WORKSHOP ON

CAUSAL DATA SCIENCE

22 OCT | 9 - 2 PM

University of Barcelona

LEARN THE BASICS AND GET HANDS-ON WITH REAL EXAMPLES

With Jordi Vitrià (UB), Jordi Mur (UB)
and Enrique Mora (Nestle)

Agenda

Time	Contents
09:00	Arrival and registration
09:30	Jordi Vitrià: **Welcome** and **Introduction to Causal Inference**
11:00	*Coffee break - First floor*
11:30	Jordi Mur-Petit: **Potential Outcomes framework and Propensity Score methods**
12:30	Enrique Mora: **A practical introduction to *DoWhy* — an end-to-end library for causal inference**
13:30	Hands-on work and discussion
13:55	Closing remarks
14:00	*Lunch - First floor*

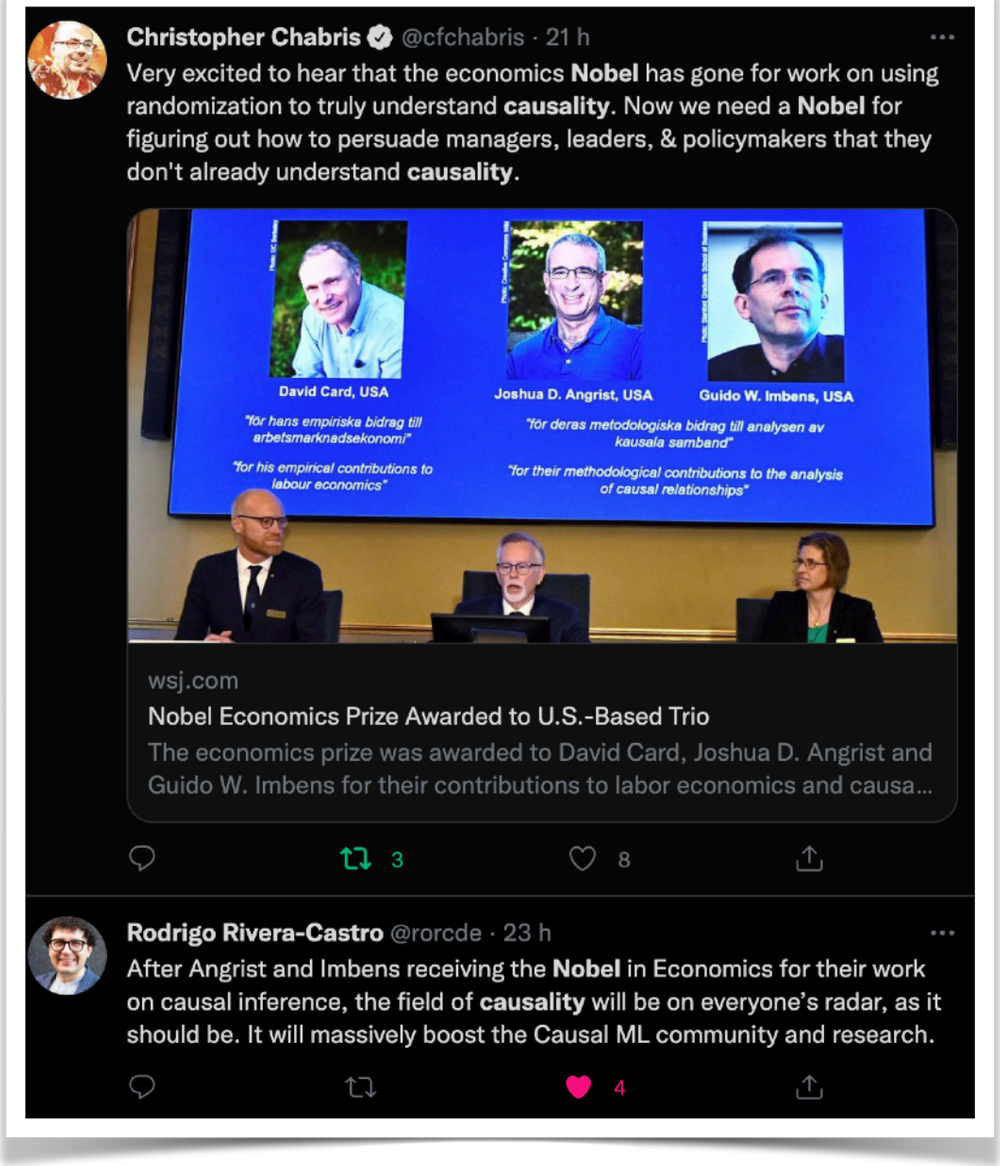
https://github.com/santboia/Curs_CausalDS_oct21

UNIVERSITAT DE BARCELONA

wifi.ub.edu

Identificador	agnlis.tmp
Contrasenya	usuo55

News! Causality is a hot topic!



Two weeks ago Nobel Prize in Economics brought joy to researchers working w/ quasi-experimental designs **to establish causality in topics that cannot be easily studied in experiments**.

Data Science Facts

Applicable to all your projects — We predict to act

Statistics and ML — **Predictive models can be accurate without being correct.**

Empirical risk is not the only feature to optimize. — Machine Learning

Sometimes — **We can extract causal information from observational data.**

Causal inference techniques are mature enough for addressing real-world problems.

The data science toolbox must be updated with causal inference techniques.

Ingredients: DAGs, Identifiability, Non parametric SCMs, Potential Outcomes, Stratification, Matching, Double ML.

Jordi Vitrià, October 2021

UNIVERSITAT DE BARCELONA

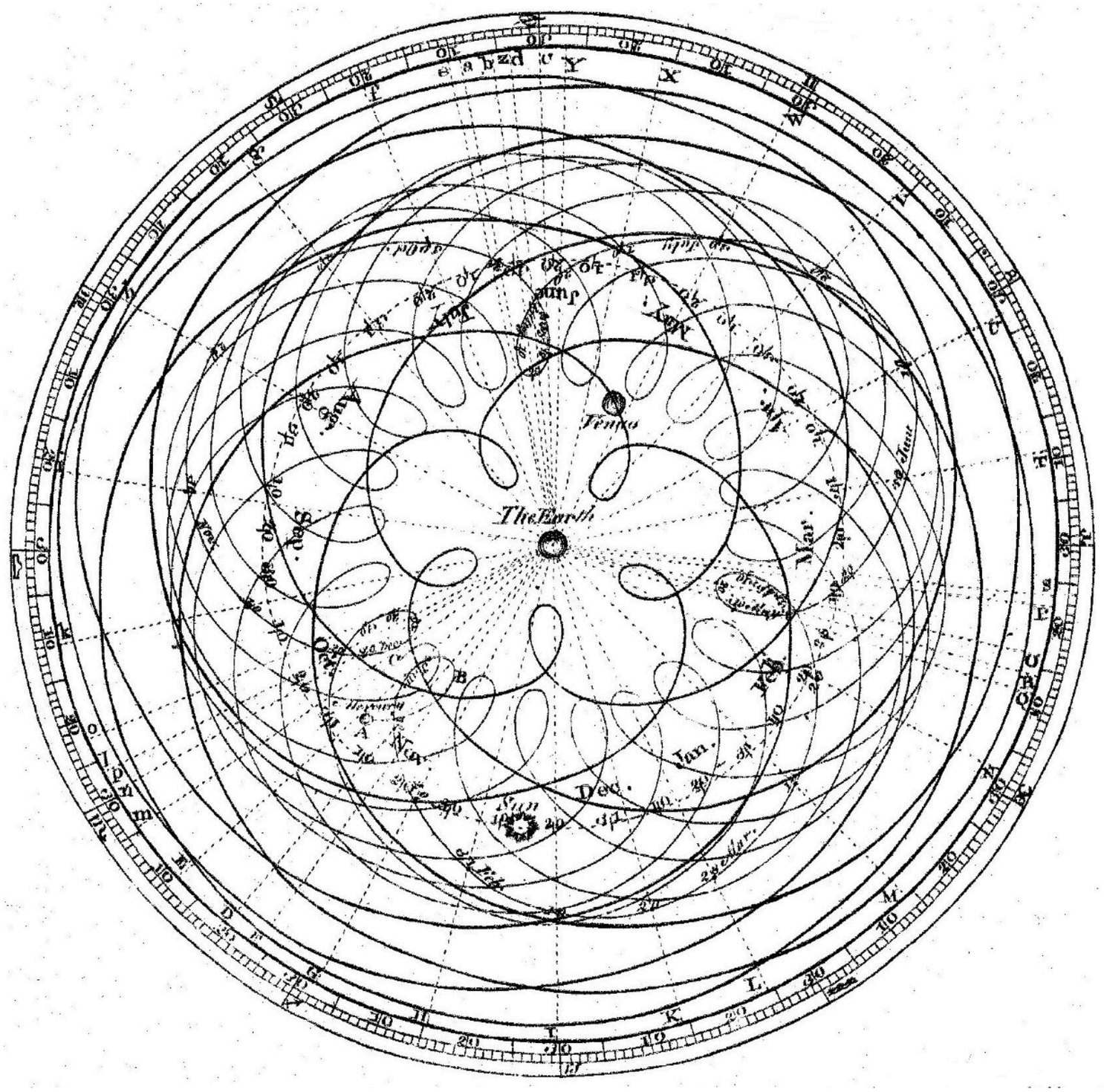
Predicting observations vs predicting interventions.

Basic concepts of causality for ML practitioners.

Can we predict what we see in the sky?

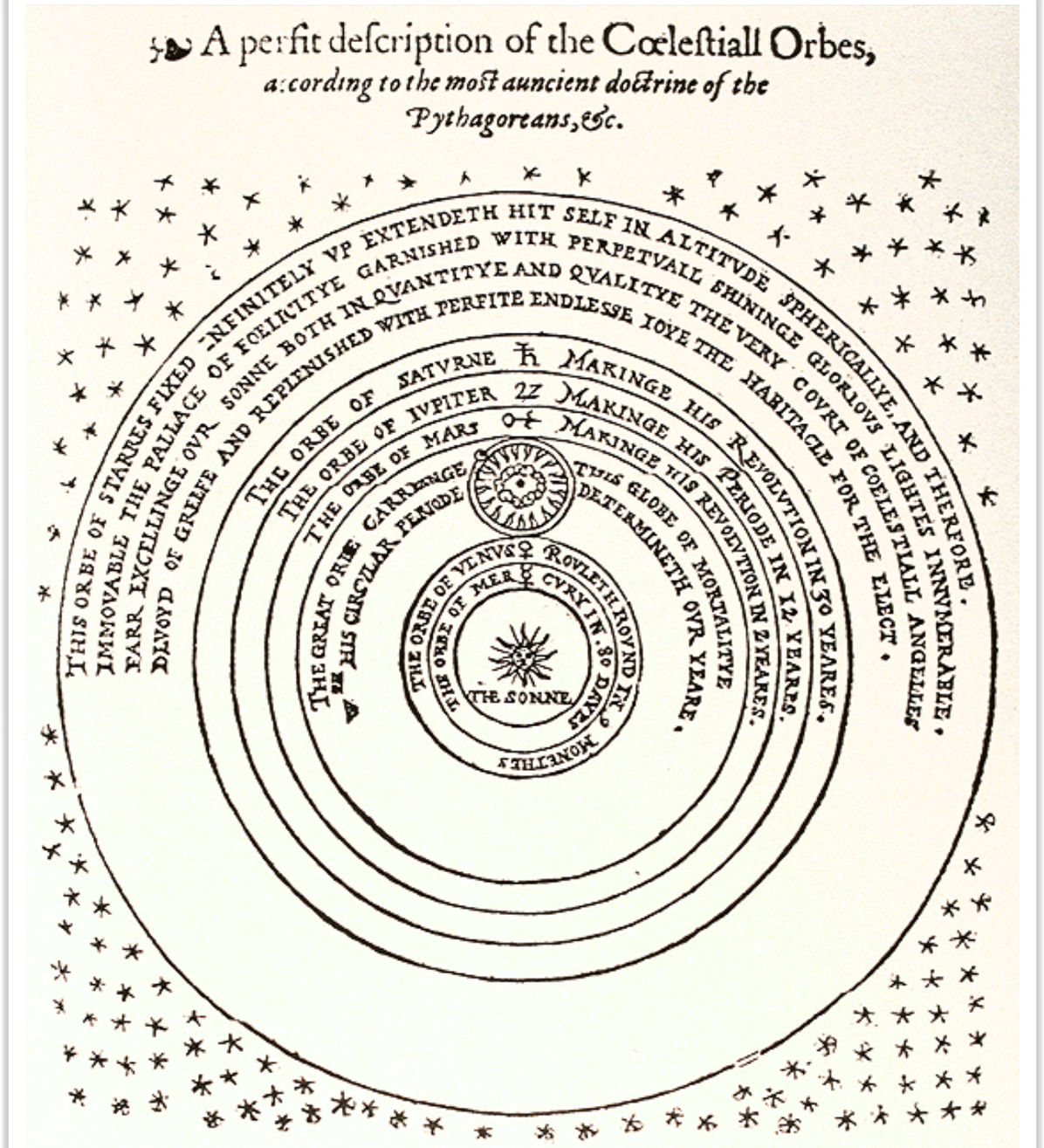
Prediction problem = Getting data + Building a data-driven model

Ptolemaic model (circular orbits, geocentric)



Can we predict what we see in the sky?

Copernican model (heliocentric, fewer causes)



Mikołaj Kopernik (1473–1543): heliocentric model of the solar system.
He is famous for his heliocentric model of the solar system, Kopernik argued for replacing the geocentric model, because the heliocentric model was more “harmonious”.
The model was not more accurate than the geocentric model. The heliocentric model made exactly the same predictions as the geocentric model.

Ptolomeus and Kopernik build models with
high predictive power. But they were false!

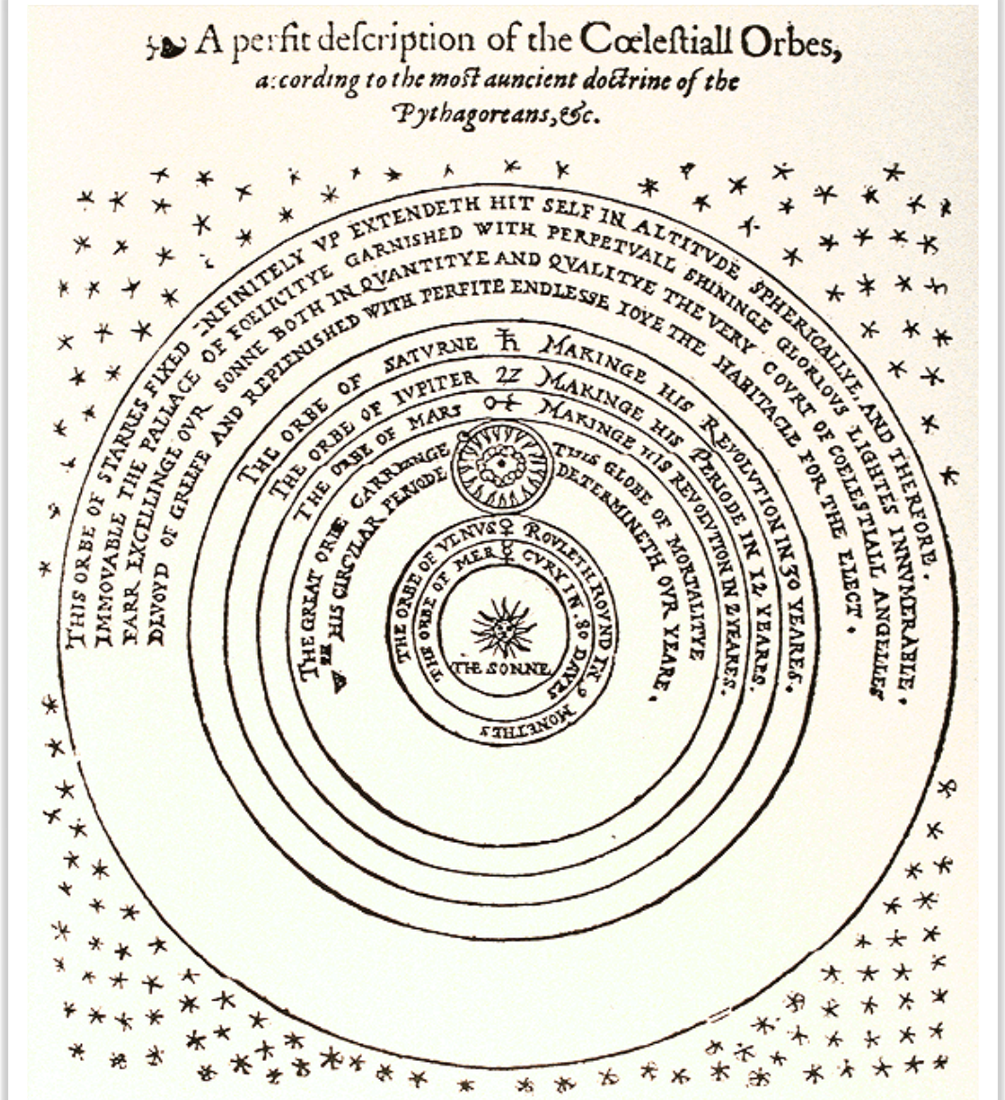
It was not false in the "predictive" (statistical) sense,
but in the "interventional" (scientific/causal) sense.

Statistical inference and machine learning methods are designed to **predict observations**.

They are based on modeling data to answer **associative questions**.

What can I say about Y given that I have observed X?

Copernican model (heliocentric, harmony = fewer causes)



Intervention

What about moving Mars?

Harmony = few causes, but the right ones!

Models can be accurate without being correct

FACT 1

Data does not speak by itself.

Given a dataset, there are multiple predictive models that can be used for prediction.

Why is causality a hot topic now in Data Science?

Artificial Intelligence, Machine Learning are at the heart of the **digital revolution** that citizens, businesses and administrations are experiencing in recent times.

During the first years of this revolution, the great application of these technologies has been **to transform SERVICES into PREDICTION problems**, such as:

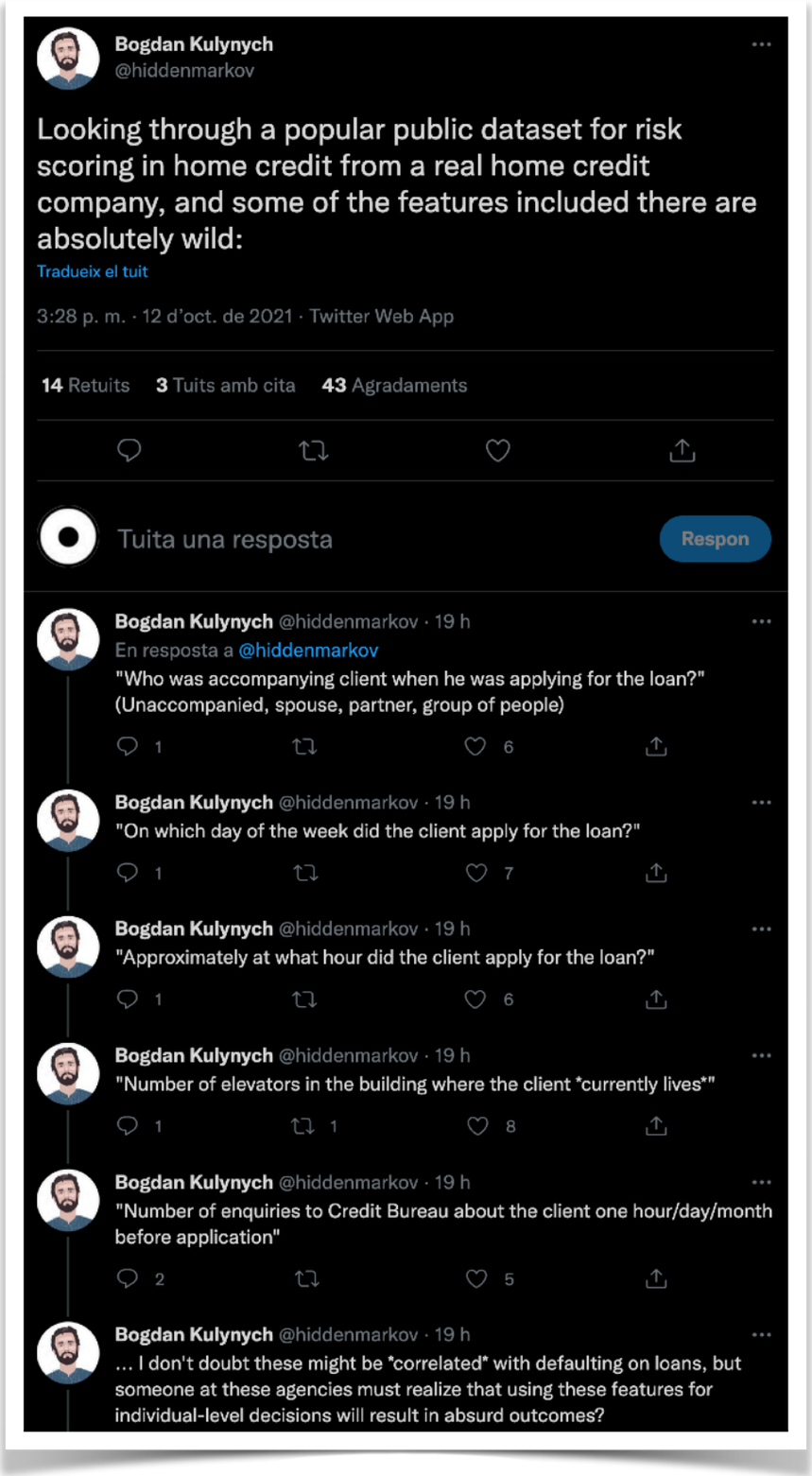
- banks have replaced employees with AI systems that determine whether or not you will get a loan based on your financial history.
- virtual bookstores recommend the books we might like based on a prediction based on our shopping history.
- online travel platforms recommend us the hotels best suited to our profile based on our previous trips.
- hospitals use AI programs to determine a patient's diagnosis based on their medical data.

Artificial Intelligence and Machine Learning are being used widely to predict events and to inform decisions but, in some cases, **not in the right way**!

- Prediction accuracy is not the only concern.
- Some predictions do not fulfill ML assumptions.
- Some decisions must take into account the effect of **interventions**!

It is necessary to revise the set of DS good practices.

Example: Uncontrolled use of spurious correlations



Which variables must we consider for prediction?

Which variables must we consider for prediction?

Given an i.i.d. set of samples $\{(\mathbf{x}^i, y^i)\}$, which features $(x_1, \ldots, x_n)$ from $\mathbf{x}$ must I consider for predicting $\mathbb{E}(y \mid \mathbf{x})$?

Which variables must we consider for prediction?

It depends on the needs:

- **I want to minimize the Empirical Risk.** $R_{emp}(f) = \frac{1}{n}\sum_{i=1}^{n} L(f_\theta(\mathbf{x}_i), y_i)$
- I want to maximize robustness against changes in $p(\mathbf{X}, Y)$.
- I want to maximize robustness against adversarial attacks.
- I want to be able of explaining my predictions.
- I want to measure and mitigate unwanted biases (discrimination).
- I want to use the prediction to inform a decision that can change $p(\mathbf{X})$.
- Etc.

All these considerations involve causal thinking.

Causal Inference is designed to **predict interventions**.

It is based based on our knowledge about the **structure of the world** + modeling specific **data** relationships.

THE NEED FOR CAUSAL THINKING

PREDICTIVE systems are useful and suitable for many tasks, but lately their limitations and shortcomings have begun to become apparent when they have been used to make DECISIONS:

- Predictive systems have very limited utility during decision making if the goal of the DECISION is to change the PREDICTION.

  For example, when a company predicts the possible loss of a customer (**churn prediction**), the decision on the best course of action to avoid such abandonment should not be based on the predictive model!

Intervention

THE NEED FOR CAUSAL THINKING

These systems can show a lack of TRANSPARENCY when used to make DECISIONS that affect people.

For example, any customer of a bank has the right to know the reasons why an automated system has denied him a loan, and in most cases predictive systems do not provide this information. **The best explanations are contrastive (counterfactual) explanations!**

THE NEED FOR CAUSAL THINKING

These systems can DISCRIMINATE against certain groups when they are used to make DECISIONS that affect people, due to the bias in the data that has been used to train them.

For example, a system of support for hiring people by a company, which predicts the best candidates for a job among those who have applied, can discriminate against women because of the existence of historical data that show a predilection for men when it comes to taking a job. **Discrimination is a causal effect, not a correlation.**

In all these cases we need to complement the data and tools of STATISTICS/MACHINE LEARNING with CAUSAL INFERENCE techniques that allow us to integrate in the solution of the problem the knowledge associated with the problem treated and also **measure the impact of the different options on which we can decide**.

FACT 2

Causal inference is the right way to predict interventions but it is also necessary when predicting observations.

Some prediction features can only be understood from a causal point of view.

Defining Causation

Definition: In the "interventionist" definition of causality, we say that an event A causes another event B if we observe a difference in B's value after changing A, keeping everything else constant.

A is the **cause** or **treatment** and B is the **outcome**.

An **intervention** refers to any action that actively changes the value of a treatment variable independently of the data generation process.

Defining Causation

"While keeping other variables constant" (or "controlling") may seem intuitive, it is unclear about which variables to include.

We can obtain a more precise definition by utilizing the second key concept of causal reasoning, **counterfactuals**.

Defining Causation

For any intervention, we can imagine two worlds, identical in every way up until the point where a some “treatment” occurs in one world but not the other. Any **subsequent difference in the two worlds is then logically, a consequence of this treatment**.

The first one is the observed, **factual** world, while the second one is the unobserved, **counterfactual** world.

The counterfactual world, identical to the factual world except for the intervention, provides a precise formulation to the “keeping everything else constant” maxim. **The value a variable takes in this world is called a counterfactual value.**

Counterfactual Value: The (hypothetical) value of a variable under an event that did not happen.

Defining Causation

Putting together counterfactuals and interventions, the **causal effect** **of an intervention** can be defined as the **difference between the observed outcome after an intervention and its counterfactual outcome**.

Defining Causation

Putting together counterfactuals and interventions, the **average treatment effect (ATE) of an intervention** can be defined as the **mean of the difference between the observed outcomes after an intervention and their counterfactual outcomes**.

The Gold Standard for measuring ACE: Randomized Experiment

Person	T	$Y_{T=1}$	$Y_{T=0}$
P1	1	0.4	0.3
P2	0	0.8	0.6
P3	1	0.3	0.2
P4	0	0.3	0.1
P5	1	0.5	0.5
P6	0	0.6	0.5
P7	0	0.3	0.1

Observed Outcome

Counterfactual Outcome

This may seem straightforward, but the fundamental challenge is that this calculation requires taking the difference between an observed outcome and a counterfactual that we cannot observe.

The Gold Standard for measuring ATE: Randomized Experiment

Person	T	$Y_{T=1}$	$Y_{T=0}$
P1	1	0.4	0.3
P2	0	0.8	0.6
P3	1	0.3	0.2
P4	0	0.3	0.1
P5	1	0.5	0.5
P6	0	0.6	0.5
P7	0	0.3	0.1

Observed Outcome

Counterfactual Outcome

In general, $\mathbb{E}(Y_{T=1} - Y_{T=0}) \neq \mathbb{E}(Y_{T=1}) - \mathbb{E}(Y_{T=0})$

The Gold Standard for measuring ACE: Randomized Experiment

Causal reasoning took a major advance in the early twentieth century when Fisher discovered a conceptually straightforward way to conduct an intervention such that the there is no systematic difference between the treated and untreated groups.

We simply gather one large population of people and randomly split them into two groups ($G = 0$ or $G = 1$), one of whom will receive the treatment and the other will not.

By randomly assigning individuals to receive or not receive treatment, we ensure that, on average, there is no difference between the two groups.

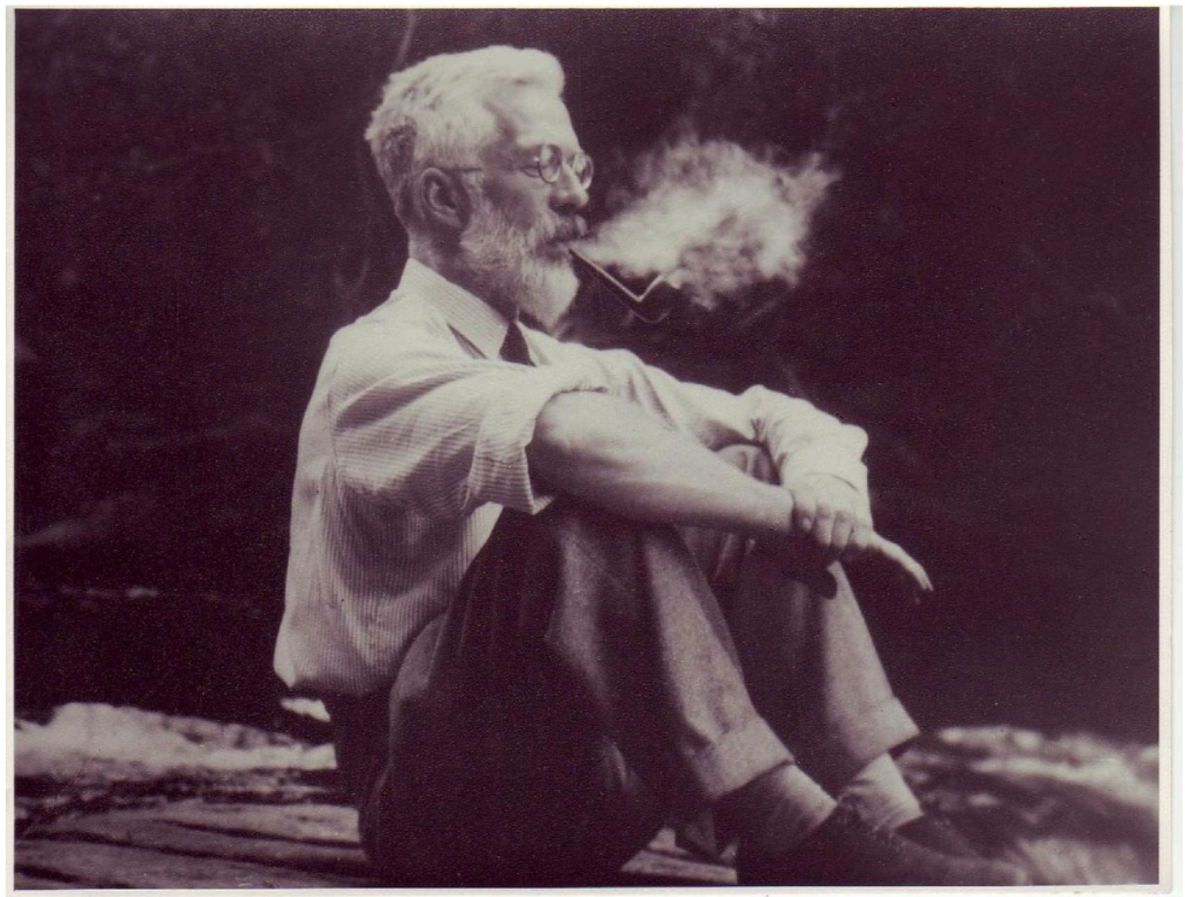
But experiments can be expensive or non-ethicals....

Can we get an answer to these questions by using observational data?

Why the Father of Modern Statistics Didn't Believe Smoking Caused Cancer

By Ben Christopher

Share Tweet



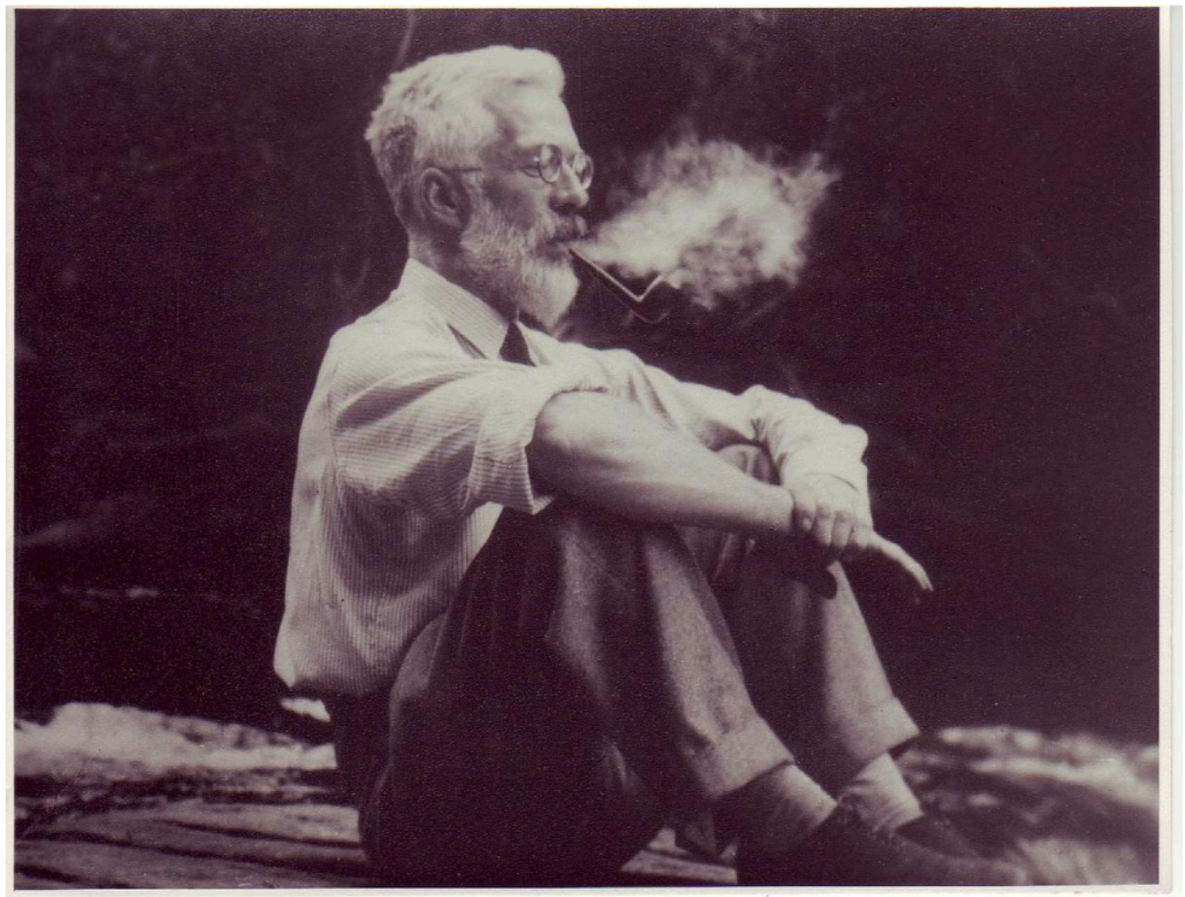
Ronald A. Fisher, father of modern statistics, enjoying his pipe.

Can we get an answer to these questions by using observational data?

Why the Father of Modern Statistics Didn't Believe Smoking Caused Cancer

By Ben Christopher

Share Tweet



Ronald A. Fisher, father of modern statistics, enjoying his pipe.

Sometimes…

Causal Inference Process

Asking a causal query
Gathering data and knowledge
Building a causal model
Identifying the causal query
Estimating the causal effect
Validating the result

Causal Data Science

Data science tasks:

- **Description** is using data to provide a quantitative summary of certain features of the world.
  Descriptive tasks include, for example, computing the proportion of individuals with diabetes in a large healthcare database and representing social networks in a community.

- **Prediction** (or association) is using data to map some features of the world (the inputs) to other features of the world (the outputs) in an observational setting.

- **Causal Inference** is using data to predict certain features of the world if the world had been different.
  An example of causal inference is the estimation of the mortality rate that would have been observed if all individuals in a study population had received screening for colorectal cancer vs. if they had not received screening.

OBSERVATIONAL VS INTERVENTIONAL DISTRIBUTION

Let's suppose we have i.i.d. data sampled from some joint $f(x, y, z, \dots)$. Say we are interested in how variable y behaves given x.

Machine Learning setting: What is the distribution of Y given that I **observe** variable X takes value x?

What we usually estimate in supervised machine learning is a conditional distribution $p(Y \mid X)$.

Causal setting: What is the distribution of Y if I were to **set** the value of X to x?

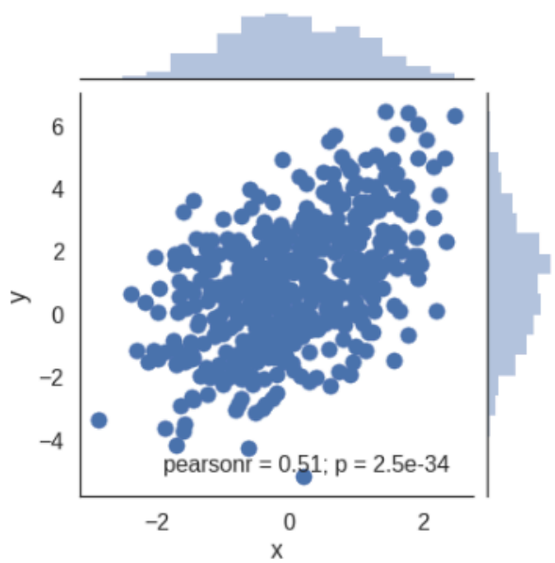
This describes the distribution of Y I would observe if I intervened in the data generating process by artificially forcing the variable X to take value x, but otherwise simulating the rest of the variables according to the original process that generated the data, $p(Y \mid \text{do}(X))$

OBS $p(y \mid x)$

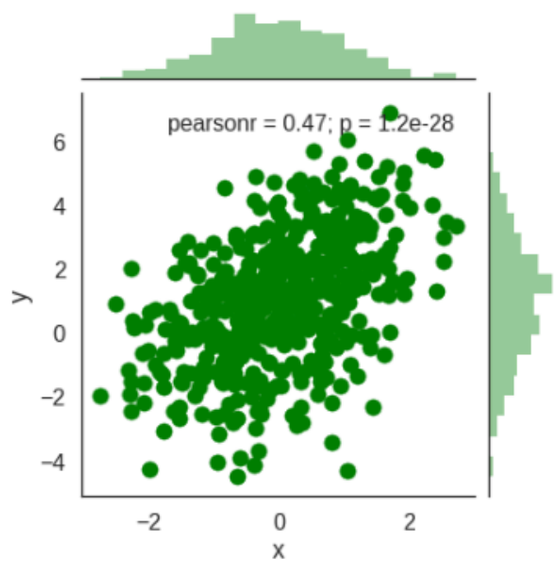
INT $p(y \mid do(x))$

OBS and INT are not generally the same!
Let's consider three generative models
corresponding to the same $p(x, y)$

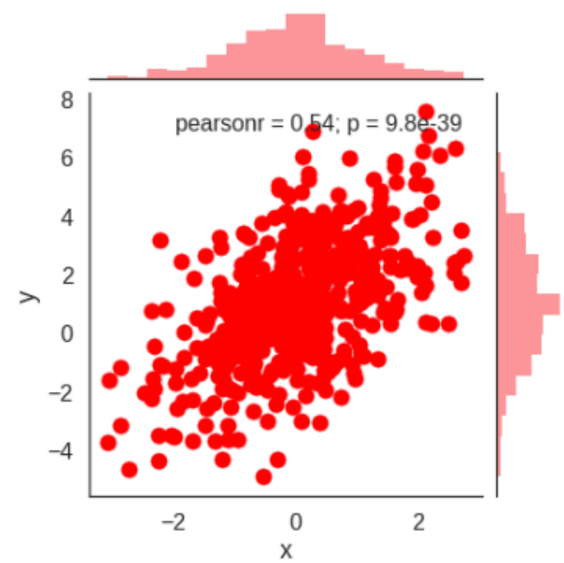
```
x = randn()
y = x + 1 + sqrt(3)*randn()
```




```
y = 1 + 2*randn()
x = (y-1)/4 + sqrt(3)*randn()/2
```




```
z = randn()
y = z + 1 + sqrt(3)*randn()
x = z
```

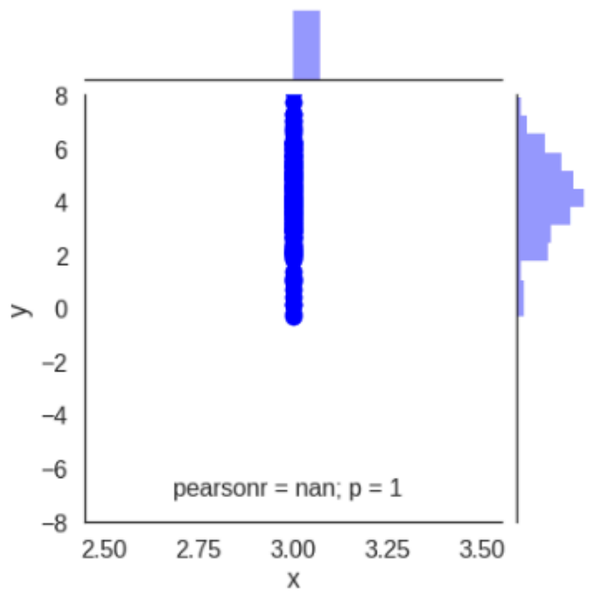


OBS $p(y \mid x)$

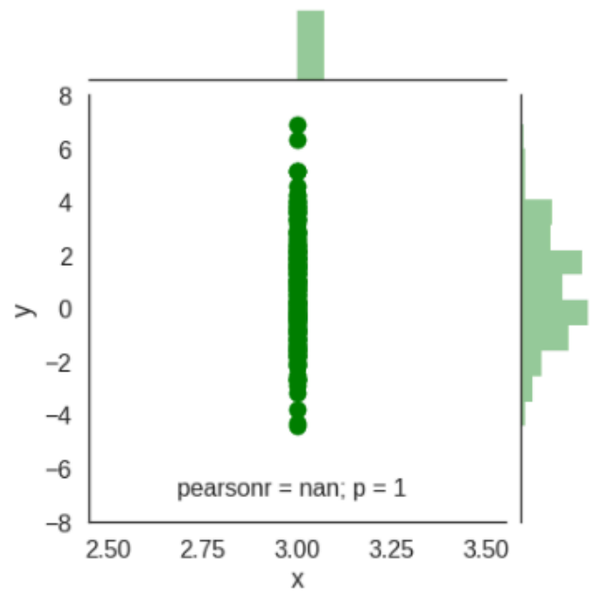
INT $p(y \mid do(x))$

OBS and INT are not generally the same!
Let's consider three generative models
corresponding to the same $p(x, y)$

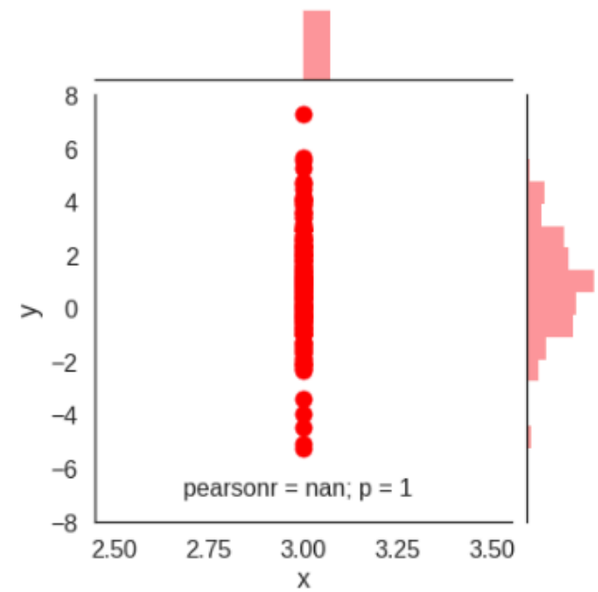
```
x = randn()
x = 3
y = x + 1 + sqrt(3)*randn()
x = 3
```




```
y = 1 + 2*randn()
x = 3
x = (y-1)/4 + sqrt(3)*randn()/2
x = 3
```




```
z = randn()
x = 3
x = z
x = 3
y = z + 1 + sqrt(3)*randn()
x = 3
```



OBS $p(y \mid x)$

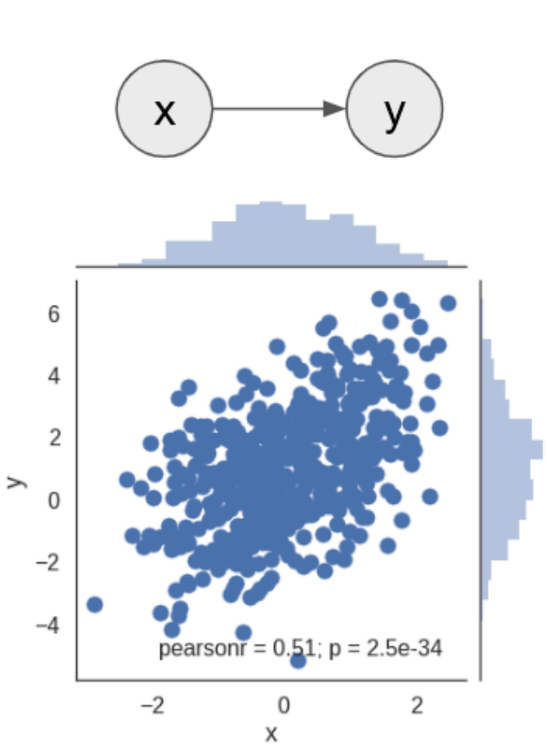
INT $p(y \mid do(x))$

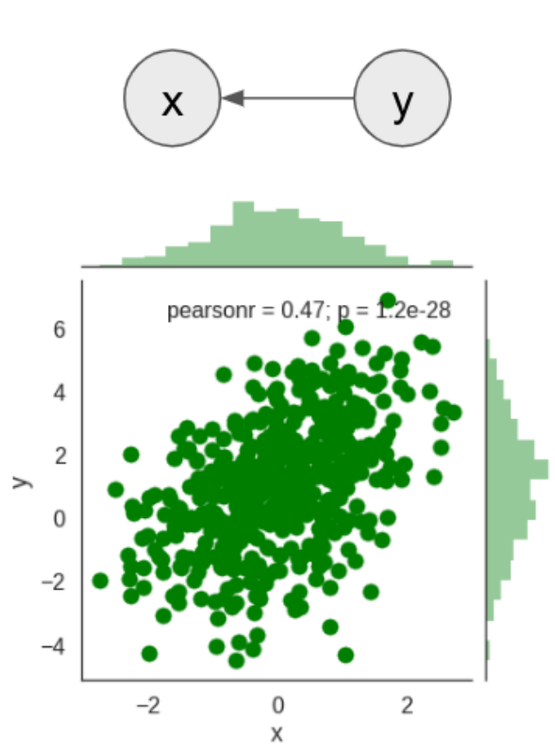
OBS and INT are not generally the same!
Let's consider three generative models
corresponding to the same $p(x, y)$

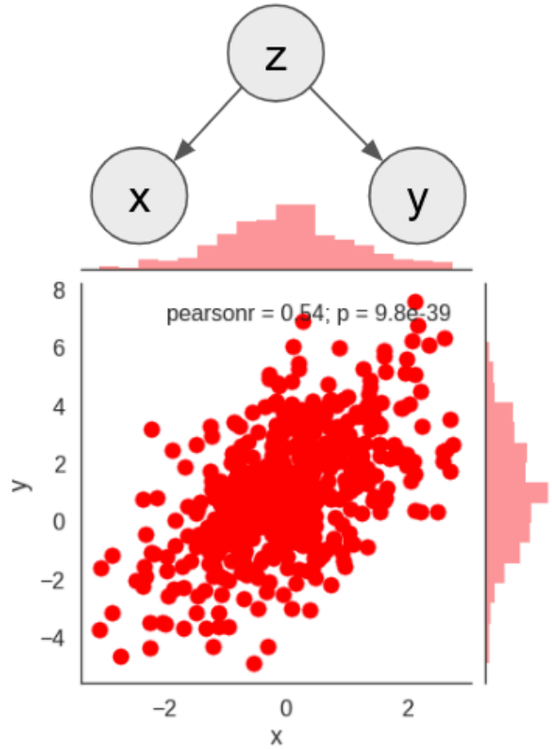
```
x = randn()
y = x + 1 + sqrt(3)*randn()
```

```
y = 1 + 2*randn()
x = (y-1)/4 + sqrt(3)*randn()/2
```

```
z = randn()
y = z + 1 + sqrt(3)*randn()
x = z
```

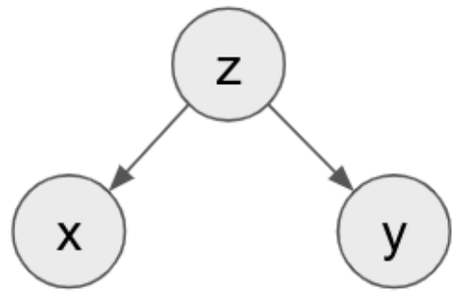






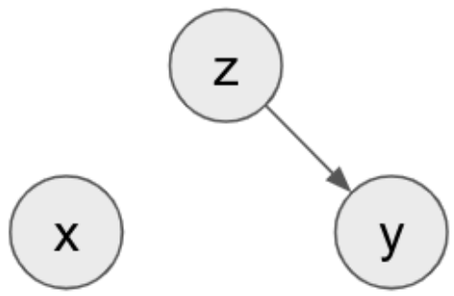
What is an intervention?

Graphically, to simulate the effect of an intervention, you **mutilate** the graph by removing all edges that point into the variable on which the intervention is applied, in this case x.



$$P(y|do(X)) = p(y|x)$$

$$P(y|do(X)) = p(y)$$



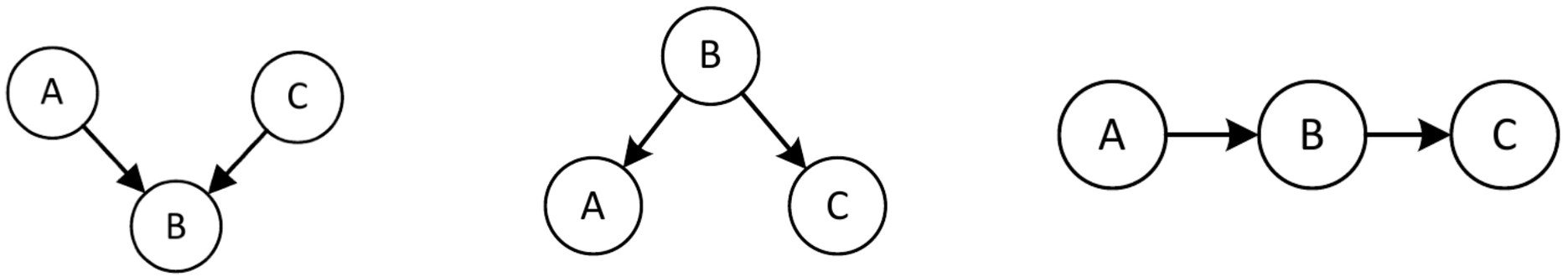
$$P(y|do(X)) = p(y)$$

By only looking at the causal diagram, we are now able to predict how the scripts are going to behave under the intervention $X = 3$.

Causal Graphs

The primary language for modeling causal mechanisms and expressing our assumptions is the language of *causal graphs*.

Causal graphs encode our domain knowledge about the causal mechanisms underlying a system or phenomenon under study.

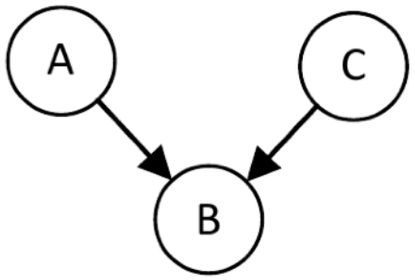


Causal graphs are assumed to be acyclic. This is why they are called DAGs (Directed Acyclic Graphs).

Causal Graphs

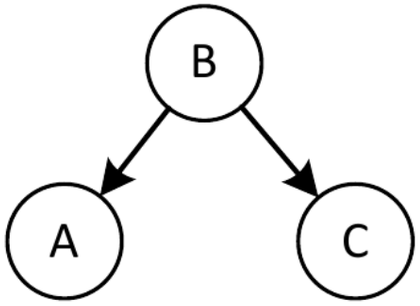
Fundamentally, a causal graph describes a non-parametric data-generating process over its nodes.

By specifying independence and dependence between the nodes, the graph constrains relationship between generated variables corresponding to those nodes.



B is a **collider** for A and C

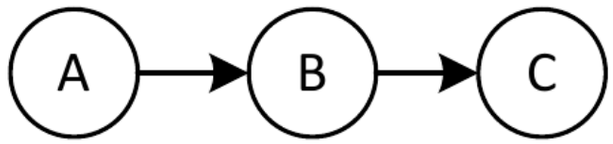
A and C are statistically independent



B is a **confounder**
B creates a **fork** to A and C

A and C are not independent.

A and C are independent conditional on B

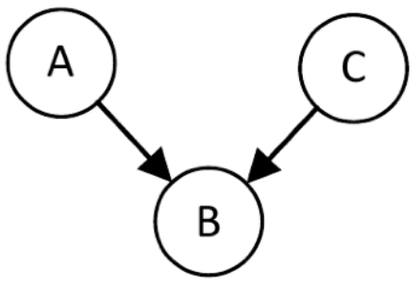


B forms a **chain** from A to C

A and C are conditionally independent given B

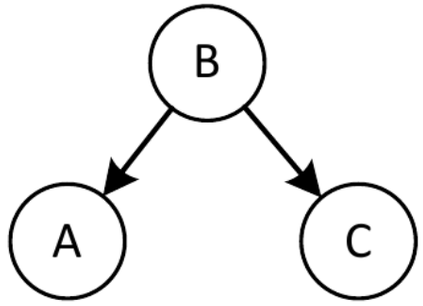
Causal Graphs

Fundamentally, a causal graph describes a non-parametric data-generating process over its nodes.



B is a **collider** for A and C

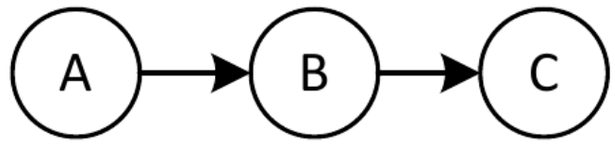
A and C are statistically independent



B is a **confounder**
B creates a **fork** to A and C

A and C are not independent.

A and C are independent conditional on B



B forms a **chain** from A to C

A and C are conditionally independent given B

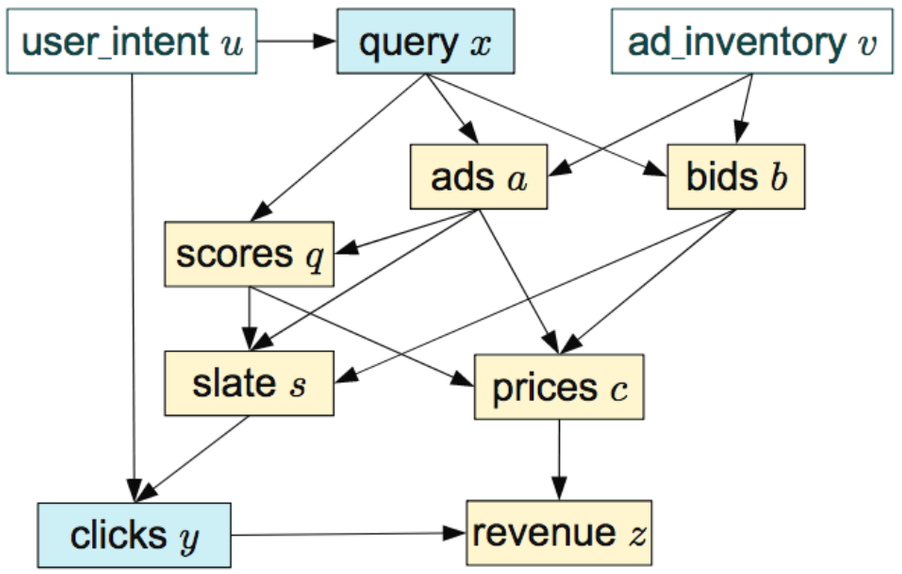
The three basic structures can be extended to determine statistical independence in any graph.

Causal Graphs

A DAG provides enough extra-data information (in terms of conditional independences) **to answer many causal queries**, even with the data generating process hidden.

More in a moment…

Structural Causal Models



DAG

$$
\begin{aligned}
x &= f_1(u, \varepsilon_1) && \text{Query context } x \text{ from user intent } u. \\
a &= f_2(x, v, \varepsilon_2) && \text{Eligible ads } (a_i) \text{ from query } x \text{ and inventory } v. \\
b &= f_3(x, v, \varepsilon_3) && \text{Corresponding bids } (b_i). \\
q &= f_4(x, a, \varepsilon_4) && \text{Scores } (q_{i,p}, R_p) \text{ from query } x \text{ and ads } a. \\
s &= f_5(a, q, b, \varepsilon_5) && \text{Ad slate } s \text{ from eligible ads } a, \text{ scores } q \text{ and bids } b. \\
c &= f_6(a, q, b, \varepsilon_6) && \text{Corresponding click prices } c. \\
y &= f_7(s, u, \varepsilon_7) && \text{User clicks } y \text{ from ad slate } s \text{ and user intent } u. \\
z &= f_8(y, c, \varepsilon_8) && \text{Revenue } z \text{ from clicks } y \text{ and prices } c.
\end{aligned}
$$

Structural Causal Model

Identification

Once we have captured our causal assumptions in the form of a model, the second stage of causal analysis is *identification*.

In this stage, our goal is to analyze our causal model—including the causal relationships between features and which features are observed—to determine whether we have enough information to answer a specific causal inference question.

The task of causal identification is to determine an expression, the **causal estimand**, that expresses our target value as a function of the **observable correlational relationships in our system**.

Identification

Using **do-calculus** (developed by J.Pearl), we can derive simple methods (**algorithms**) for causal identification in many situations. In some cases, the causal query can be non identifiable.

pedemonte96 / causaleffect Public

Watch 3 | Star 20 | Fork 0

Code | Issues 3 | Pull requests | Actions | Projects | Wiki | Security | Insights

main | 2 branches | 2 tags | Go to file | Add file | Code

pedemonte96 Create CONTRIBUTING.md — 320d16f on 12 Jul — 19 commits

.github/ISSUE_TEMPLATE	Update issue templates	3 months ago
causaleffect	improved verbose d-separation	4 months ago
documentation	improved documentation	4 months ago
examples	fixed example and added documentation	4 months ago
images	updated readme	4 months ago
tests	add id tests	3 months ago
.gitignore	causal effect added	4 months ago
CODE_OF_CONDUCT.md	Create CODE_OF_CONDUCT.md	3 months ago
CONTRIBUTING.md	Create CONTRIBUTING.md	3 months ago
LICENSE	Create LICENSE	3 months ago
README.md	removed pycairo dependency	4 months ago
pyproject.toml	build done	4 months ago
requirements.txt	removed pycairo dependency	4 months ago
setup.py	Update setup.py	3 months ago

About

Python package to compute conditional and non-conditional causal effects.

Readme

MIT License

Releases 2

v0.0.2 Latest
on 19 Jun

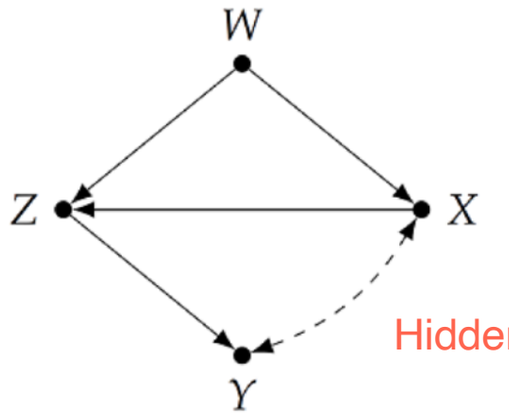
+ 1 release

Packages

No packages published
Publish your first package

Languages

Python 100.0%



Hidden Confounder

```
import causaleffect

G = causaleffect.createGraph(['X<->Y', 'Z->Y', 'X->Z', 'W->X', 'W->Z'])
causaleffect.plotGraph(G)
```

```
P = causaleffect.ID({'Y'}, {'X'}, G)
P.printLatex()
```

The code above computes the causal effect, and returns a string encoding the distribution in LaTeX notation:

```
'\sum_{w, z}P(w)P(z|w, x)\left(\sum_{x}P(x|w)P(y|w, x, z)\right)'
```

This string, in LaTeX, is

$$\sum_{w,z} P(w)P(z|w,x)\left(\sum_{x} P(x|w)P(y|w,x,z)\right)$$

Estimation

Once we have found these causal quantities, we need to choose how to **estimate them** by using statistical/ML methods.

- ML provides a **systematic framework** for learning the form of the conditional expectation function from the data!
- It allows us to do causal inference with **minimal assumptions** about the functional form of our model.
- It allows us to do causal inference with **high-dimensional data**.

Example

Example 3.1. ***AdBot*** *Consider an online advertising agent attempting to maximizing clickthroughs, with* $X \in \{0, 1\}$ *representing two ads,* $Y \in \{0, 1\}$ *whether or not it was clicked upon, and* $Z \in \{0, 1\}$ *the sex of the viewer. A marketing team collects the following data on purchases following ads shown to focus groups to be used by AdBot:*

	Ad 0	**Ad 1**
Male	108/120 (90%)	340/400 (85%)
Female	266/380 (70%)	65/100 (65%)
Total	374/500 (75%)	405/500 (81%)

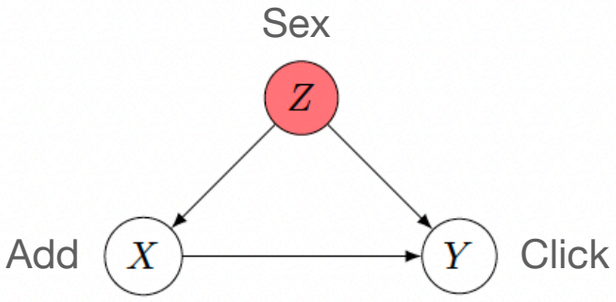
Table 1. Clickthroughs in the AdBot setting striated by the ad shown to participants in a focus group, and the sex of the viewer.

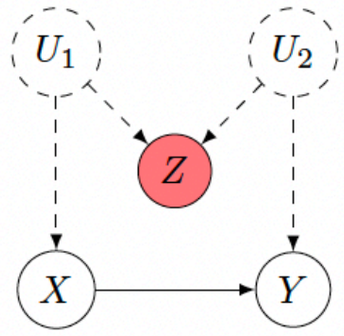
If the sex of a viewer is not know, which ad is the best choice?

From “Causal Inference in AI Education: A Primer”

Example

These are two different interventional stories:

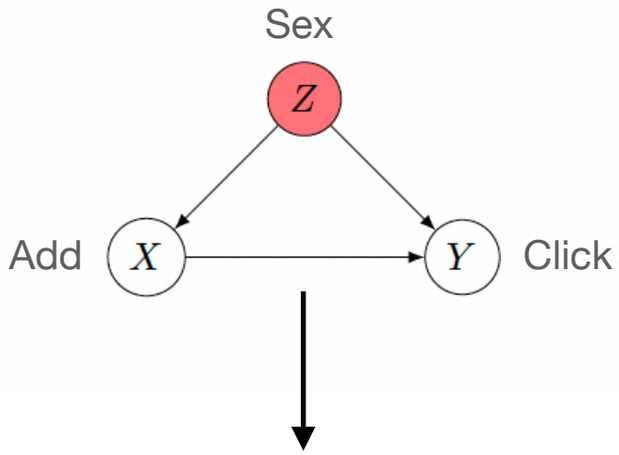




From “Causal Inference in AI Education: A Primer”

Example

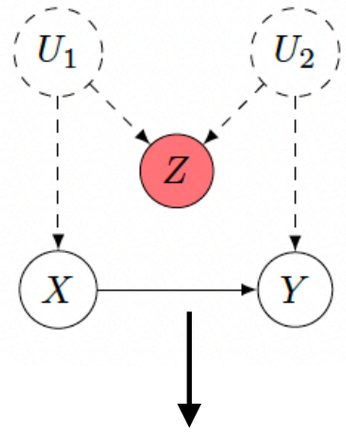
These are two different interventional stories:




```
1 G = causaleffect.createGraph(['X->Y', 'Z->Y', 'Z->X'])
2 P = causaleffect.ID({'Y'}, {'X'}, G)
```

$$p(Y \mid \mathrm{do}(X)) = \sum_{Z} P(Y \mid X, Z) P(Z)$$

If (a) is our explanation of the data, then AdBot should display Ad0.




```
1 G = causaleffect.createGraph(['Z<->Y', 'Z<->X', 'X->Y'])
2 P = causaleffect.ID({'Y'}, {'X'}, G)
```

$$p(Y \mid \mathrm{do}(X)) = P(Y \mid X)$$

If (b) is our explanation of the data, then AdBot should display Ad1.

From “Causal Inference in AI Education: A Primer”

Python package failing | pypi v0.6 | python 3.6 | 3.7 | 3.8

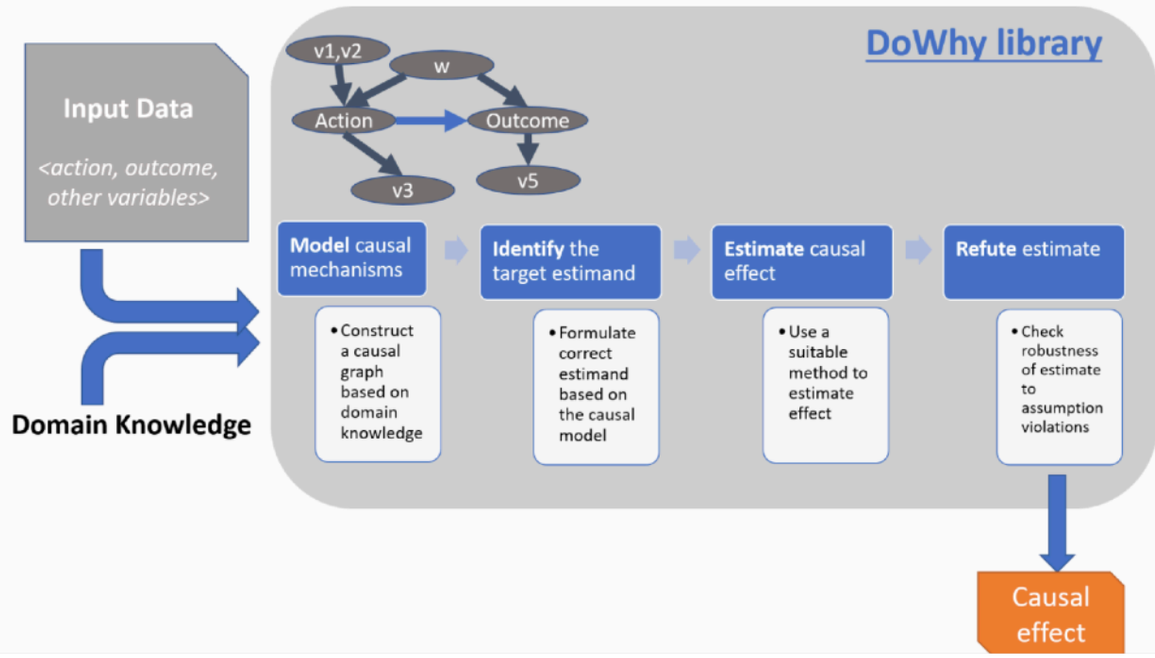
DoWhy | An end-to-end library for causal inference

Amit Sharma, Emre Kiciman

Introducing DoWhy and the 4 steps of causal inference | Microsoft Research Blog | Video Tutorial | Arxiv Paper | Slides

Read the docs | Try it online! launch binder

Case Studies using DoWhy: Hotel booking cancellations | Effect of customer loyalty programs | Optimizing article headlines | Effect of home visits on infant health (IHDP)



README.md

Azure Pipelines never built | pypi v0.12.0 | wheel yes | python 3.6 | 3.7 | 3.8

EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation

EconML is a Python package for estimating heterogeneous treatment effects from observational data via machine learning. This package was designed and built as part of the ALICE project at Microsoft Research with the goal to combine state-of-the-art machine learning techniques with econometrics to bring automation to complex causal inference problems. The promise of EconML:

- Implement recent techniques in the literature at the intersection of econometrics and machine learning
- Maintain flexibility in modeling the effect heterogeneity (via techniques such as random forests, boosting, lasso and neural nets), while preserving the causal interpretation of the learned model and often offering valid confidence intervals
- Use a unified API
- Build on standard Python packages for Machine Learning and Data Analysis

+ 25 releases

Packages

No packages published

Contributors 17

+ 6 contributors

Languages

Jupyter Notebook 81.5%
Python 16.9%
Cython 1.5%
Other 0.1%

README.md

pypi package 0.11.1 Build Status docs passing downloads 512k
cii best practices passing

Disclaimer

This project is stable and being incubated for long-term support. It may contain new experimental code, for which APIs are subject to change.

Causal ML: A Python Package for Uplift Modeling and Causal Inference with ML

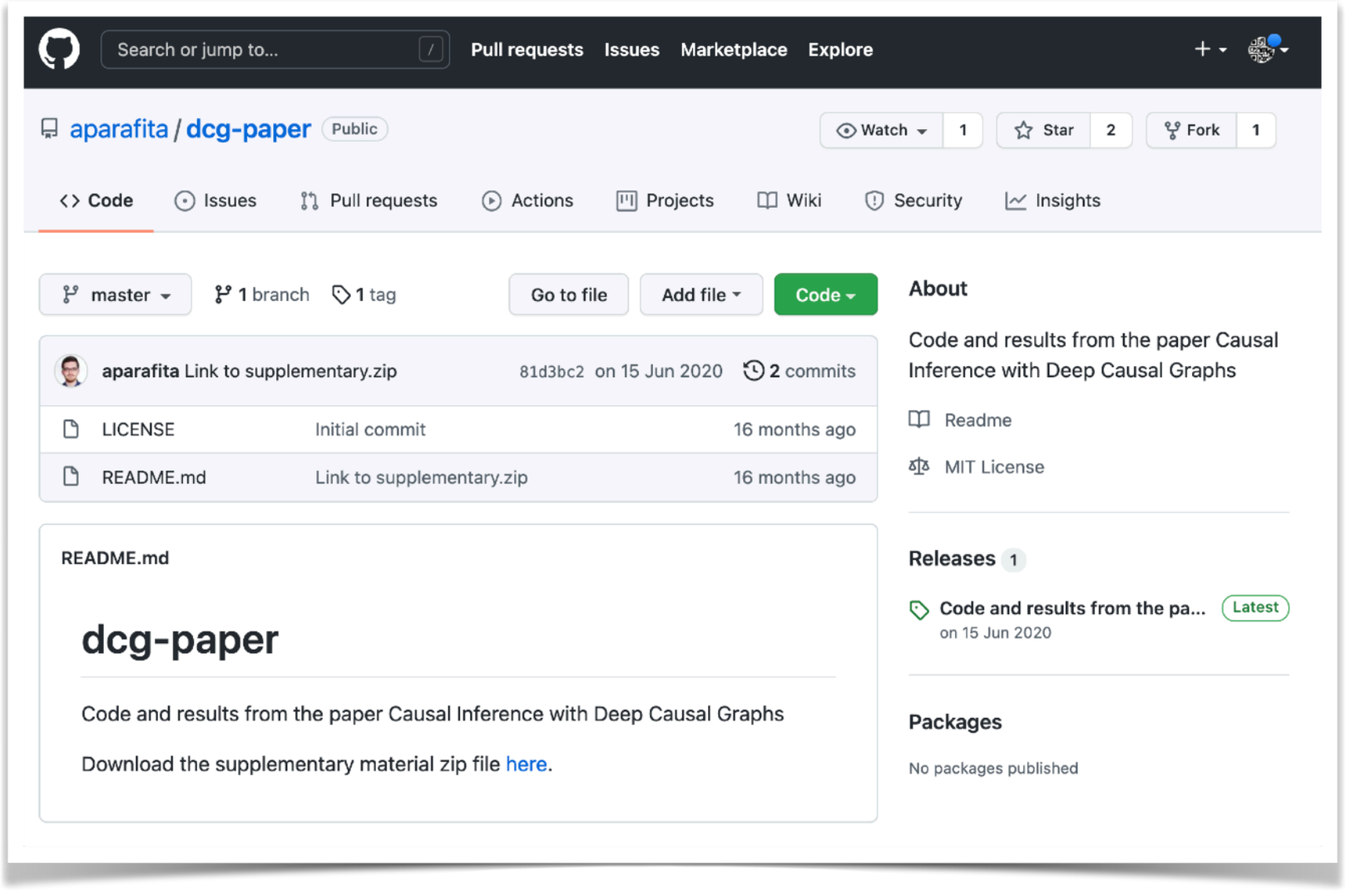
Causal ML is a Python package that provides a suite of uplift modeling and causal inference methods using machine learning algorithms based on recent research. It provides a standard interface that allows user to estimate the Conditional Average Treatment Effect (CATE) or Individual Treatment Effect (ITE) from experimental or observational data. Essentially, it estimates the causal impact of intervention `T` on outcome `Y` for users with observed features `X`, without strong assumptions on the model form. Typical use cases include

github-pages Active

Languages

Python 97.2% Cython 2.8%

Common causal inference techniques estimate a different model for each causal query.



Can we estimate a single model to estimate any identifiable query?