A Brief Introduction to Causal Discovery and Causal inference

Part 2

LiNGAM (Linear Non-Gaussian Acyclic Model)

- Goal: Discover the causal graph from observational data only.
- Motivation:
  - In linear Gaussian models, causal direction is not identifiable (Markov equivalence).
  - LiNGAM resolves this by assuming non-Gaussian independent noise.
- Key Model:

$$X_i = \sum_{j \in Pa(i)} b_{ij} X_j + e_i$$

where
  - Graph is acyclic
  - Structural relations are linear
  - Noise terms e_i are independent and non-Gaussian

- Implication: Non-Gaussianity introduces asymmetry → causal direction becomes recoverable.

Shimizu, Shohei, et al. "A linear non-Gaussian acyclic model for causal discovery." *Journal of Machine Learning Research* 7.10 (2006).

How LiNGAM Works

- Rewrite the model in matrix form:

$$X = BX + e \quad \Rightarrow \quad X = Ae$$

where $A = (I - B)^{-1}$ is a mixing matrix.

Because e has independent non-Gaussian components,

→ Apply Independent Component Analysis (ICA) to recover A uniquely (up to permutation).

- Then:

$$B = I - A^{-1}$$

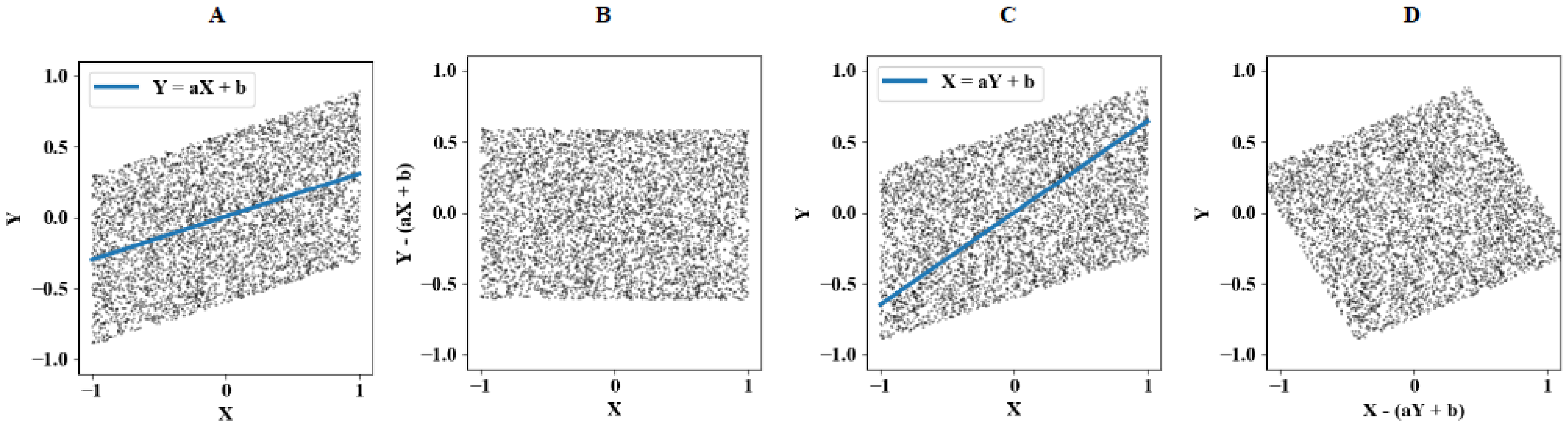
which directly gives the causal adjacency matrix.

Example

- Given the linear structural equations

$X = U_X$ and $Y = X + U_Y$ such that $U_Y \perp U_X$

- If U_X or U_Y is non-Gaussian
- Then the causal direction $X \rightarrow Y$ is identifiable



NOTEARS (Neural Optimization for DAG Structure Learning)

- NOTEARS is a causal discovery method that learns a DAG from observational data using continuous optimization.
- Key Idea:
  - Treat structure learning as optimizing a weight matrix W (like in linear regression),but enforce the DAG constraint via a smooth, differentiable function.
- Model

$$X = XW + \epsilon$$

  - $W_{ij} \neq 0$ means $j \to i$
  - ϵ are independent noise terms
- Goal: Find W that
  - Fits data well (low reconstruction error)
  - Forms a DAG (no directed cycles)

Zheng, Xun, et al. "Dags with no tears: Continuous optimization for structure learning." *Advances in neural information processing systems* 31 (2018).

Continuous DAG Constraint

- Traditional DAG constraint is combinatorial → NP-hard.
- NOTEARS introduces a smooth acyclicity constraint:

$$h(W) = Tr(e^{W \circ W}) - d = 0$$

where

  - $W \circ W$ is elementwise square
  - d is number of variables
  - W encodes a DAG if and only if $h(W) = 0$

- Optimization problem

$$\min_{\mathrm{W}} \frac{1}{2n} \left|\left|X - XW\right|\right|_F^2 + \lambda \left|\left|W\right|\right|_1 \qquad s.t.\ h(W) = 0$$

  - Use augmented Lagrangian to solve efficiently

Why NOTEARS Matters

Property	Traditional Causal Discovery	NOTEARS
Search space	Discrete (exponential)	Continuous (matrix space)
Optimization	Combinatorial, often greedy	Gradient-based, global optimization
DAG constraint	Hard to enforce	Smooth, differentiable ($h(W) = 0$)
Scalability	Limited	Scales to high-dimensional problems

- Extensions
  - Nonlinear NOTEARS: Replace linear model with neural nets (MLP)
  - NOTEARS-ICA: Handles non-Gaussian noise
  - NOTEARS-ADMM: Distributed optimization
- Takeaway
  - NOTEARS reframed causal structure learning as a continuous optimization problem, allowing DAGs to be learned with gradient-based methods.

NCM (Neural Causal Models)

- Neural Causal Models (NCMs) combine:
  - Structural Causal Models (SCMs): explicit causal graph + mechanisms
  - Neural Networks: flexible function approximators for structural equations
- Model Structure
  - For variables $X_1, X_2, \dots, X_i, \dots$

$$X_i = f_i(Pa(X_i), \epsilon_i)$$

  where
  - $Pa(X_i)$: parents of X_i in causal graph
  - f_i: implemented using neural networks
  - ϵ_i: independent noises
- Key Features
  - Captures nonlinear causal mechanisms
  - Supports counterfactual reasoning and interventions
  - Scales to high-dimensional and continuous data

Xia, K., Lee, K. Z., Bengio, Y., & Bareinboim, E. (2021). The causal-neural connection: Expressiveness, learnability, and inference. *Advances in Neural Information Processing Systems*, *34*, 10823-10836.

Learning and Using NCMs

- Learning

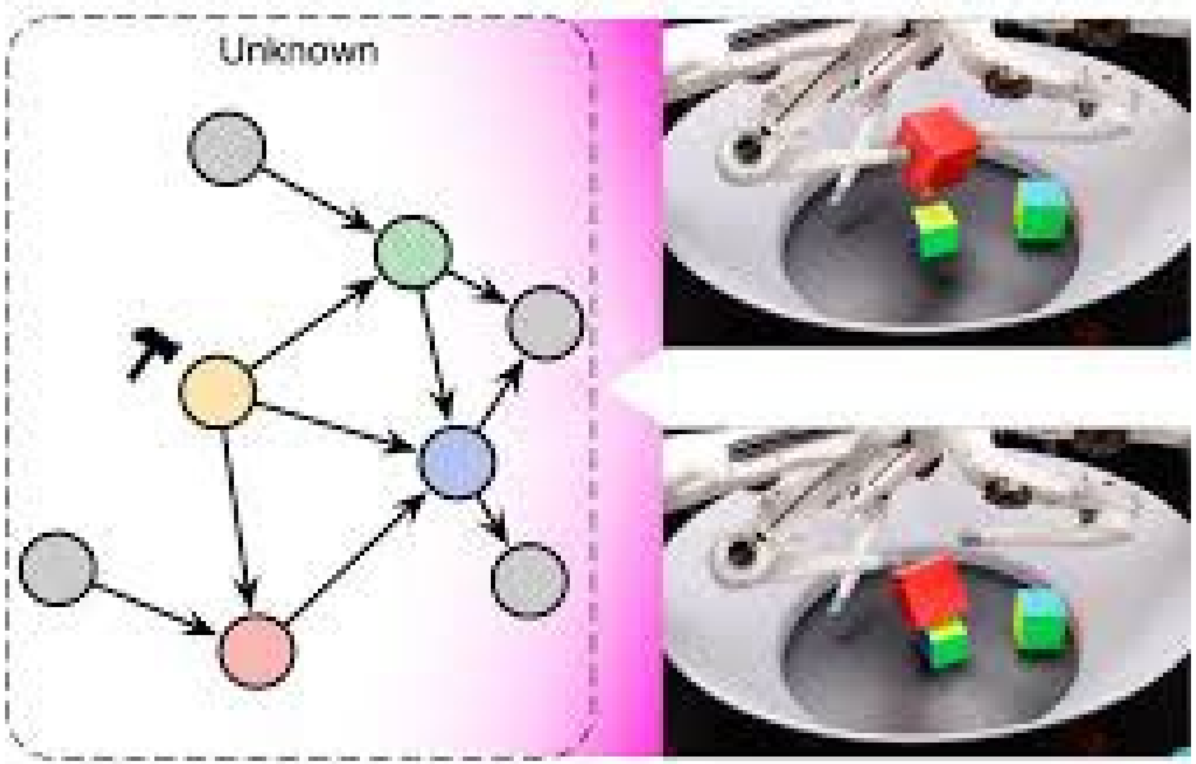
1. Discover graph structure
    - PC, NOTEARS, LiNGAM, DAG-GNN
2. Learn neural mechanisms as **generative models**:

$$f_i(\cdot) = NN(Pa(X_i), \epsilon_i)$$

- Counterfactual Reasoning

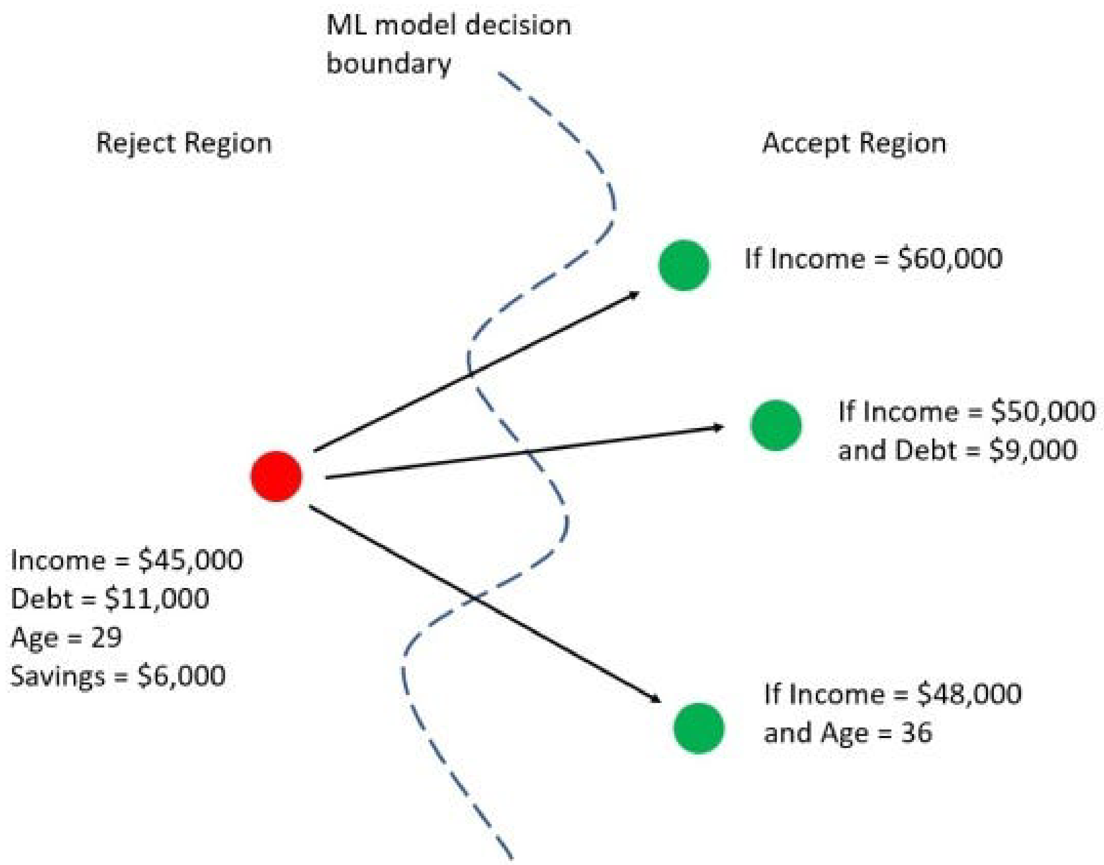
1. Intervention: Change noise or mechanism
2. Counterfactual: Re-generate outcomes

Causal Representation Learning



Schölkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A., & Bengio, Y. (2021). Toward causal representation learning. *Proceedings of the IEEE*, *109*(5), 612-634.

Explainable ML – Counterfactual Explanations



Many other applications

- Fair machine learning
- Reinforcement learning
- Transfer learning and multi-task learning
- Robust machine learning
- …