Causal discovery

AI4ER MRes Workshop

So far…

So far we have assumed we know the causal graph (the DAG)!

This may not always be true!!

The question is…can we discover it from data…?

This is a real leap…how are we going to do it?!!

This is the world of causal discovery

- First thing to note - **some people think this is impossible and should never even be entertained.**
- There are a number of methods to do this which will be outlined here.
- I, for one, am undecided…

Firstly, some assumptions we have to make

1. **Acyclicity** — causal structure can be represented by DAG, G
2. **Markov Property** — all nodes are independent of their non-descendants when conditioned on their parents
3. **Faithfulness** — all conditional independencies in true underlying distribution p are represented in G
4. **Sufficiency** — any pair of nodes in G has no common external cause

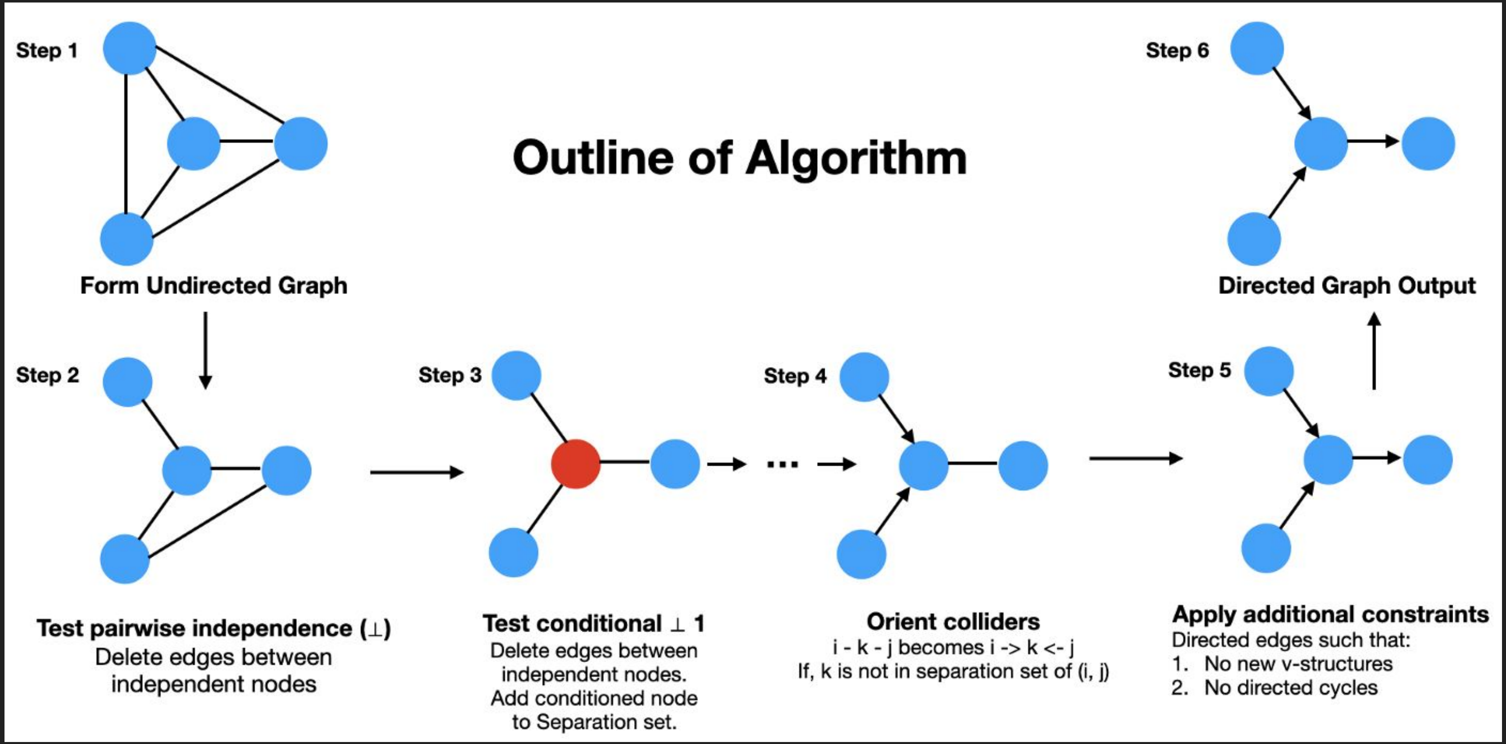
Any ideas?

- Does anyone have any idea how we might construct our causal graph from data?

Method 1: Conditional Independence Testing (constraints)

- This method relies on the idea that two statistically independent variables are not causally linked.

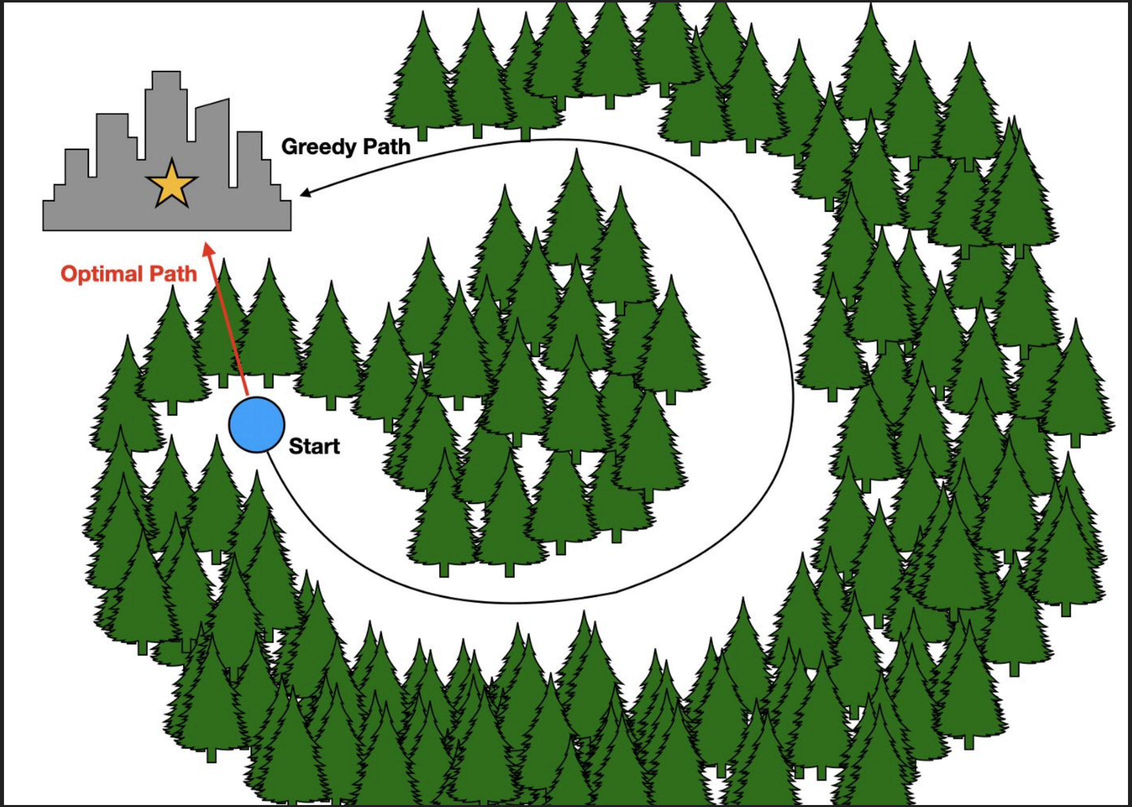
Outline of how these algorithms work!



Method 2: Greedy Search of Graph Space (score-based)

- There are three main elements to this method: a graph, a graph space, and a greedy search.
- A graph space is a collection of graphs. This is just a fancy way to formalize that there are many possible graphs for a given number of vertices and edges. For example, a DAG with 2 vertices and 1 edge could take the forms: $A \rightarrow B$ or $B \rightarrow A$.
- Finally, a greedy search is a way to navigate a space such that you always move in a direction that seems most beneficial based on the local surroundings.

Greedy Search



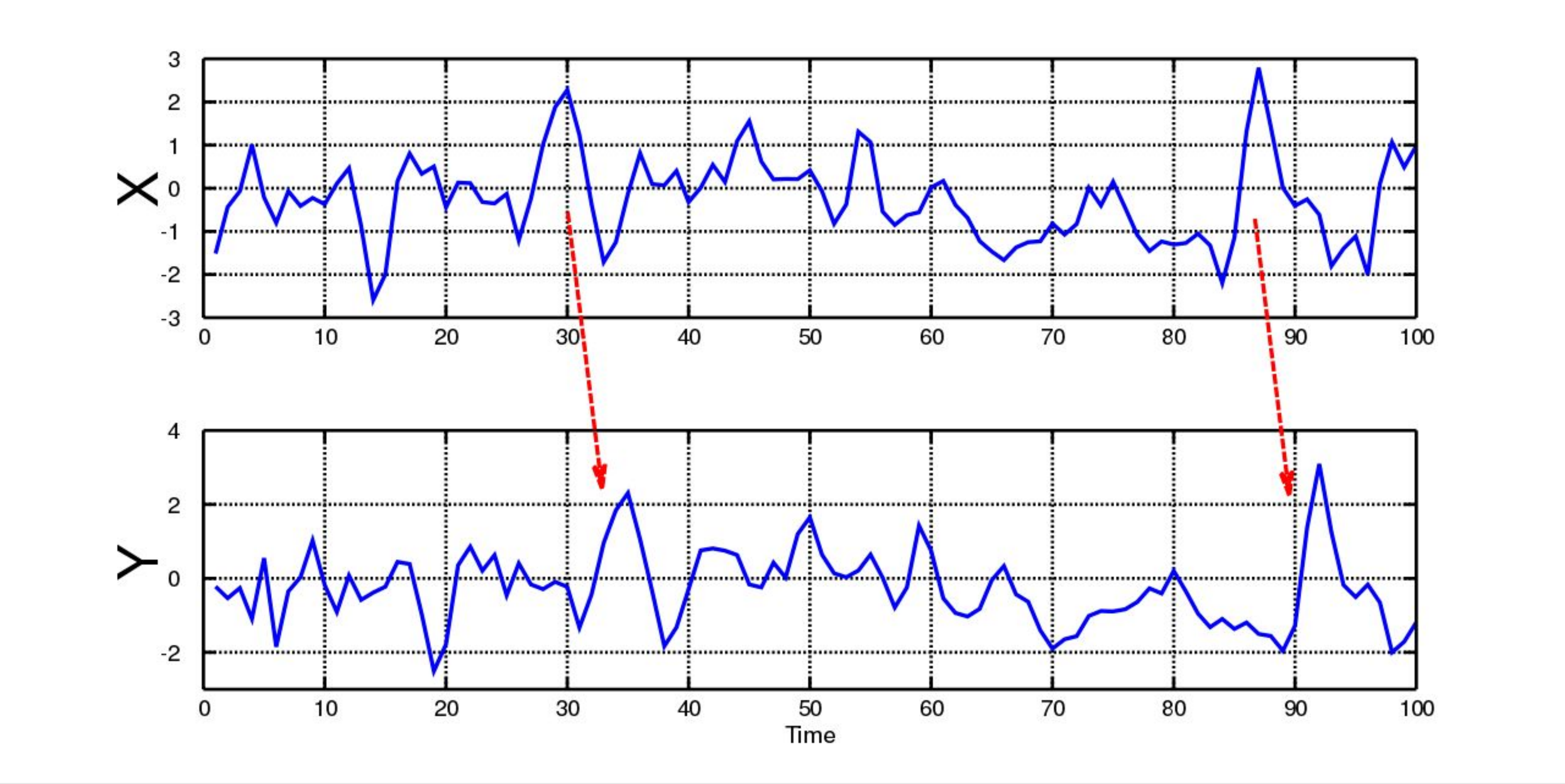
Greedy Equivalence Search

- The Greedy Equivalence Search (GES) algorithm uses this trick. GES starts with an empty graph and iteratively adds directed edges such that the improvement in a model fitness measure (i.e. score) is maximized. An example score is the Bayesian Information Criterion (BIC).

Method 3: Exploiting Asymmetries

- A fundamental property of causality is asymmetry. A could cause B, but B may not cause A. There is a large space of algorithms that leverage this idea to select between causal model candidates.
- One possible asymmetry is **time**. A cause must precede the effect!
- This sits at the core of **Granger causality**. Although Granger causality is not sufficient to claim causality, it leverages the idea that causes precede effects.
- It does this in the two variable case (e.g. X and Y), by quantifying the gain in predicting Y given past information of Y and X, as opposed to past information of Y alone.

Example of Granger Causality



Causal discovery for time series!

- Granger causality points at causal discovery for time series.
- This is highly relevant for environmental problems, and as such we will explore in more depth in code.