kaggle   Search kaggle   🔍   **Competitions**   **Datasets**   **Kernels**   **Discussion**   **Jobs**

Cause-effect pairs

Given samples from a pair of variables A, B, find whether A is a cause of B.

$10,000 · 266 teams · 4 years ago

Overview   Data   Discussion   Leaderboard   Rules   Late Submission

Overview

Description

Evaluation

Prizes

Help Credits

Timeline

**Come to our NIPS workshop (dec 9 or 10 in Tahoe).**

The problem of attributing causes to effects is pervasive in science, medicine, economy and almost every aspects of our everyday life involving human reasoning and decision making. What affects your health? the economy? climate changes? The gold standard to establish causal relationships is to perform randomized controlled experiments. However, experiments are costly while non-experimental "observational" data collected routinely around the world are readily available. Unraveling potential cause-effect relationships from such observational data could save a lot of time and effort.

Consider for instance a target variable B, like occurence of "lung cancer" in patients. The goal would be to find whether a factor A, like "smoking", might cause B. The objective of the challenge is to rank pairs of variables {A, B} to prioritize experimental verifications of the conjecture that A causes B.

As is known, "correlation does not mean causation". More generally, observing a statistical dependency between A and B does not imply that A causes B or that B causes A; A and B could be consequences of a common cause. But, is it possible to determine from the joint observation of samples of two variables A and B that A should be a cause of B? There are new algorithms that have appeared in the literature in the past few years that tackle this problem. This challenge is an opportunity to evaluate them and propose new techniques to improve on them.

We provide hundreds of pairs of real variables with known causal relationships from domains as diverse as chemistry, climatology, ecology, economy, engineering, epidemiology, genomics, medicine, physics. and sociology. Those are intermixed with controls (pairs of independent variables and pairs of variables that are dependent but not causally related) and semi-artificial cause-effect pairs (real variables mixed in various ways to produce a given outcome).

This challenge is limited to pairs of variables deprived of their context. Thus constraint-based methods relying on conditional independence tests and/or graphical models are not applicable. The goal is to push the state-of-the art in complementary methods, which can eventually disambiguate Markov equivalence classes. If you are skeptical that this is possible, try this quiz: Examine the plot below of values of variable B plotted as a function of values of variable A. Can you guess which one is a cause of the other? Hint: Some non-linear functions are non-invertible.
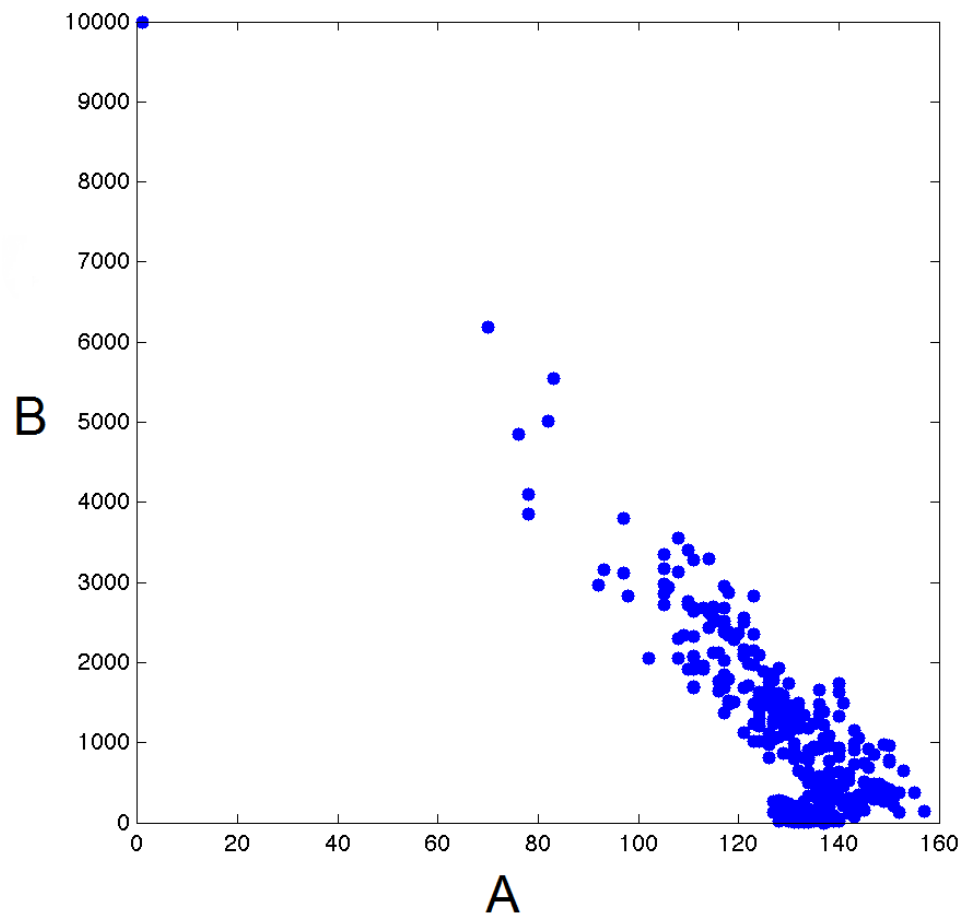
(example provided by D. Janzig).
Answer: A is temperature and B is altitude of German cities, hence B causes A

July 1: A new data release was made to address a normalization problem and the deadline was extended. Scores on the public leaderboard prior to July 1 were decreased by 0.5. Please make new submissions with the new validation set. The competition is open to new teams.

Leaderboard

1	ProtoML
2	jarfo
3	HiDLoN
4	FirfiD
5	mouse
6	Domcastro & Sayani
7	nor
8	LucaToni

77 discussion topics

Have the winners published their method publi...
1 reply · 10 months ago

Workshop on challenges at NIPS 2014
0 replies · 3 years ago

is this even possible?
10 replies · 4 years ago

track 1, looking for teamates
0 replies · 4 years ago

Code of top ranking participants
6 replies · 4 years ago

Launch
5 years ago

Close
4 years ago

266 **313**
Teams Competitors

Points **This competition awarded standard ranking points**
Tiers **This competition counted towards tiers**

Tags   custom metric   small

Our Team Terms Privacy Contact/Support