



UiO : **University of Oslo**

# Causality in Machine Learning

For the IFI MLS research seminar, May 22, 2020

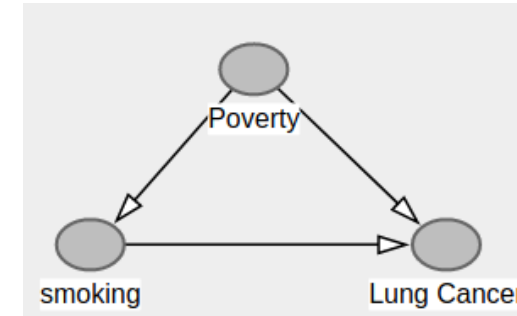
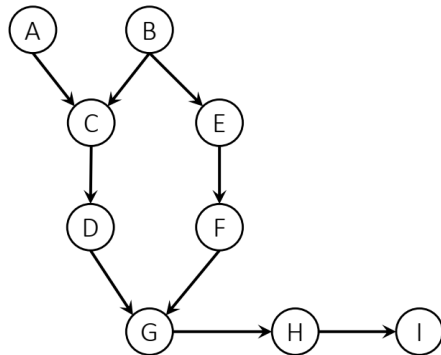


# Causality is relevant in several areas of machine learning

## 1. Moving from ML towards AI



## 2. Making better predictions



## 3. Learning causal relationships from data

# Making algorithms aware of causality could be an important step towards a true artificial intelligence

ML algorithms crunch data based on an objective

Humans reason about causes and effects

$f(x) = \dots$

HR Information		Contact	
Position	Salary	Office	Extn.
Accountant	\$162,700	Tokyo	5407
Chief Executive Officer (CEO)	\$1,200,000	London	5797
Junior Technical Author	\$86,000	San Francisco	1562
Software Engineer	\$132,000	London	2558
Software Engineer	\$206,850	San Francisco	1314
Integration Specialist	\$372,000	New York	4804
Software Engineer	\$163,500	London	6222
Pre-Sales Support	\$106,450	New York	8330
Sales Assistant	\$145,600	New York	3990
Senior Javascript Developer	\$433,060	Edinburgh	6224



## Causality has been formalized in terms of counterfactuals and directed acyclic graphs (DAGS)

### Counterfactuals

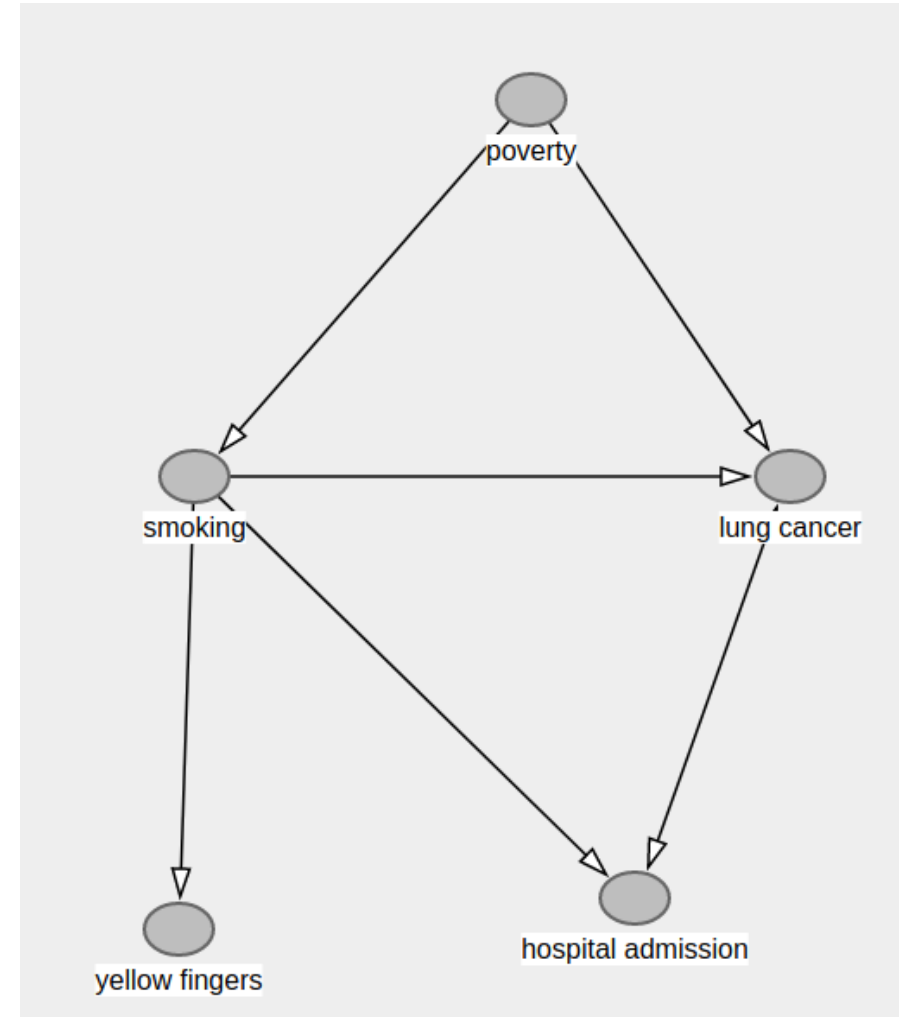
*"If patient X had not smoked she would not have gotten lung cancer"*

$$O_X^{S=1} = 0$$

*"If all Norwegians smoked 10% of them would get cancer within 10 years"*

$$E[O_{\text{Norway}}]^{S=1} = 0.1$$

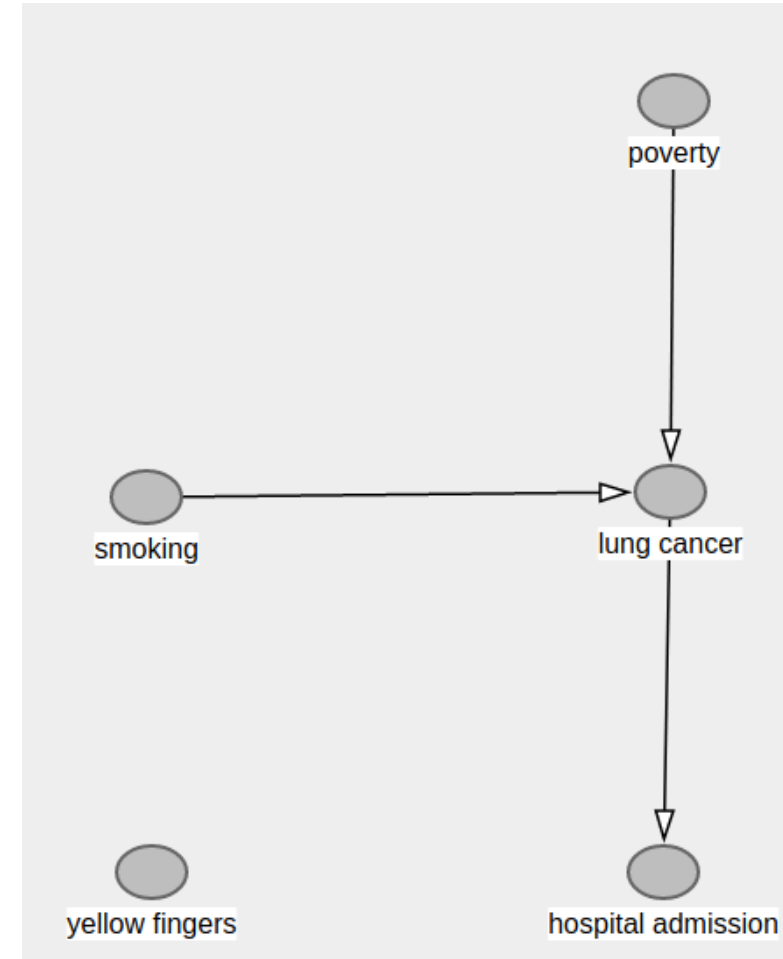
DAG



## The ideal dataset for inferring causality is a randomized trial

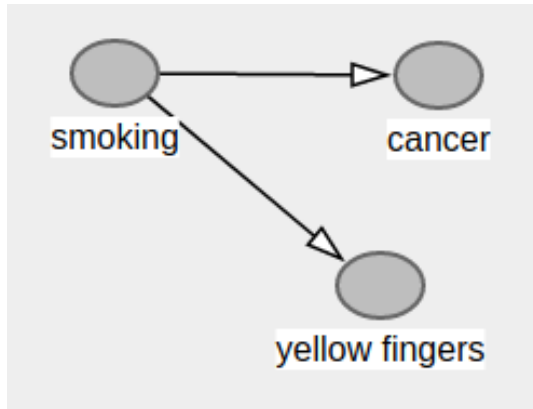


50/50

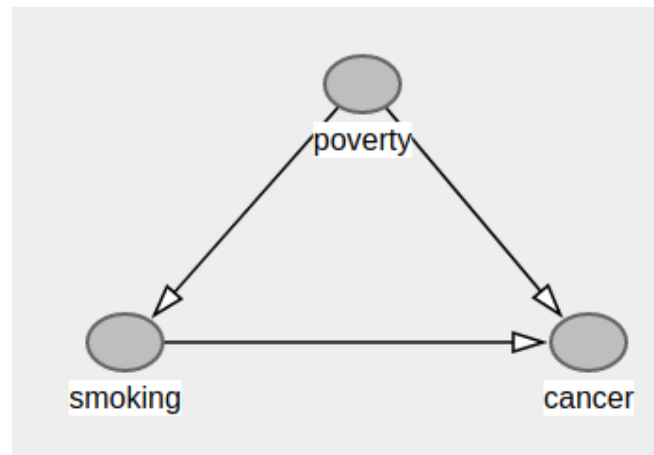


# Observational datasets contain causal relationships which can make or break your ML predictions

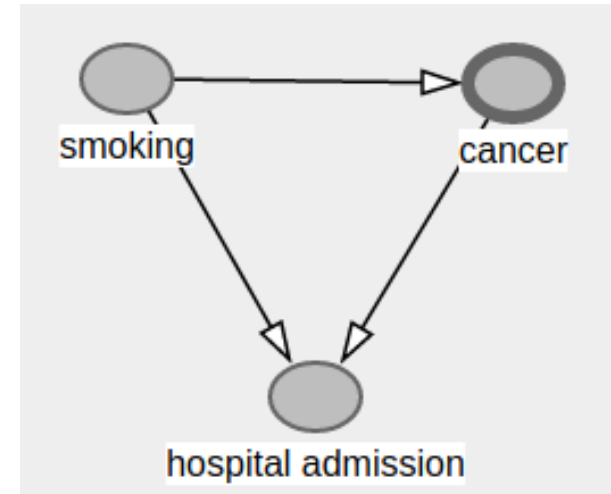
side effect



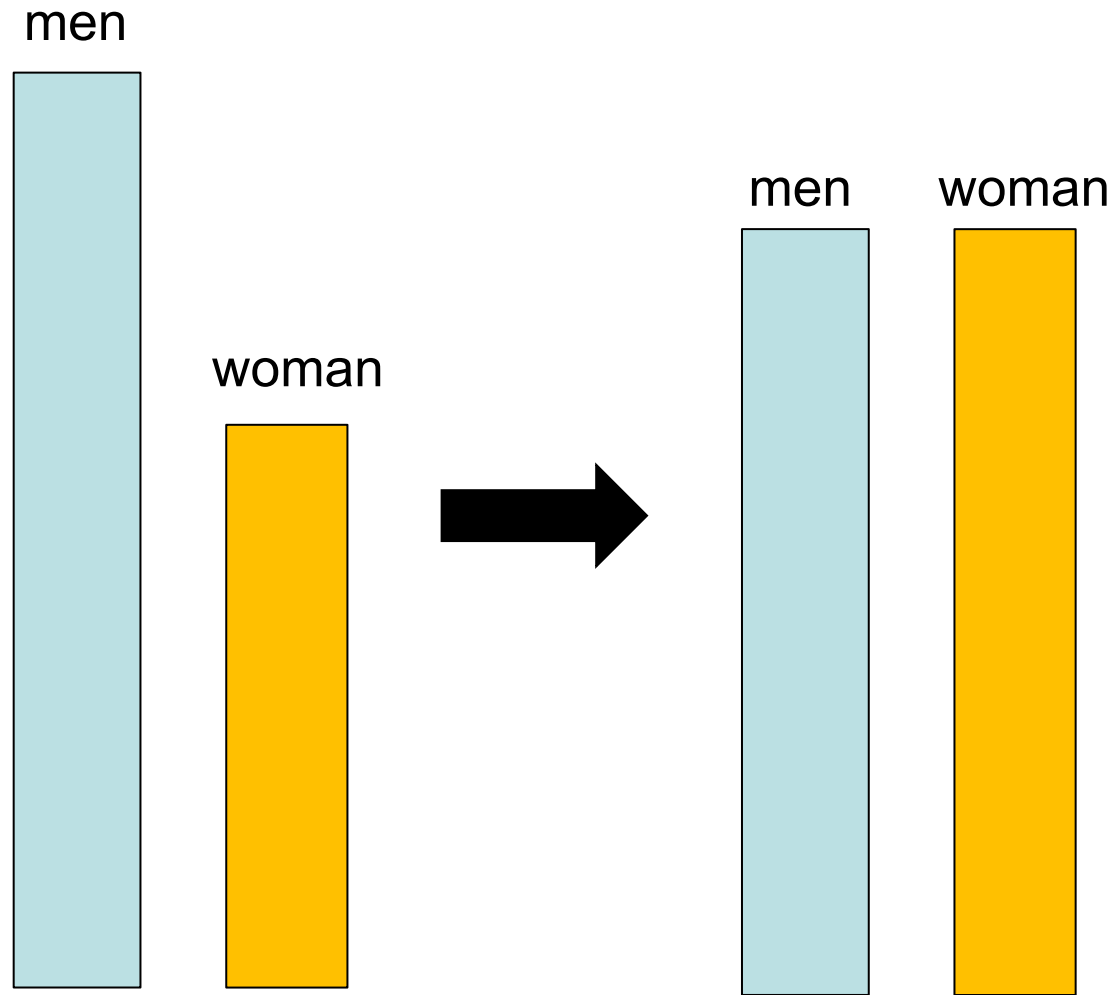
confounder



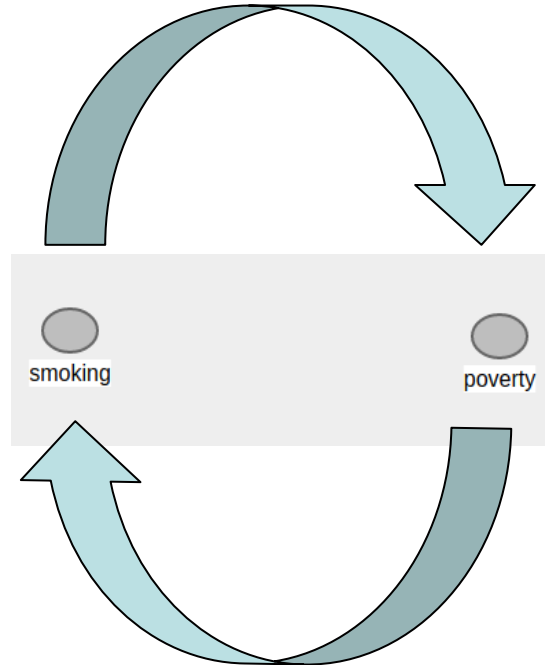
collider



## Reweighting or resampling datasets can help to address confounding and selection issues



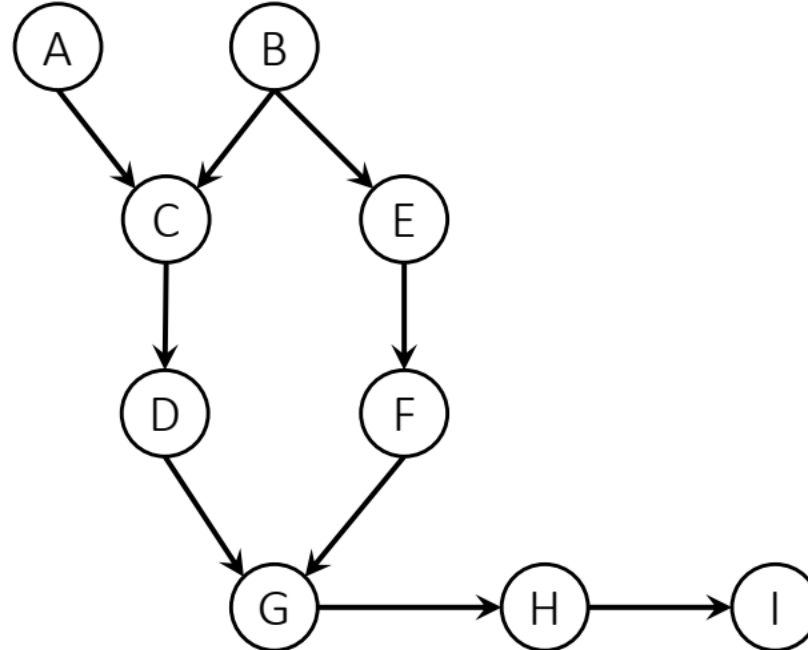
**The real world contains many feedback loops, so the timing of measurements is important**





## The field of causal discovery deals with inferring causal relationships from data

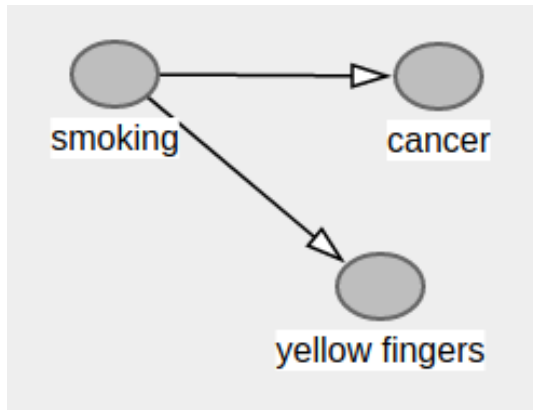
How to learn the arrows if you are given just data A-I ?



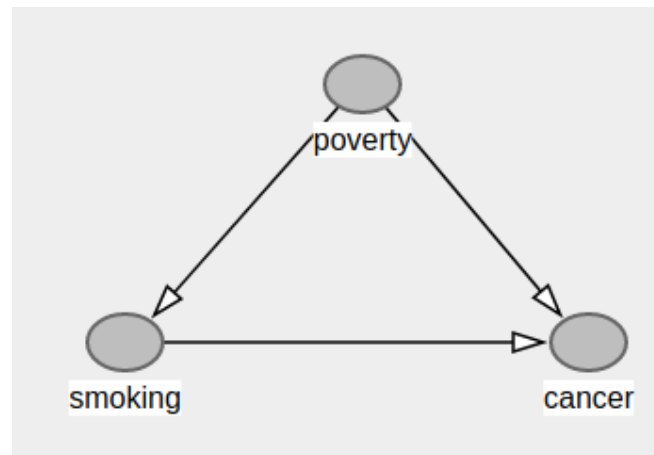
*Check out the review article by Bernhard Schölkopf  
<https://arxiv.org/pdf/1911.10500.pdf>*

# Are there any causality issues in the ML problems that you work with?

side effect



confounder



collider

