Efficient Neural Causal Discovery without Acyclicity Constraints

Phillip Lippe\textsuperscript{1}, Taco Cohen\textsuperscript{2}, Efstratios Gavves\textsuperscript{1}

\textsuperscript{1}University of Amsterdam, QUVA lab; \textsuperscript{2}Qualcomm AI Research, The Netherlands

### Problem statement
- Learn causal relations between variables as a directed, acyclic graph (DAG) from observational and interventional data
- Assumptions: interventions are sparse (only one variable at a time), soft (distribution over values), perfect (new distribution independent of original parents), and available for all variables.
- Continuous-approximation methods are promising due to their efficiency, but acyclicity needs to be ensured by constrained optimization methods which are slow and sensitive to hyperparameters, or regularizers without guarantees

### ENCO – Efficient Neural Causal Discovery

- **Distribution fitting**
  - \( \hat{L} = -E_{x,x'}(E_{y/y'}(E_{y/x}(E_{y/x'}(y/x')))) + \lambda_{\text{sparse}} \sum_{i,j} \sum_{x'} \sigma(\gamma_{ij}) \cdot \sigma(\theta_{ij}) \)
  - Fit NNs to conditional, observational distributions

- **Graph fitting**
  - Learn edge and orientation parameters based on fitted distributions

### Graph optimization
- Unbiased, low-variance gradient estimator for \( \theta \) and \( \gamma \) via REINFORCE and Monte-Carlo sampling
- Intuition: sample interventional data and \( K \) graphs, and check for each edge whether its existence improved the child’s likelihood estimate
- Phenomenon can be detected by recording observational and interventional gradients on \( y \) separately, and combine in a score function:
  \( \Delta = \gamma \cdot \theta \)
- Orientations only updated on interventions

### Latent confounders
- Latent confounders between two observable variables without direct causal relation cause a unique pattern in the graph parameters
- An edge between the two variables is disadvantaged on interventional data but beneficial when intervening on any other variable
- Continuous-approximation methods are promising due to their efficiency, but acyclicity needs to be ensured by constrained optimization methods which are slow and sensitive to hyperparameters, or regularizers without guarantees

### Experiments
- **Synthetic graphs**
  - 25 nodes
  - 5k observational, 200 interventional samples
  - ENCO recovers 4 out of 6 without less than one mistake on average (SHD)

### Takeaways
- Splitting graph parameters into edge existence and orientation for greater control over graphs without acyclicity constraints
- Efficient graph optimization using low-variance gradient estimators by testing generalization to interventions
- Convergence guarantees can be given when interventions on all variables are provided
- ENCO reliably and efficiently recovers causal graphs with up to 1000 variables, including latent confounders