_{\text{NC}} \rightarrow Sleep$ Blocked
- $Turkey \leftarrow \underbrace{Travel}_{\text{NC}} \rightarrow \underbrace{Gravy}_{\text{NC}} \rightarrow Sleep$ Open

Causal Discovery

How might we check conditional independence of X and Y given Z?

Causal Discovery

How might we check conditional independence of X and Y given Z? If variables are normally distributed

- Regress X and Y onto Z

Causal Discovery

How might we check conditional independence of X and Y given Z? If variables are normally distributed

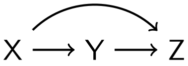
- Regress X and Y onto Z
- Get residuals from each regression

Causal Discovery

How might we check conditional independence of X and Y given Z? If variables are normally distributed

- Regress X and Y onto Z
- Get residuals from each regression
- Check if residuals are correlated

Causal Discovery

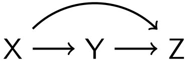


- $X \perp\!\!\!\perp Y$?
- $X \perp\!\!\!\perp Z$?
- $Z \perp\!\!\!\perp Y$?
- $X \perp\!\!\!\perp Y \mid Z$?
- $Y \perp\!\!\!\perp Z \mid X$?
- $X \perp\!\!\!\perp Z \mid Y$?

$$\mathrm{X} \longrightarrow \mathrm{Y} \longrightarrow \mathrm{Z}$$

- $X \perp\!\!\!\perp Y$?
- $X \perp\!\!\!\perp Z$?
- $Z \perp\!\!\!\perp Y$?
- $X \perp\!\!\!\perp Y \mid Z$?
- $Y \perp\!\!\!\perp Z \mid X$?
- $X \perp\!\!\!\perp Z \mid Y$?

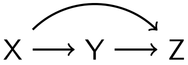
Causal Discovery



- $X \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Z$? No
- $Z \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Y \mid Z$? No
- $Y \perp\!\!\!\perp Z \mid X$? No
- $X \perp\!\!\!\perp Z \mid Y$? No

- $X \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Z$? No
- $Z \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Y \mid Z$? No
- $Y \perp\!\!\!\perp Z \mid X$? No
- $X \perp\!\!\!\perp Z \mid Y$? Yes

Causal Discovery



- $X \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Z$? No
- $Z \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Y \mid Z$? No
- $Y \perp\!\!\!\perp Z \mid X$? No
- $X \perp\!\!\!\perp Z \mid Y$? No

$$\mathrm{X} \longrightarrow \mathrm{Y} \longrightarrow \mathrm{Z}$$

- $X \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Z$? No
- $Z \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Y \mid Z$? No
- $Y \perp\!\!\!\perp Z \mid X$? No
- $X \perp\!\!\!\perp Z \mid Y$? Yes

If there is an edge between two nodes, they cannot be made conditionally independent!

Rule 1

- Start with (undirected) edges between every pair of nodes
- If you can find a set L such that $X \perp\!\!\!\perp Y \mid L$, take away the edge between X and Y

Rule 1

- Start with (undirected) edges between every pair of nodes
- If you can find a set L such that $X \perp\!\!\!\perp Y \mid L$, take away the edge between X and Y

Allows us to find where the edges are, but not necessarily direction

Rule 1

- Start with (undirected) edges between every pair of nodes
- If you can find a set L such that $X \perp\!\!\!\perp Y \mid L$, take away the edge between X and Y

Allows us to find where the edges are, but not necessarily direction

A **skeleton** is the DAG where we have made all edges undirected

$$\text{DAG} : X \rightarrow Y \rightarrow Z \qquad\qquad \text{Skeleton} : X - Y - Z$$

Causal Discovery

Can we also tell which direction an edge points?

Causal Discovery

$$X \longrightarrow Y \longrightarrow Z$$

- $X \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Z$? No
- $Z \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Y \mid Z$? No
- $Y \perp\!\!\!\perp Z \mid X$? No
- $X \perp\!\!\!\perp Z \mid Y$? Yes

$$X \longrightarrow Y \longleftarrow Z$$

- $X \perp\!\!\!\perp Y$?
- $X \perp\!\!\!\perp Z$?
- $Z \perp\!\!\!\perp Y$?
- $X \perp\!\!\!\perp Y \mid Z$?
- $Y \perp\!\!\!\perp Z \mid X$?
- $X \perp\!\!\!\perp Z \mid Y$?

Causal Discovery

$X \longrightarrow Y \longrightarrow Z$

- $X \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Z$? No
- $Z \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Y \mid Z$? No
- $Y \perp\!\!\!\perp Z \mid X$? No
- $X \perp\!\!\!\perp Z \mid Y$? Yes

$X \longrightarrow Y \longleftarrow Z$

- $X \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Z$? Yes
- $Z \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Y \mid Z$? No
- $Y \perp\!\!\!\perp Z \mid X$? No
- $X \perp\!\!\!\perp Z \mid Y$? No

Colliders can sometimes tell us the direction of an edge

Rule 2

- Suppose we have $X - Y - Z$ and no edge between X and Z
- Suppose $X \not\perp\!\!\!\perp Y \mid L$ for some set L that does not contain Y

Rule 2

- Suppose we have $X - Y - Z$ and no edge between X and Z
- Suppose $X \not\perp\!\!\!\perp Y \mid L$ for some set L that does not contain Y
- Then, $X \rightarrow Y \leftarrow Z$

Rule 2

- Suppose we have $X - Y - Z$ and no edge between X and Z
- Suppose $X \not\perp\!\!\!\perp Y \mid L$ for some set L that does not contain Y
- Then, $X \rightarrow Y \leftarrow Z$
- **Unshielded collider**: $X \rightarrow Y \leftarrow Z$ and X and Z do not have an edge

Causal Discovery

How far can we go?
Can we fully determine the graph from data?

Causal Discovery

$$X \longrightarrow Y \longrightarrow Z$$

- $X \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Z$? No
- $Z \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Y \mid Z$? No
- $Y \perp\!\!\!\perp Z \mid X$? No
- $X \perp\!\!\!\perp Z \mid Y$? Yes

$$X \longleftarrow Y \longleftarrow Z$$

- $X \perp\!\!\!\perp Y$?
- $X \perp\!\!\!\perp Z$?
- $Z \perp\!\!\!\perp Y$?
- $X \perp\!\!\!\perp Y \mid Z$?
- $Y \perp\!\!\!\perp Z \mid X$?
- $X \perp\!\!\!\perp Z \mid Y$?

Causal Discovery

$$X \longrightarrow Y \longrightarrow Z$$

- $X \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Z$? No
- $Z \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Y \mid Z$? No
- $Y \perp\!\!\!\perp Z \mid X$? No
- $X \perp\!\!\!\perp Z \mid Y$? Yes

$$X \longleftarrow Y \longleftarrow Z$$

- $X \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Z$? No
- $Z \perp\!\!\!\perp Y$? No
- $X \perp\!\!\!\perp Y \mid Z$? No
- $Y \perp\!\!\!\perp Z \mid X$? No
- $X \perp\!\!\!\perp Z \mid Y$? Yes

Some graphs have the exact same set of conditional independence statements and cannot be distinguished from data alone!

Rule 3

Graphs have the same conditional independence statements if

- Same skeleton: edges in the same location, but possibly different direction (from Rule 1)
- Same unshielded colliders: $X \to Y \leftarrow Z$ and X and Z do not share an edge (from Rule 2)

Causal Discovery

- Under few assumptions, we can test for conditional independence in data

Causal Discovery

- Under few assumptions, we can test for conditional independence in data
- This allows us to find the skeleton

Causal Discovery

- Under few assumptions, we can test for conditional independence in data
- This allows us to find the skeleton
- We can orient some edges

Causal Discovery

- Under few assumptions, we can test for conditional independence in data
- This allows us to find the skeleton
- We can orient some edges
- Need experiments or more assumptions to get further

Causal Discovery

- Under few assumptions, we can test for conditional independence in data
- This allows us to find the skeleton
- We can orient some edges
- Need experiments or more assumptions to get further
- Happy Thanksgiving!