Robust Causal Graphs: Extensions to the NOTEARS Algorithm

Shaan Desai[1,2], Philip Pilgerstorfer[2], Roxana Pamfil[2], Konstantinos Georgatzis[2], Stephen J. Roberts[1]
[1]University of Oxford, Department of Engineering
[2]QuantumBlack (a Mckinsey Company)

Introduction

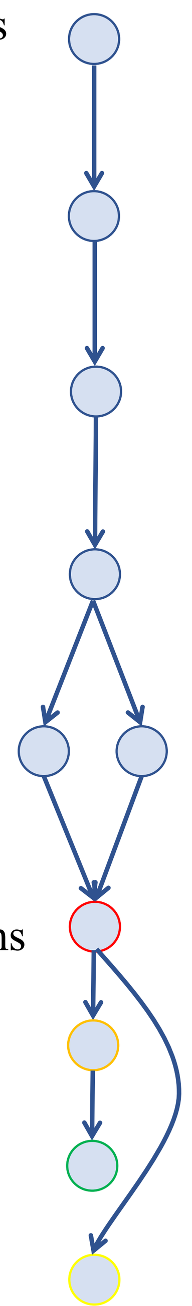
Causal Inference helps us draw conclusions about causal links in our data.

Directed Acyclic Graphs (DAGs) capture these links by representing a set of variables and their conditional dependencies.

Current state of the art in structure learning - **NOTEARS algorithm**.

Challenges:
- presence of discrete data
- sensitivity to dataset e.g. number of samples

Solutions:
- *Bootstrapping*
  - Visualizing variance of learned graphs
- *Statistical Tests*
  - Confidence measures based on bootstraps
  - Negative control tests
- *Extended Functionality*
  - Binary solver + combined solver



NOTEARS for Binary Data

Original NOTEARS (Gaussian likelihood) algorithm fails on discrete data.

Logistic NOTEARS (Bernoulli likelihood) solves this for binary data.

Can combine logistic with original to solve for mixed datatypes by splitting continuous variables and discrete (binarized).

Original NOTEARS

$$\min_{W\in\mathbb{R}^{dxd}} F(W) = \frac{1}{2n}||X - XW||_F^2 + \frac{\rho}{2}h(W)^2 + \alpha h(W) + \lambda||W||_1$$

New Logistic Loss

$$Loss_{binary} = \frac{1}{n}\left[-Tr(X^T * \log(\sigma(XW))) - Tr\left((1-X)^T * \log(1-\sigma(XW))\right)\right]$$

Logistic NOTEARS

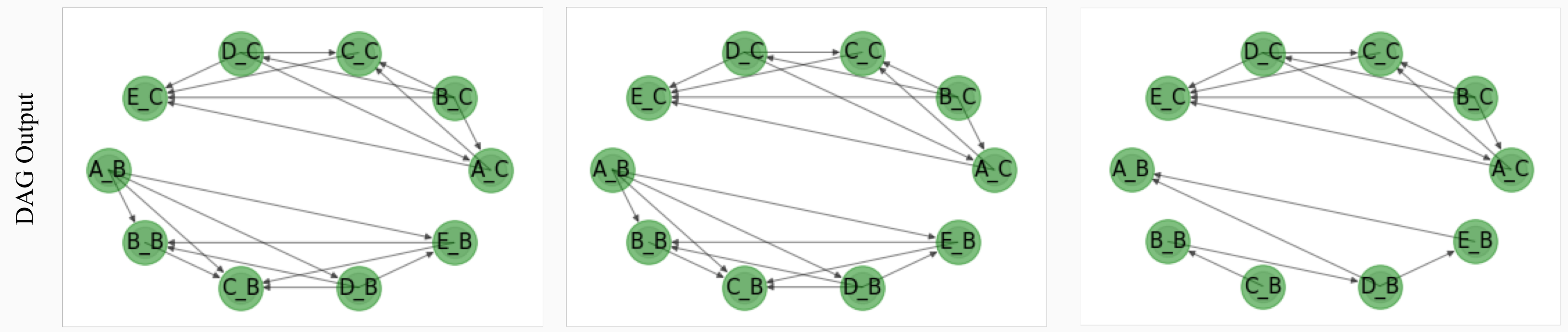
$$\min_{W\in\mathbb{R}^{dxd}} F(W) = Loss_{binary} + \frac{\rho}{2}h(W)^2 + \alpha h(W) + \lambda||W||_1$$

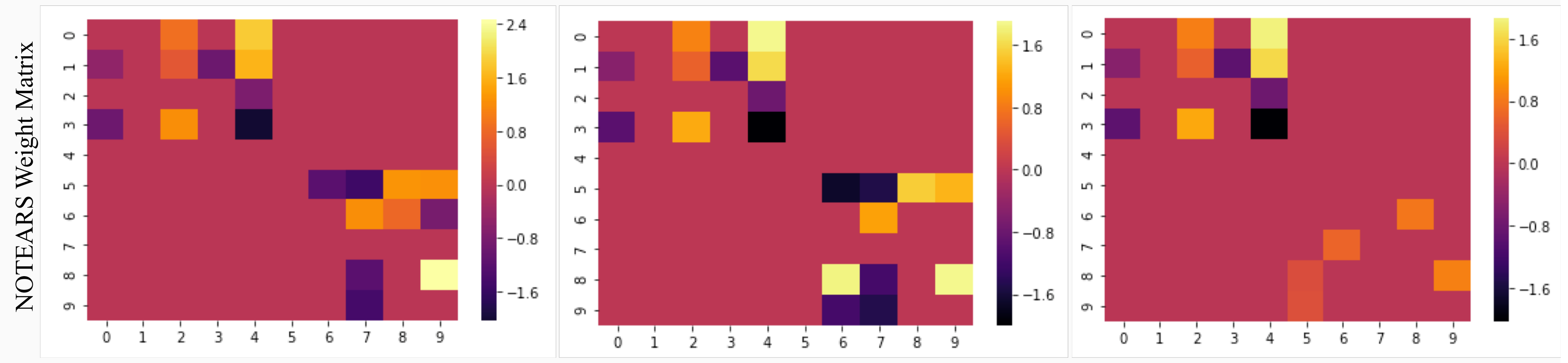
NOTEARS for Mixed Data

Combined NOTEARS

$X = [X_c, X_b]$ $\qquad$ $W = [W_c, W_b]$

$$\min_{W\in\mathbb{R}^{dxd}} F(W) = \frac{1}{2n}||X_c - XW_c||_F^2 + \frac{1}{n}\left[-Tr(X_b^T * \log(\sigma(XW_b))) - Tr\left((1-X_b)^T * \log(1-\sigma(XW_b))\right)\right] + ..$$



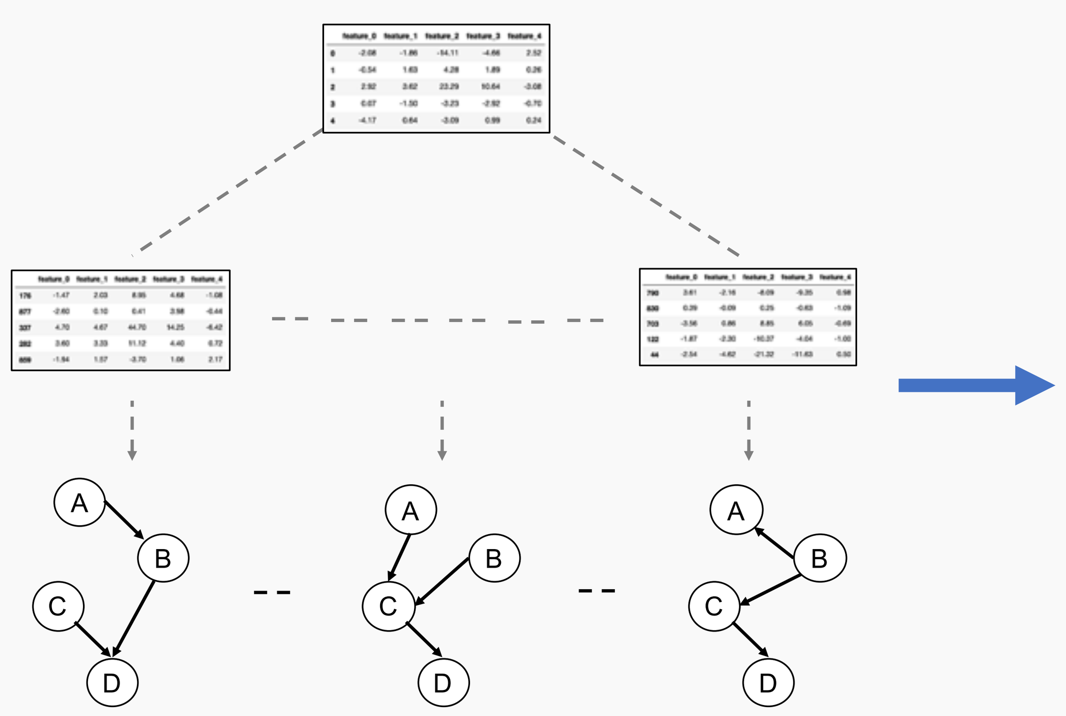


Ground Truth | Combined NOTEARS | Original NOTEARS

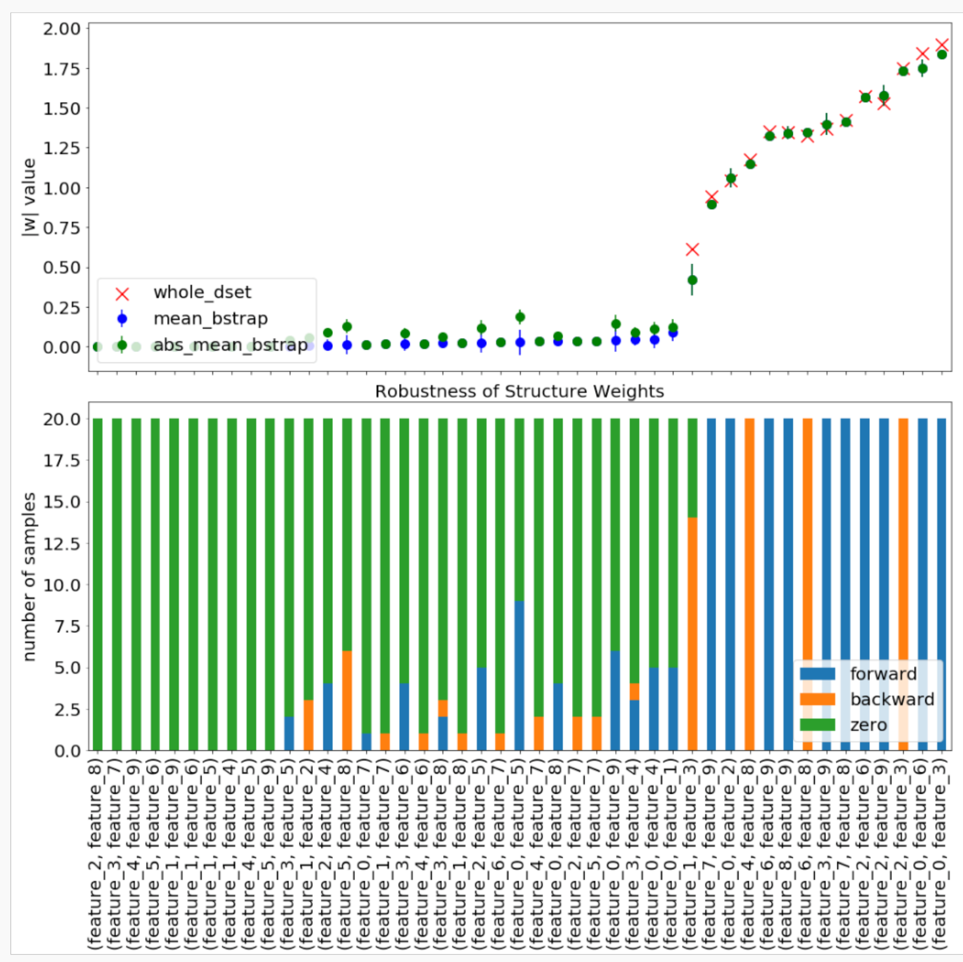
Bootstrap

Can use bootstrapping to generate multiple datasets and graphs
Each graph generated from NOTEARS has a weight matrix which encodes:
- **Direction of an edge**
- **Weights/coefficents of an edge**

Plot illustrates variance in the above properties
Key Point – bootstrapping gives us more information about our solver across samples



Bootstrapping



Variance Plots

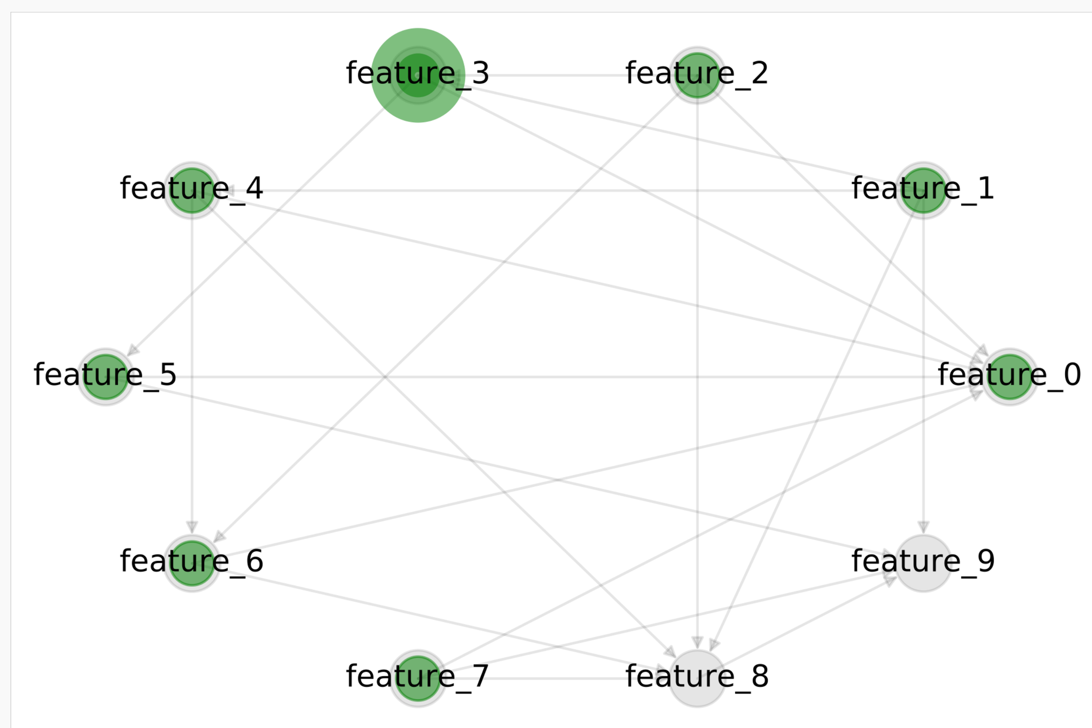
Statistical Tests

$$P(f) = \frac{1}{B}\sum_{i=1}^{B} f(G_i)$$

where:

$$f \in \{0,1\}$$

B is # of bootstraps.
Feature *f* can be any question which can be converted to binary output e.g.
- Does edge A-B exist?



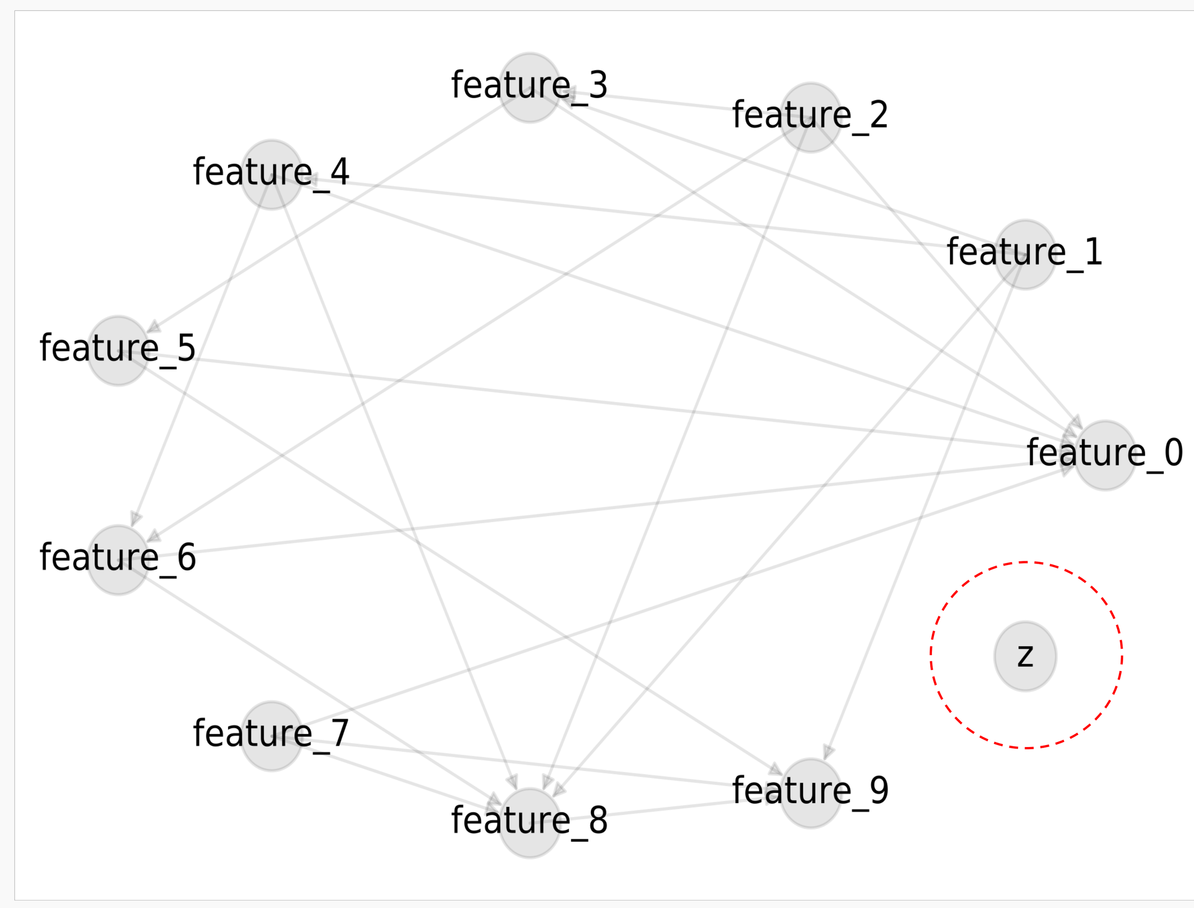
Markov Blanket of feature 3 ($p > 0.75$)

Negative Control

What happens when we have a variable which is independent of our dataset but models the covariance of other variables?

A single run of NOTEARS will not give us any sense for whether this exists

Bootstrapping thus helps us detect negative control features



Negative control feature Z with no links