IMPERIAL

An Introduction to Causal Inference

Pre-plenary

Ellie Van Vogt
02/09/2024

Introduction

This is a primer talk for Keynote Speaker, Karla Diaz-Ordaz

From Causal Inference to Machine Learning and Back: a Two-way Street towards Better Science
10am Tuesday 3rd September, Oxford Hall

Today's talk will cover the fundamentals of causal inference, with a short intro to machine learning

Causal Inference

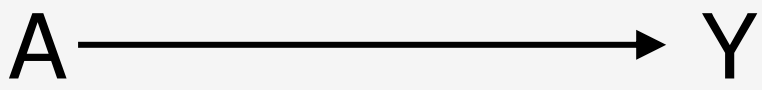
What do we mean by that?

Across many research areas (healthcare, economics, socio-political research...) There is a common quest to understand and learn cause-effect relationships

However, so much research out there has limitations that they can only draw associations

Despite this, researchers would like to think of these associations as causal

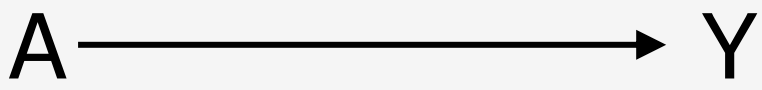
BUT to interpret associations causally, extra steps are required



Causal Language

Probability and statistics allow us to describe (joint) distributions of variables in our dataset

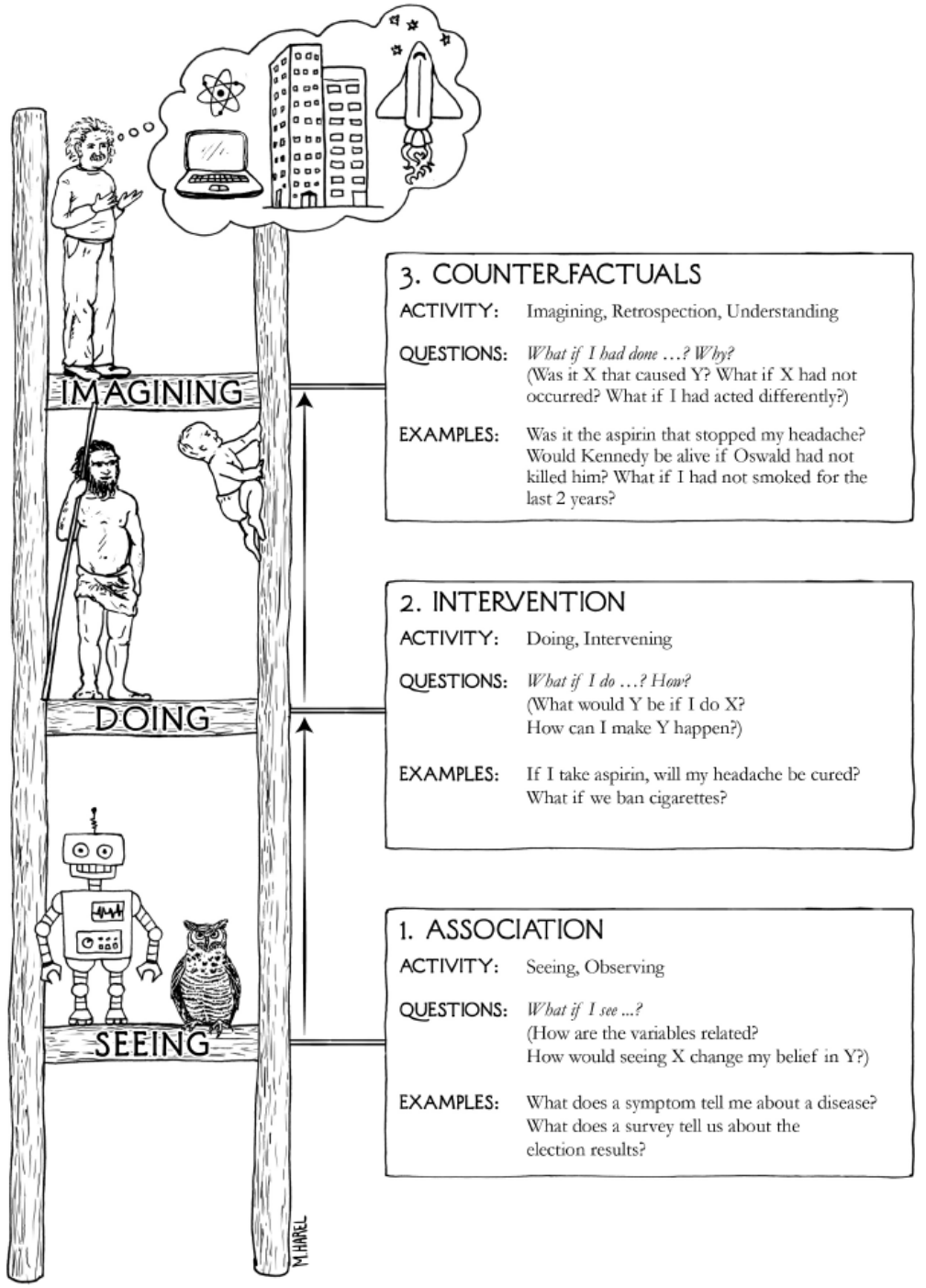
However, this is not sufficient to explain how distributions change in response to external stimuli



The Ladder of Causation

The Book of Why - Pearl

As we go up this ladder, we are asking causal questions in a more sophisticated way

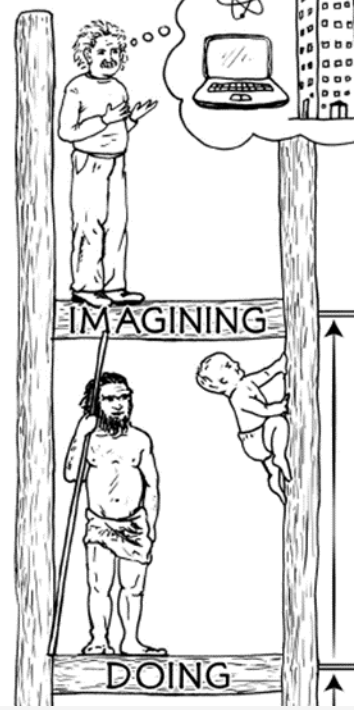


Do-calculus

One set of notation for representing the external manipulation of variables

$Do(A = a)$ shows that we are actively forcing $A = a$

To look at the effect on Y of externally manipulating A we are examining $P(Y|\ do(A))$



Causal graphs

Causal graphs are a common way to visualise relationships between various factors and understand data-generating processes

Graphs can be used for causal effect estimation and causal discovery

The most common causal graph is the directed-acyclic graph (DAG)

Directed – arrows depict causal relationships A -> Y

Acyclic – no feedback loops or double-sided arrows

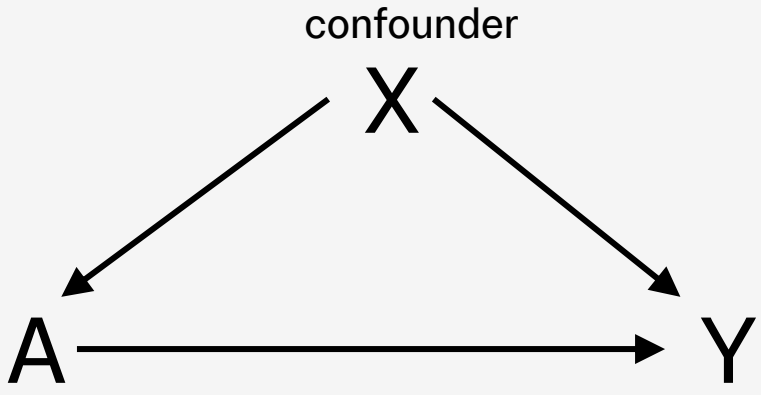
Causal Directed Acyclic Graphs

DAGs

Include both observed and unobserved variables

We can look at the variables that lie on a causal pathway between an action and an outcome (mediators)

We can also identify factors that influence both actions and outcomes (confounders)

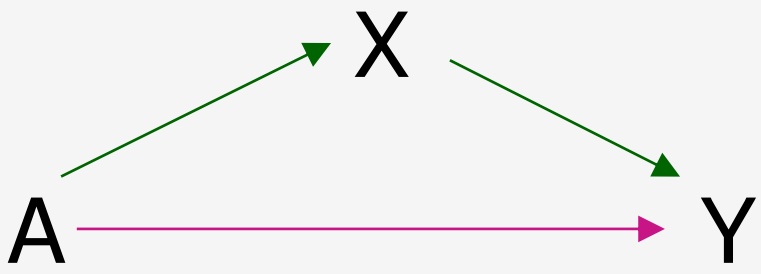


Untangling Mediators

From this DAG, A appears to act on Y directly and indirectly through X

Therefore, the total effect of A on Y can be decomposed into direct and indirect effects

If we are interested in this decomposition, we can use mediation analysis to split these effects



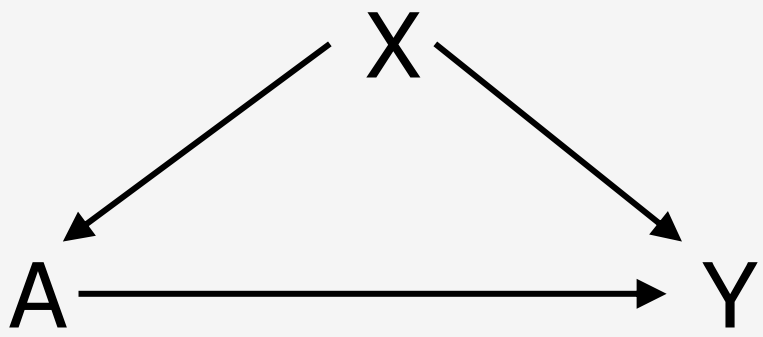
Confounders

X can be a confounder if:

- it is a risk factor for the outcome Y
- It is associated with the exposure A
- X is not on the causal pathway between A and Y

Confounders create backdoor paths between Y and A

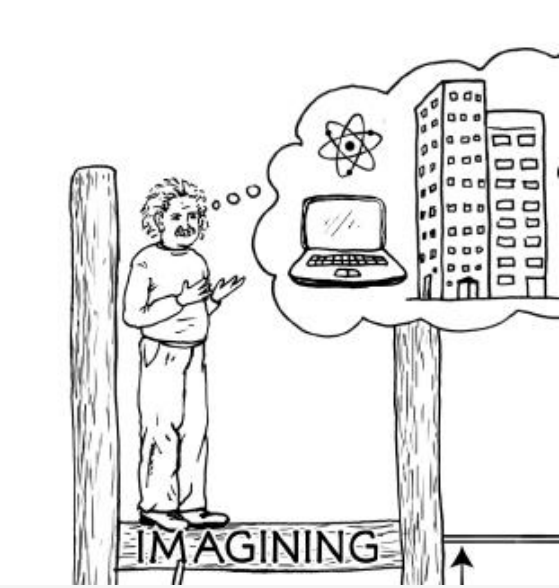
Issue: distortion of the causal effect we seek to estimate, if not properly considered then incorrect conclusions may be drawn



Combating Confounding

Restriction
Restrict population to have identical values of a confounder
Stratification
Divide data into levels of confounding variables before analysis
Multivariate modeling
Including potential confounders in the outcome model
Matching
Pairing treated and untreated observations based on values of confounders

Counterfactuals



In the case of a binary treatment:

Observed: the outcome Y_i receiving treatment $\mathrm{A} = 1$

Counterfactual: the outcome Y_i if instead received $A = 0$

The fundamental problem of causal inference:

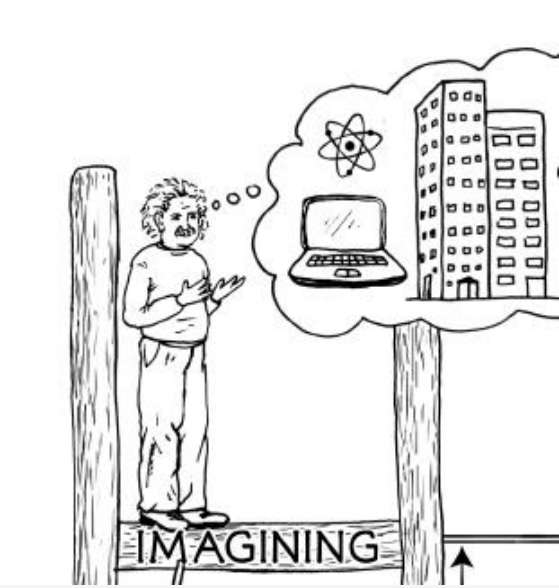
We can only assign one treatment; the counterfactual is always unobserved.

Potential Outcomes

What would the outcome be if we fixed A = 1?

What would the outcome be if we fixed A = 0?

We formalise $Y_i(a)$ as the potential outcome under treatment $A = a$



Causality vs prediction

Prediction
looking at associations in what we have already observed, we can use statistics to describe the joint distributions of observed variables

We are looking at the conditional distribution:
Y|A

Causality
Understanding the change in distributions after external manipulation

The new distribution we are caring about:

Y(A)

NOTE:

In general, $E[Y|A = 1] - E[Y|A = 0] \neq E[Y(1)] - E[Y(0)]$

Research questions

What is my risk of a cardiac event in the next five years, given my characteristics?

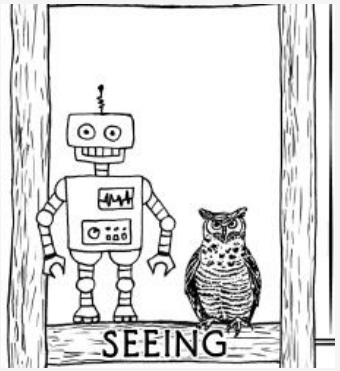
What is the expected claim amount for insured driver X?

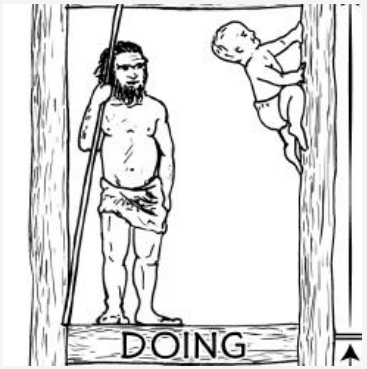
what is the effect of taking statins on risk of cardiac event?

Does quitting smoking improve lung capacity?

What would my salary be if I were a man?

If all school meals had been free, would educational attainment have been better?





Causation vs association example

Educational attainment

Prediction models are often used to give estimates for A-level grades in the UK

During the COVID-19 pandemic, exams could not take place so students were assigned their predicted grades

This raises an issue, now predictions are being used for decision-making – universities, jobs, etc

Has the error of these predictions properly been considered?

To use predictions for a causal purpose, biases in historical predictions would need to be taken into account

Causal Estimands

Average Treatment Effects

A common goal of experimental studies is to estimate the average treatment effect (ATE)

$$ATE = E[Y(1) - Y(0)]$$

To ensure we are targeting a causal effect, several assumptions are required

Assumptions for ATE Estimation

No interference	the treatment assignment of one unit does not affect the outcomes of other units $$Y_i(A) \perp Y_j(A) \; \forall \; i \neq j$$
Consistency	the observed outcome is equal to the potential outcome under that treatment $$Y_i = Y_i(A_i) \; \forall \; i$$
Unconfoundedness / ignorability / exchangeability	there are no unobserved confounders that can influence either the treatment assignment or the outcome $$Y(A) \perp A \mid X$$
Overlap / positivity	for all covariates, there is a non-zero probability of receiving the intervention or control $$pr(A = 1 \mid X = x) \; \in (0,1) \; \forall \; x \in X$$

No interference

For $i \neq j$, individual i's potential outcome under $A_i = a$ does not depend on individual j's treatment A_j

Also known as "no interaction between units"

Examples of violations:

- Vaccine studies: infection dependent on proportion of vaccinated people in household/community
- Education: individual student interventions might influence other students in the same class

Possible relaxation, vectors of units

Consistency

$Y_i(a) = Y_i$ for an individual i with $A_i = a$

The observed outcome Y_i equals the potential outcome $Y_i(a)$ if individual i receives treatment a

If the possible treatments are $A_i \in \{0,1\}$, the observed outcome Y_i is often written as

$$Y_i = A_i \cdot Y_i(1) + (1 - A_i) \cdot Y_i(0)$$

Exchangeability

$Y(a) \perp A \mid X \; \forall \; a$

Conditional on the variables X, the treatment is independent of each potential outcome

Exchangeability relates to the potential outcome $Y(a)$, not the observed outcome Y

$Y \perp A \mid X$ would imply A has *no causal effect* on Y given X

In causal inference, we must control for confounders X that influence both treatment assignment A and potential outcomes $Y(a)$

If assignment is completely at random, then exchangeability holds with X={∅}.

When interventions are not randomized (e.g., observational studies), this assumption is more challenging to assess.

Positivity

$0 < P(A = a \,|X = x) < 1 \; \forall \, a, x$

Conditional on the covariates X, there is a probability greater than zero of being assigned to each of the treatment levels

We need to choose a set of variables X that satisfies exchangeability.

On the selected variables X, the positivity assumption must also hold.

In RCTs, positivity is usually guaranteed by design for the randomised treatment of interest.

Satisfied these assumptions?

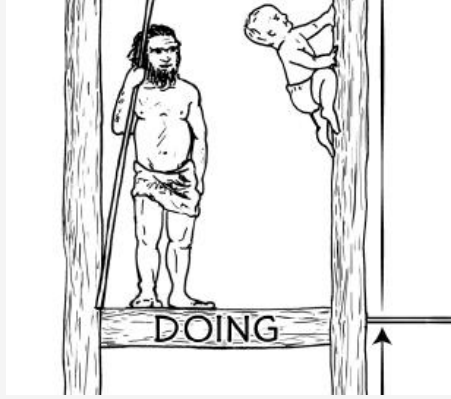
$E[Y(1) - Y(0)] = E[Y(1)|A = 1] - E[Y(0)|A = 0]$ by exchangeability

$E[Y(1) - Y(0)] = E[Y|A = 1] - E[Y|A = 0]$ by consistency + no interference, positivity ensures both terms are defined

That is, under our assumptions, the difference in average outcomes under $A = 1$ vs $A = 0$ is a consistent estimator for $E[Y(1) - Y(0)]$

Gold standard: RCTs

Assumptions are satisfied by design



Random treatment assignment

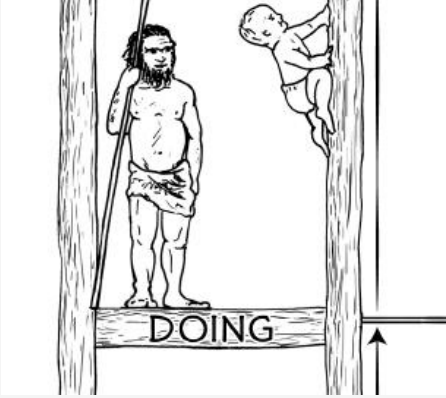
Treatment Group

New treatment

Control Group

Usual treatment / placebo

Are RCTs the only option?



RCTs are costly and time-consuming, and there an abundance of observational data

Not all causal questions can be answered through RCTs
- ethical/practical considerations

Even the best-run RCTs can have things going wrong
- treatment switching/discontinuation

Post-randomisation events may induce confounding

We still need causal methods to address these issues that come up in trials

Estimating ATE in non-perfect RCTs / Real World Data

Inverse probability weighting

In RCTs, people can deviate from the protocol

Study withdrawal, non-adherence to treatment, treatment switching

Protocol deviation can affect valid estimation of the ATE

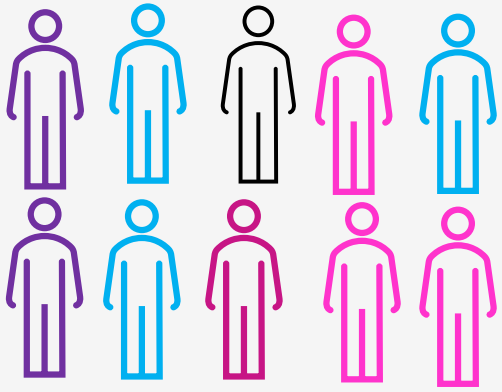
In observational settings, treatment might not be equally available or taken up, here, we can use the propensity score:

$$g(X) = pr(A = 1|X)$$

Now define a weight $\frac{1}{g(X)}$

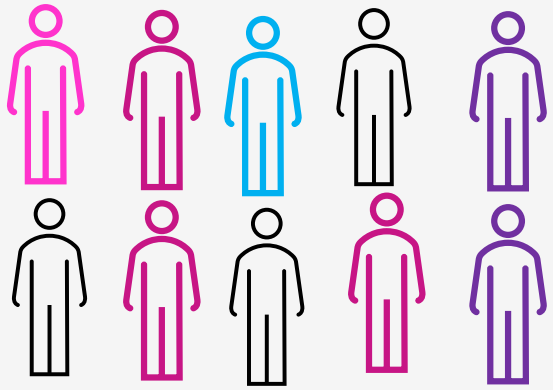
This increases the weight of underrepresented observations in our dataset, improving balance.

Treatment Group

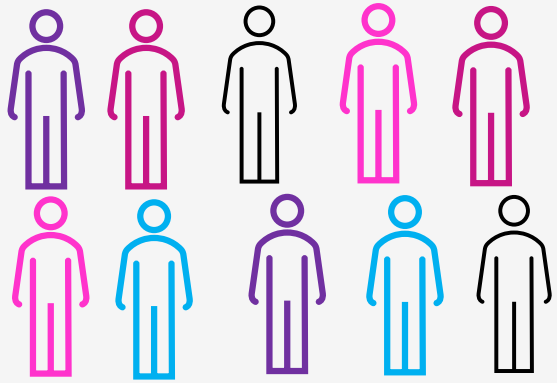


Observational setting
The people who have access to the treatment are not the same as those who are in the control

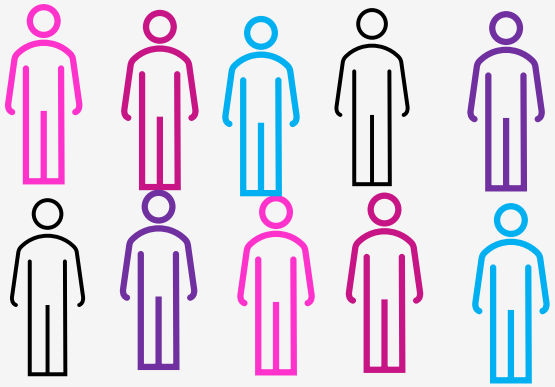
Control Group



Treatment Group



Control Group



Regression models for ATE

In RCT settings, we can use our favourite multivariate regression models and marginalisation to get direct causal effects of A on Y, not conditional on other covariates

In observational settings, we can use tools such as G-computation to estimate the ATE

Other candidates include instrumental variable analysis, mediation analysis...

Other causal targets are also available

CATEs – conditional ATEs on covariates

Longitudinal ATEs

Machine Learning
Associative and predictive

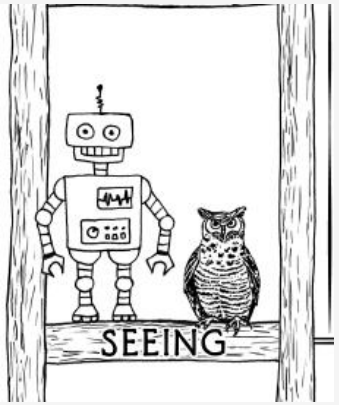
ML is focused on making predictions based on previous inputs of data to a model

Unsupervised learning: learning patterns in data without labels
clustering

Supervised learning: learning on prior collected labelled data how to predict labels on future data

Spotify recommendations

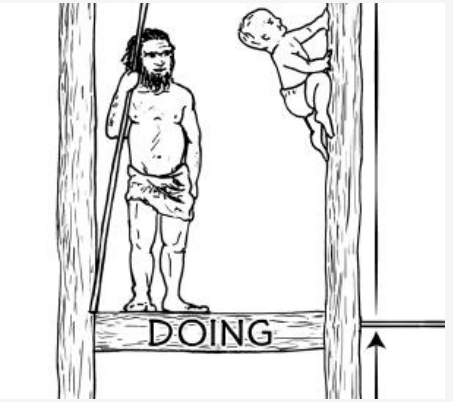
How many ICU beds will be filled next week?



A step towards causal inference:

Reinforcement learning

An action-reward cycle motivates an adaptive model



Common ML Models
Random forests (RF)

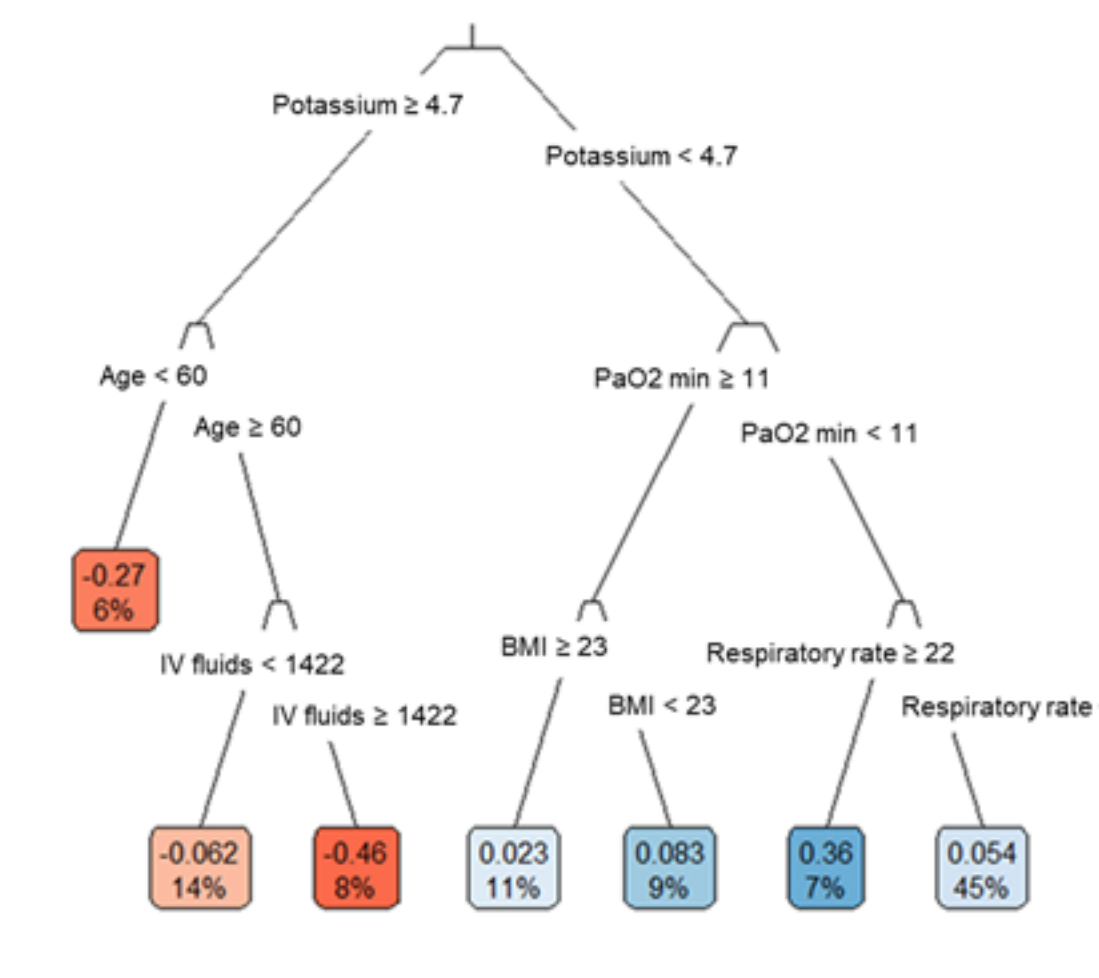
Ensembles of decision trees

Define a metric for splitting create many trees and average

Classification or regression

Explainability:

Permutation variable importance



Common ML models

Generalised Additive Models (GAMs) and ensemble methods

Similar to splines

Sum of smooth predictors for each covariate

Good for capturing non-linear relationships while keeping interpretability

More on this in the Keynote

Even more general:

Use many ML predictors as base learners, feed into a another model to combine predictors

Idea: cons of base learners cancel each other out

Machine Learning + Causal Inference?

More on this tomorrow in the Keynote Talk!

References

Hernán M. A. and Robins J. M. (2020). Causal inference: What if. Boca Raton: Chapman & Hall/CRC.

Pearl, J. (2009). Causal inference in statistics: An overview. Statistics Surveys, 3, 96-146.

Ogburn, E. L. and VanderWeele, T. J. (2014). Causal diagrams for interference. Statistical Science, 29(4), 559-578.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies.Journal of Educational Psychology,66(5), 688.

Cox, D. R. (1958). Planning of experiments. Wiley.

Pearl, J. (1995) Causal diagrams for empirical research., Biometrika, 82:669–710.

IMPERIAL

Thank you

Causal Inference Pre-plenary
02/09/2024

IMPERIAL