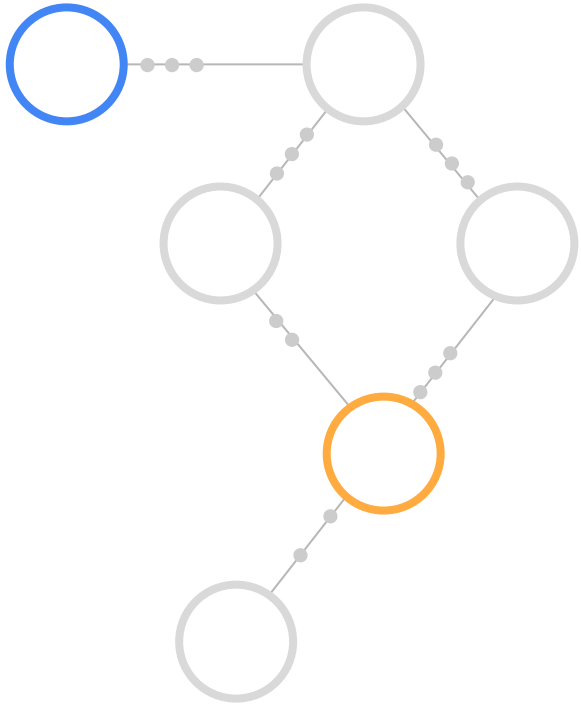
from correlation to causality in AI



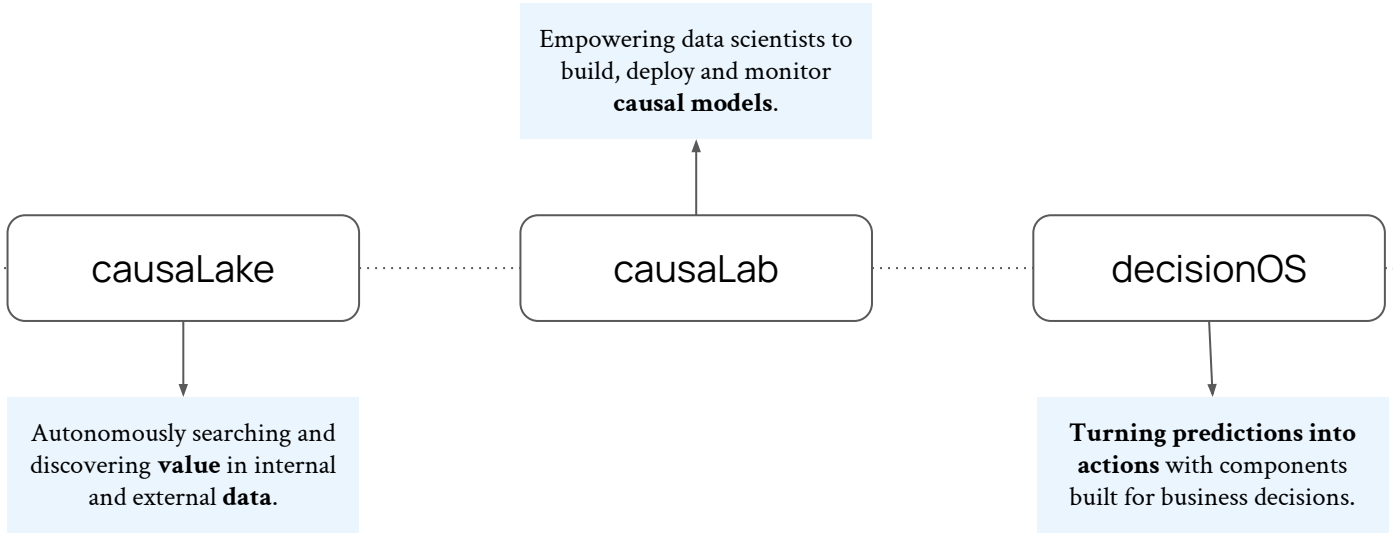
causaLens

causaLens

creating a world in which humans can trust machines with the greatest challenges in the economy, society and healthcare.

towards robust, fair and trustworthy AI

product


Empowering data scientists to build, deploy and monitor **causal models**.
causaLake
causaLab
decisionOS
Autonomously searching and discovering **value** in internal and external **data**.
Turning predictions into actions with components built for business decisions.

counterfactuals

"creating possible alternatives to life events that have already occurred; something that is contrary to what actually happened."

what if...?

counterfactuals

envisioning new worlds that enable us to test various scenarios

clinical trials in the desktop

causal understanding

Evaluating counterfactuals require specific understanding of **cause and effect relationships**.

Predicting is not sufficient!

what if I change X…?

how does X affect Y…?

causal understanding

We need to understand **how to act** in order to optimise for a desired outcome.

Predicting		Intervening
How do I predict churn?	?	How do I retain my customers?
How do I predict the next item in the shopping cart?	→	Which products should I recommend?
What is the expected return of an advertising campaign?		How do I optimise my marketing allocation?

causal understanding

We need to understand **how to act** in order to optimise for a desired outcome.

causal understanding

What **causes** customer loyalty?

What **drives** a customer to buy the next product?

What is the **impact** of adding into an advertising channel?

Intervening

How do I retain my customers?

Which products should I recommend?

How do I optimise my marketing allocation?

causaLens

foundational concepts

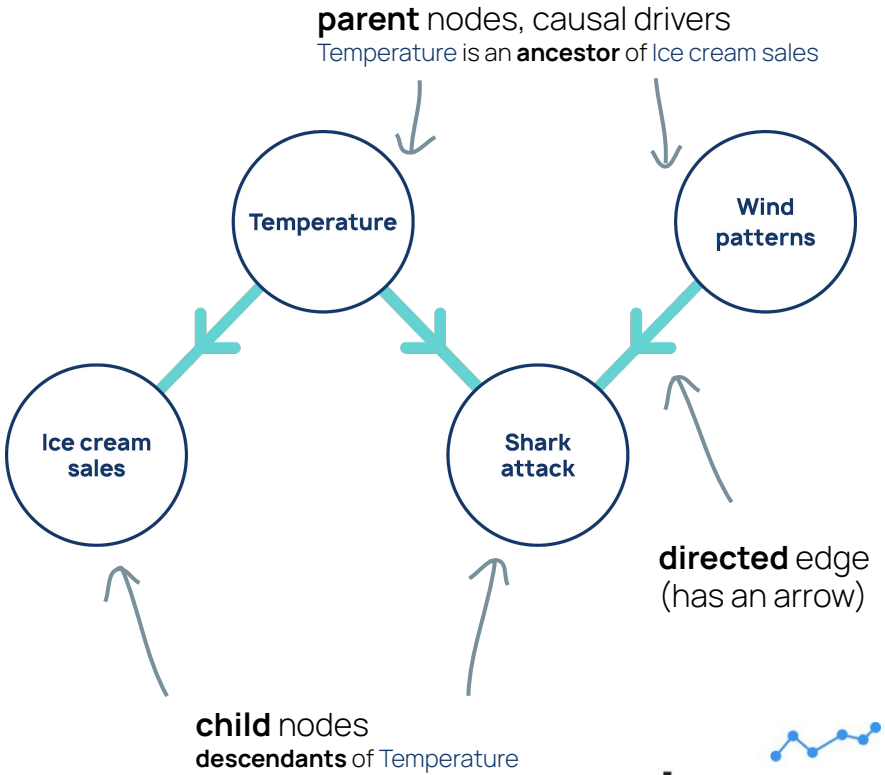
what do we mean by causal AI, how to infer causality from data and what are the downstream applications

Foundational concepts in Causal AI

Causal graphs describe causal relations

A key concept in Causal AI is the **causal graph**, which we use to describe causal relationships.

- We see a graph (right) for the ice cream sales and shark attacks example.
- **Nodes** (circles) = the variables.
- **Edges** (lines) = relationships - arrows show the direction from cause to effect.
- **Directed Acyclic Graph** or **DAG**, since all edges have arrows and no cycles exist - no directed path leads from a node to itself.



causaLens

Foundational concepts in Causal AI

Structural causal models describe the functional relationships

Structural causal models or **SCMs** describe how the relevant features of the world interact with each other.

An SCM $\{U, V, F\}$ is fully described by exogenous variables, U, endogenous variables, V, and a set of functions, F, that assign values to variables in V based on other variables in the model.

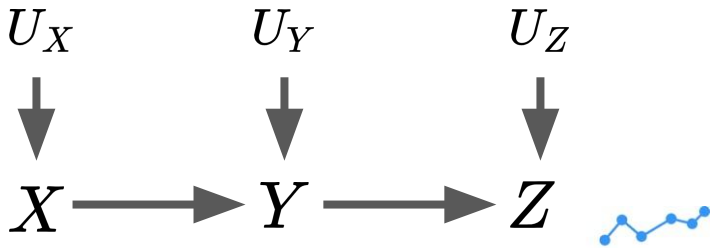
- A variable A is a **direct cause** of another B if it appears in the function which assigns B its value.
- Each SCM has an associated causal graph - but it contains **more information on causal relationships** than a causal graph alone.

A structural causal model

$$V = \{X, Y, Z\},$$
$$U = \{U_X, U_Y, U_Z\},$$
$$F = \{f_X, f_Y, f_Z\}$$

$$f_X : X = U_X$$
$$f_Y : Y = \frac{X}{3} + U_Y$$
$$f_Z : Z = \frac{Y}{16} + U_Z$$

Corresponding causal graph



Foundational concepts in Causal AI

Causal graphs imply statistical relations

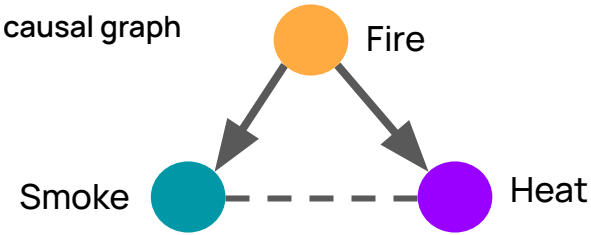
If we know the causal graph, then we can use it to **infer statistical properties of the relationships** among variables.

- Variables are **independent** if knowing one tells us nothing about another - e.g. shark attacks and Tesla's stock price - Pr(shark attacks | Tesla's price) = Pr(shark attacks).
- From a causal graph, we can infer **conditional independencies** → variables which are uncorrelated given the presence of another variable or variables.

Usually, we don't know the causal graph - we need to **infer causal knowledge from data** to draw the graph.

- As we'll see more later, **Conditional Independence Tests** can be used to find a *set* of graphs (**Markov Equivalence Class**) consistent with conditional independence relations in the data.

Assumed causal graph



Implied statistical relations

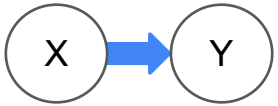
- **Smoke** and **Heat** are likely dependent (correlated)
- **Smoke** and **Heat** are **conditionally independent** (uncorrelated) given **Fire**
  - In the graph, **Fire** is a **confounder**, causing the correlation between **Smoke** and **Heat**

causaLens

Foundational concepts in Causal AI

But statistical relations do not imply specific causal graphs

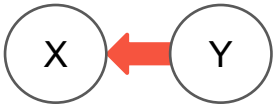
Correlation **cannot distinguish** between these 5 cases:



X causes Y

X is a **true causal driver of Y,** changes in X lead to changes in Y.

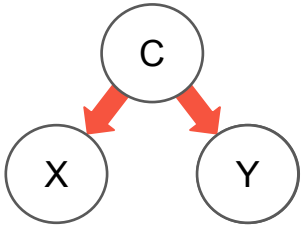
Use X to model Y



Y causes X

Now the other way around, Y is a driver of X

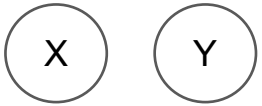
Do **not** use X to model Y



C causes both
C = 'confounder'

A 3rd variable is the driver of both of them

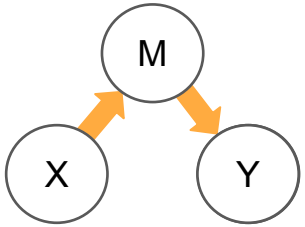
Do **not** use X to model Y



Association by chance ("spurious correlation")
No causal relationship

No causal relationship.

Do **not** use X to model Y



X has an indirect impact on Y
M = 'mediator'

X is the driver of a 3rd variable, which is the driver of Y

X *could* be used to model Y…

But M would create a better model for Y

causaLens

Foundational concepts in Causal AI

Interventions can isolate causal structures

As an example, assume we know that X and Y are correlated with one another, and that one causes the other (we don't know which).

$$p(X, Y) = p(Y \mid X)p(X)$$
$$p(X, Y) = p(X \mid Y)p(Y)$$

Both factorizations of the joint distribution are equally possible.

If we **intervene** on X, fixing it to x, only one of equations [1] and [2] is correct depending on the true direction of the causal relationship:

- In one case ([1]), **intervening changes** the conditional distribution of Y.
- In the other ([2]), the distribution of Y **remains unchanged**.

"do-calculus" showing two possible distributions after intervening on X (applying the "do" operator)

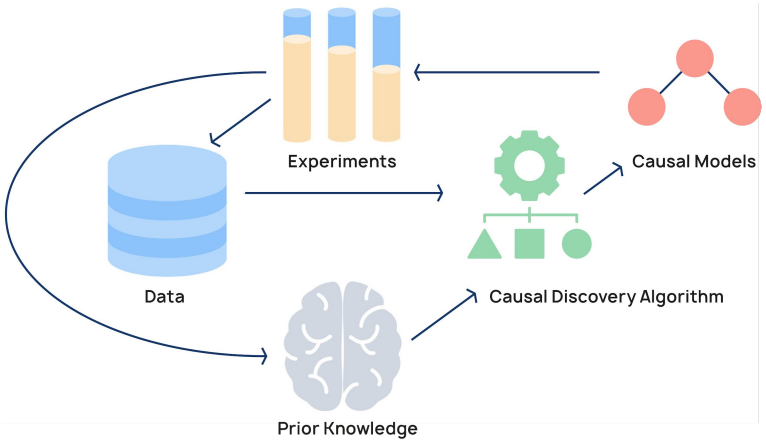
[1] $$p(Y \mid \mathrm{do}(X = x))p(\mathrm{do}(X = x)) = p(Y \mid \mathrm{do}(X = x))$$

[2] $$p(\mathrm{do}(X = x) \mid \mathrm{Y})p(Y) = p(Y)$$

causaLens

Foundational concepts in Causal AI

Acquire causal knowledge in three ways



Guiding principle of Causal AI is to **understand the causes and effects amongst variables of interest** in a system.

Three ways to acquire such causal knowledge, i.e. estimate graph structure and functional relationships:

1. **Interventional Experimentation**
2. **Causal Discovery** from observations
3. **Domain Expertise**, supplying prior knowledge - even partial knowledge

We'll discuss 2 and 3 more in this talk.

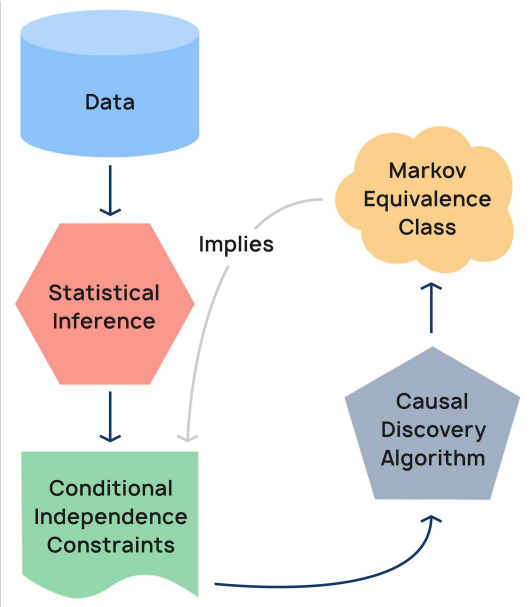
Foundational concepts in Causal AI

Causal discovery reveals causal relations in data

Causal Discovery extracts causal knowledge automatically from observations.

Two broad classes of methods:

1. **Constraint-based** - sequence of statistical tests to determine dependencies between variables, then orientate graph - e.g. PC, FCI
2. **Score-based** - directly searching space of graphs by evaluating how well each graph explains observed data - e.g. NOTEARS



Constraint-based causal discovery

Foundational concepts in Causal AI

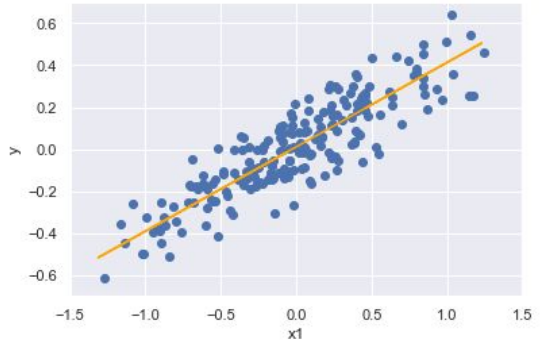
Conditional independence tests

Conditional independence tests identify whether an association between two variables, **Y** and $\mathbf{X_1}$, say, is statistically significant given the presence of another variable or variables.

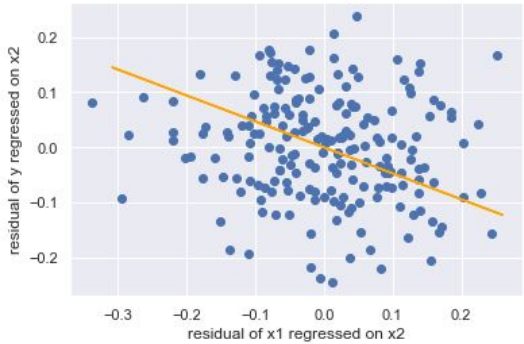
If we use **Partial Pearson** correlation → assumption: *linear* relations:

1. Identify conditioning variable(s) - in Causal AI, we condition on potential *common causes* (parents) of the two variables **Y** and $\mathbf{X_1}$.
2. Regress **Y** on the conditioning set $\{\mathbf{X_2},...\}$, and obtain the residuals - unexplained part of **Y** after controlling for $\{\mathbf{X_2},...\}$. Do the same separately for $\mathbf{X_1}$ to get its residuals.
3. Calculate Pearson correlation between residuals of **Y** and residuals of $\mathbf{X_1}$, and test it using a significance test (using a *p*-value).

Positive correlation between Y and X_1
Pearson correlation



Correlation becomes negative and weaker given X_2
Partial Pearson correlation, controlling for X_2

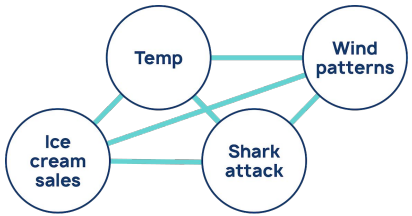


causaLens

Foundational concepts in Causal AI

PC algorithm is a classic causal discovery algorithm

1



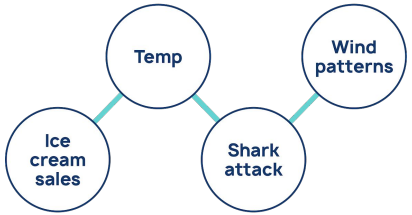
Start with a fully connected undirected graph:

- Assuming all variables are related somehow, and we don't know in which direction.

unknown causal relationship

known causal relationship

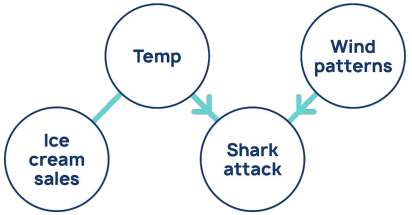
2



Conduct a series of conditional independence tests:

- Look for (unconditional) independencies, and remove edges accordingly.
- Then look for independencies conditional on one other variable, and remove arrows.
- Repeat this procedure, incrementing the number of variables you're conditioning on.

3



Orient the edges of the graph:

- Orient any "colliders" (causal structures with the form C1 → E ← C2) — these have a distinctive signature in data, and so they can be leveraged to orient the edges.
- Propagate edge orientations by following the logic that the remaining edges are not colliders.

causaLens

Foundational concepts in Causal AI

Counterfactuals, or “what-if” scenarios

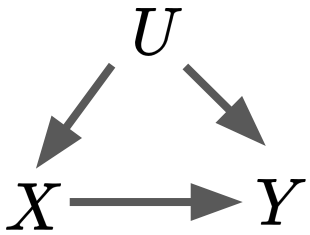
Counterfactuals take the form of:

“What would have happened if … ?“

- For example:
  What would have happen to my portfolio if oil prices had increased by 20%?

  Would I receive a loan today if I were 10 years older?
- Using the graph on the right, we can compute the following, “Y would be y had X been x, in situation $U = u$”, denoted by $Y_x(u) = y$
- Counterfactuals go beyond the *do*-operator. Here, we do not calculate the expected value of Y under one intervention or another, but the actual value of Y under the hypothesized new condition

$$X = aU$$
$$Y = bX + U$$



Foundational concepts in Causal AI

Counterfactual explanations

Counterfactual explanations (CFEs) analyze the decision-making process of an algorithm by analyzing how the algorithm's decisions change as a function of the inputs.

- Already trained function $m : \mathbf{X} \rightarrow Y$ maps from a set of features $\mathbf{x} \in \mathbf{X}$ (for instance, an individual's credit history and demographic information) to a target $y \in Y$, which could be a decision of whether to offer a loan or not.
- Counterfactual explanations answer:

  "What is the cheapest or easiest change to the individual's (changeable) characteristics, such that the decision would be inverted - from "reject" to "accept" or vice versa?"

Formal representation of counterfactual explanations - it is a cost minimization problem

Cost function *C* describes how difficult it is to change the features and can take on any form in general - its form will be domain-specific

$$\boldsymbol{\delta}^{\text{CFE}} := \underset{\boldsymbol{\delta}}{\operatorname{argmin}}\ C(\mathbf{x}^F + \boldsymbol{\delta}, \mathbf{x}^F)$$

$$\text{s.t.}\ \ m(\mathbf{x}^F + \boldsymbol{\delta}) = \mathbf{y}^*$$

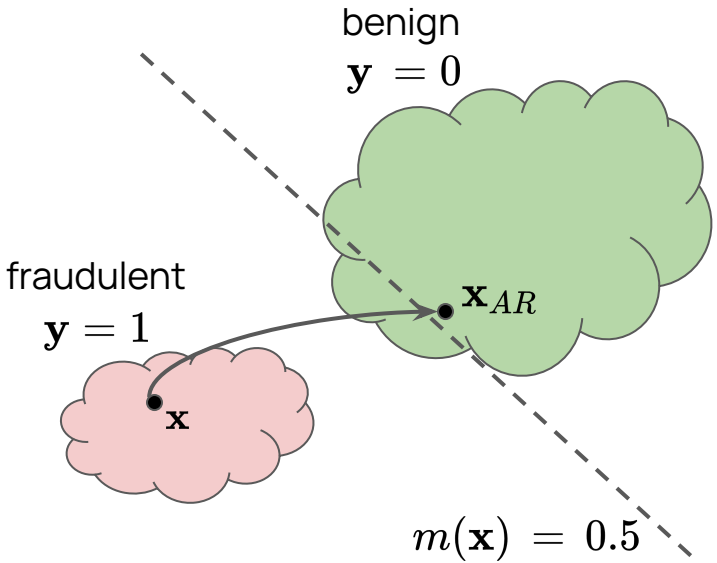
Foundational concepts in Causal AI

Algorithmic recourse

Algorithmic recourse goes beyond **CFEs** by:

1. ***Specifying how*** the desired system state may be achieved - the necessary actions or interventions
2. ***Explicitly modelling*** the downstream effects of actions on the features via a structural causal model (SCM)
   - For instance, if X_2 is related to X_1 via a deterministic function, then algorithmic recourse will adjust X_2 directly as it adjusts X_1

Already trained function $m : \mathbf{X} \rightarrow Y$ maps from a set of features $\mathbf{x} \in \mathbf{X}$ (for instance, a transaction on a bank account) to a target $y \in Y$, which could be whether the translation is fraudulent (=1) or benign (=0).



correlation-based world

In a correlation world, models are built with traditional ML techniques, optimizing for in-sample correlation and generally ignoring structural relationships within the input data.

causal AI world

In the causal AI world, a holistic understanding of the data is required: it's not enough for the input data to predict a target, but the entire relationships between all features need to be modelled. Feature importance is replaced by causal effect, and causal regularization and constraints are applied to the model.

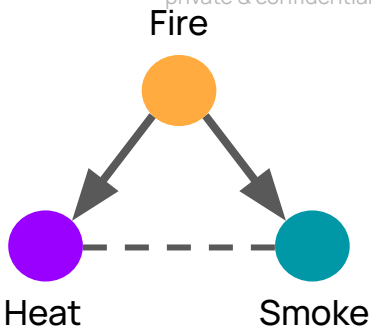
correlation-based world

"Heat can be a good predictor of smoke, so why is this a problem?"

A model that uses heat and fire as features wrongly attributes variance to heat, **overfitting** in the training set.

In correlation world, models produce **wrong counterfactuals**: they don't know what they need to do to reduce the heat in the room.

Causal discovery techniques "run clinical trials in historical data" in order to construct a causal graph, understanding cause and effect relationships.



causaLab implements **all techniques available in the literature**, as well as proprietary causaLens methods, within a simple modular framework.

```
config.causality_config.method =
CausalDiscoveryMethod.VAR_LINGAM
config.causality_config.method_params =
      CausalDiscoveryVarLingamParams(
          max_lag=5,
          threshold=0.05,
          fit_criterion=ModelFitCriteria.BIC
      )
```

causal AI world

causaLens

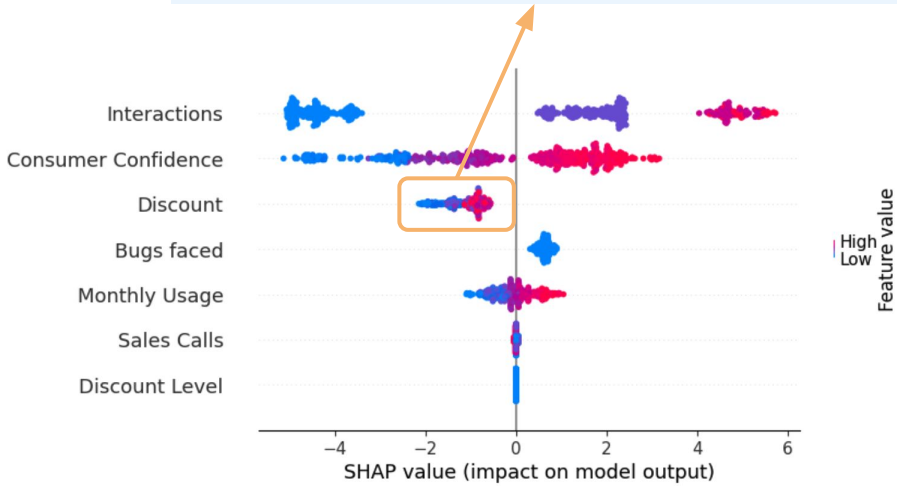
notebook #1: causal discovery from observational data

causaLens

Life in a world of correlations: Simpson's "paradox"

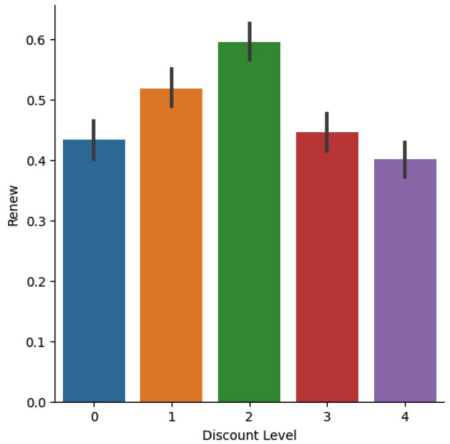
While optimizing for in-sample correlations, **ML models generalize poorly**. When faced with data outside of the training distribution, it may end up making "stupid decisions"

In this example, while predicting whether clients would renew, a gradient boosting model learned that **higher discounts correlates with churn**. Can you guess why that's the case?



causaLens

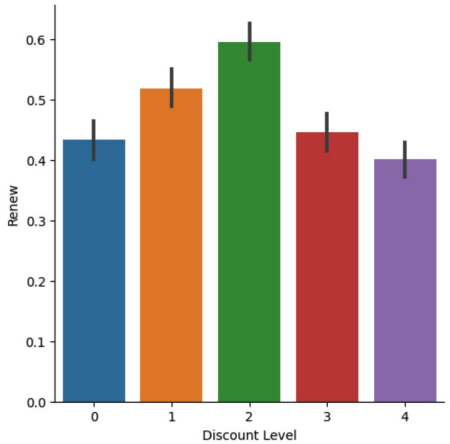
Life in a world of correlations: Simpson's "paradox"



correlation world: Customers who face bugs in products are less likely to renew. These are also the customers who are more likely to receive discounts, thus creating a **negative correlation between discount and renewals**.

Simpson's *paradox* is only a paradox in the correlation world. It exists due to the fact that confounders aren't properly taken into account. By simply measuring correlations, the model may end up suggesting - for a cohort of the population - that discounts is detrimental to customer retention.

Life in a world of correlations: Simpson's "paradox"



correlation world: Customers who face bugs in products are less likely to renew. These are also the customers who are more likely to receive discounts, thus creating a **negative correlation between discount and renewals**.

Age is a confounder between **vaccine** uptake and **hospitalization**, thus seemingly reducing the measured impact of vaccines in lowering hospitalizations.

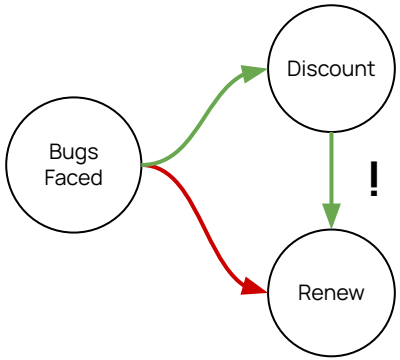
Value stocks underperform -> **growth** is a confounder, value portfolios **outperform** when controlling for growth, sector, and other factors.

There's a positive correlation between **exercise** and **cholesterol**. Again age is a confounder, and controlling for **age** we find that the causal effect is negative.

Employees who receive **raises** are more likely to churn -> **overwork** is a confounder, whose data is scarce. Models learn that paying less improves **retention**.

causaLens

Life in Causal AI world: causalNet

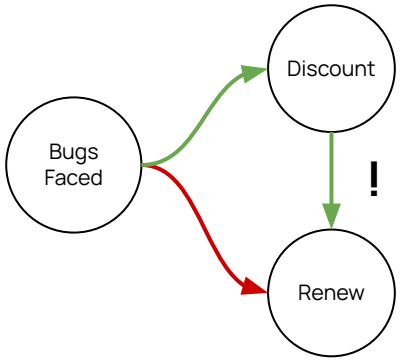


causal AI world: After controlling for bugs, we can infer that the **causal effect of applying discounts is positive**. This is only possible by understanding the causal structure of the data.

```
causalnet.add_edge('Discount', 'Renew', 'monotonic_increasing')
```

There are situations in which confounding variables will lack data, the data is fundamentally bias within the training set. **Causal Regularization ensures that any model follows this constraint, a priori**. Regardless of the input, the monotonicity condition is satisfied.

Life in Causal AI world: causalNet

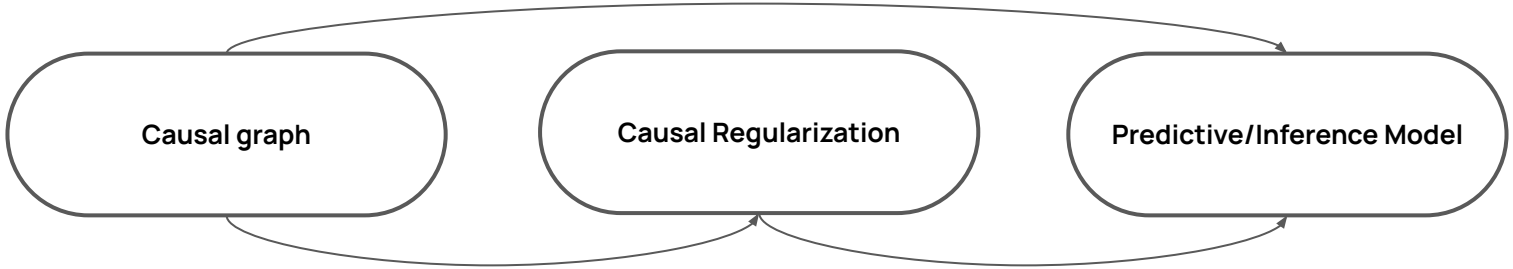


causal AI world: After controlling for bugs, we can infer that the **causal effect of applying discounts is positive**. This is only possible by understanding the causal structure of the data.

```
causalnet.add_edge('Discount', 'Renew', 'monotonic_increasing')
```

There are situations in which confounding variables will lack data, the data is fundamentally bias within the training set. **Causal Regularization ensures that any model follows this constraint, a priori**. Regardless of the input, the monotonicity condition is satisfied.

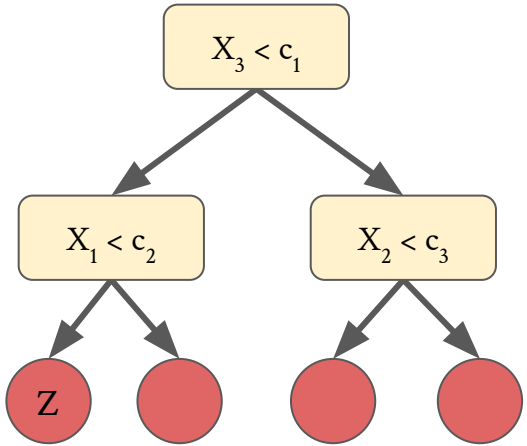
notebook #2: causal modelling and the dangers of shap




```
from causalnet_v2 import CausalNet

causalnet = CausalNet()  # empty causalnet
causalnet = CausalNet.from_graph(causal_graph)
causalnet.add_edge('treatment', 'mediator')
causalnet.add_edge('mediator', 'outcome', 'positive_linear')
causalnet.add_edge('treatment', 'outcome')
causalnet.fit( … )
```

notebook #3 causal Decision Tree

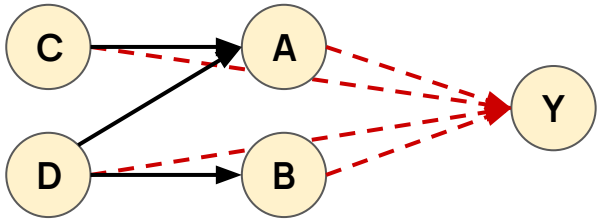


Classical decision trees do not enforce any topology in the ordering of variables, nor the directionality of the effect on the target.

1) **Causality + Kolmogorov complexity:**
when $X \rightarrow Y$, then $P(Y \mid X)\, P(X)$ is simpler to represent than $P(X \mid Y) P(Y)$.

2) **Causal Graphs / Bayesian Networks** factorize as, e.g.
$P(A, B, C) = P(A \mid B)\, P(B \mid C)\, P(C)$ for $C \rightarrow B \rightarrow A$

3) **Trees represent chains of conditional probabilities**:
$E(Z) = E(Z \mid X_1 < c_2, X_3 < c_1) P(X_1 < c_2 \mid X_3 < c_1)\, P(X_3 < c_1)$



Causal decision trees enforce the topology of the causal graph in the predictive model

causaLens

product research and future

Model risk assessment

Using causal technology to understand when models are expected to fail: from stability to fairness/bias

Causal Reinforcement Learning

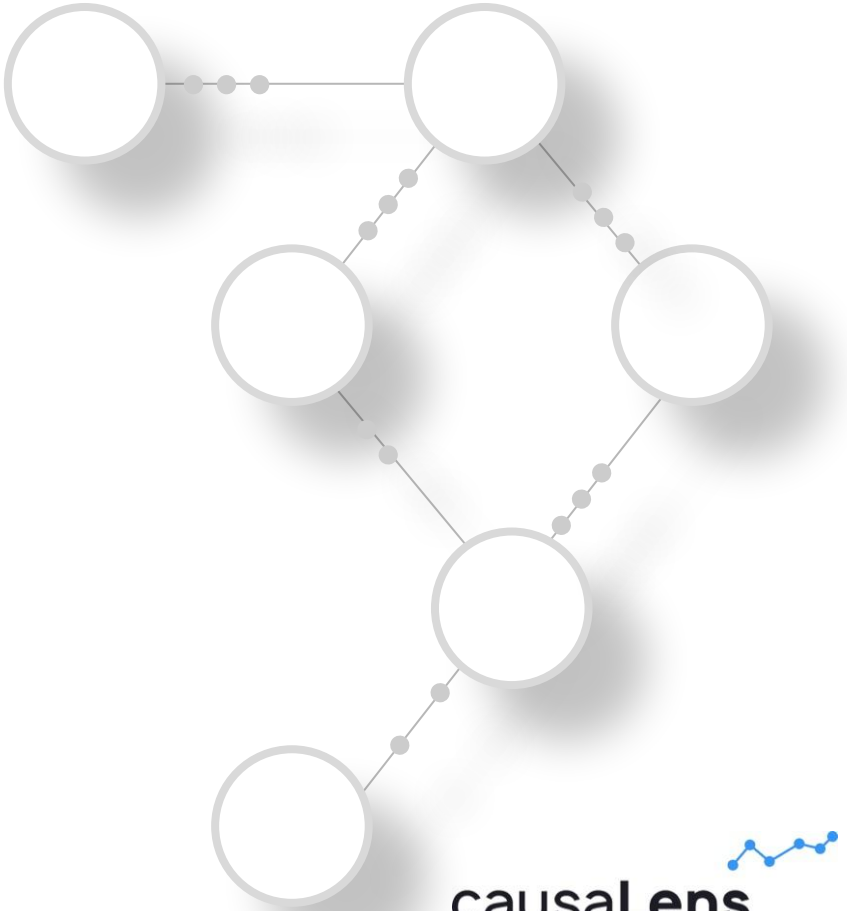
Creating RL agents that understand cause-and-effect, and can explain the decisions based on that understanding

Beyond differentiable learning

Developing learning techniques that go beyond backpropagation

Invariances and conservation laws

Exploiting the intersection between causality, invariances in the data and conservation laws for more robust models



causaLens