ACNN 2020 Talk:

Learning Directed Acyclic Graphs from Group-Structure Data

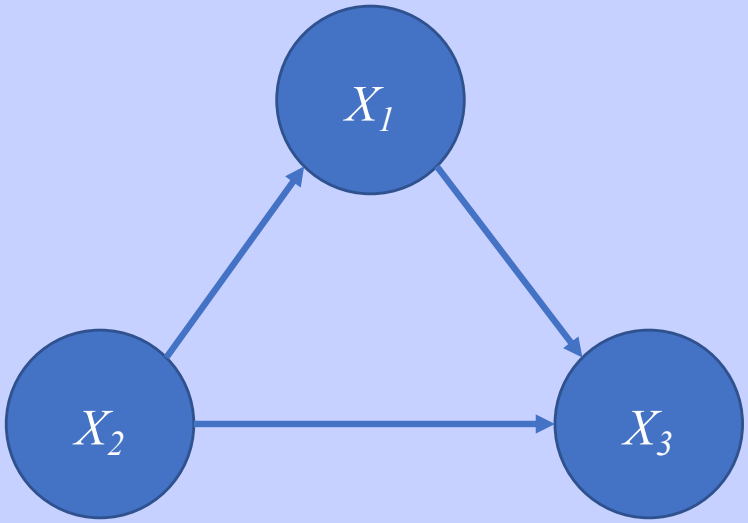
Yongsu Lee

Department of Statistics

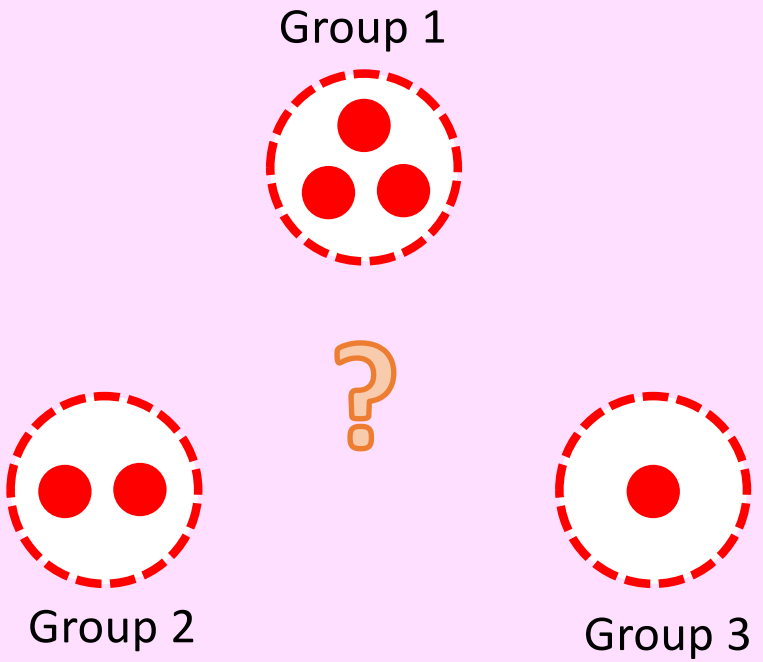
University of Wisconsin-Madison

Sep 4, 2020

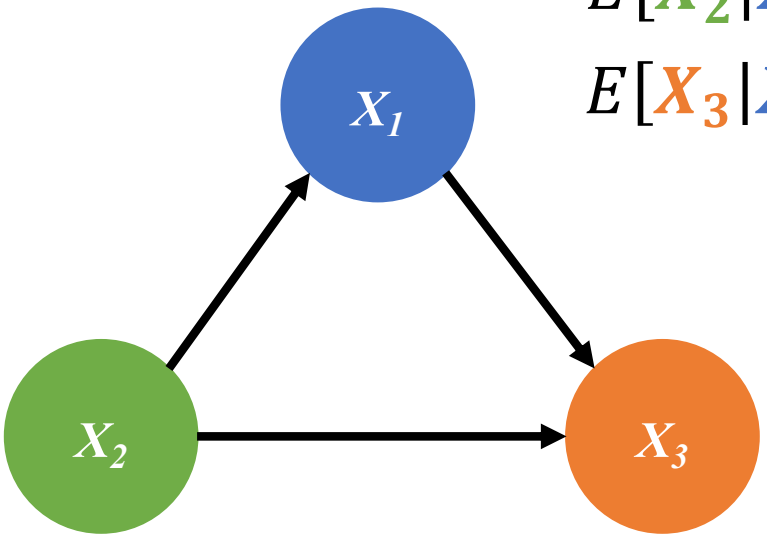
Usually, for learning DAG ...



What if ... ?



Zheng *et al.* (2018) suggests NOTEARS method …



$$E[X_1|X_2, X_3] = \beta_{01} + \beta_{21}X_2 + \beta_{31}X_3$$
$$E[X_2|X_1, X_3] = \beta_{02} + \beta_{12}X_1 + \beta_{32}X_3$$
$$E[X_3|X_1, X_2] = \beta_{03} + \beta_{13}X_1 + \beta_{23}X_2$$

$$\mathbf{W} = \begin{bmatrix} \blacksquare & \beta_{12} & \beta_{13} \\ \beta_{21} & \blacksquare & \beta_{23} \\ \beta_{31} & \beta_{32} & \blacksquare \end{bmatrix}$$

Loss function Lasso Penalized term

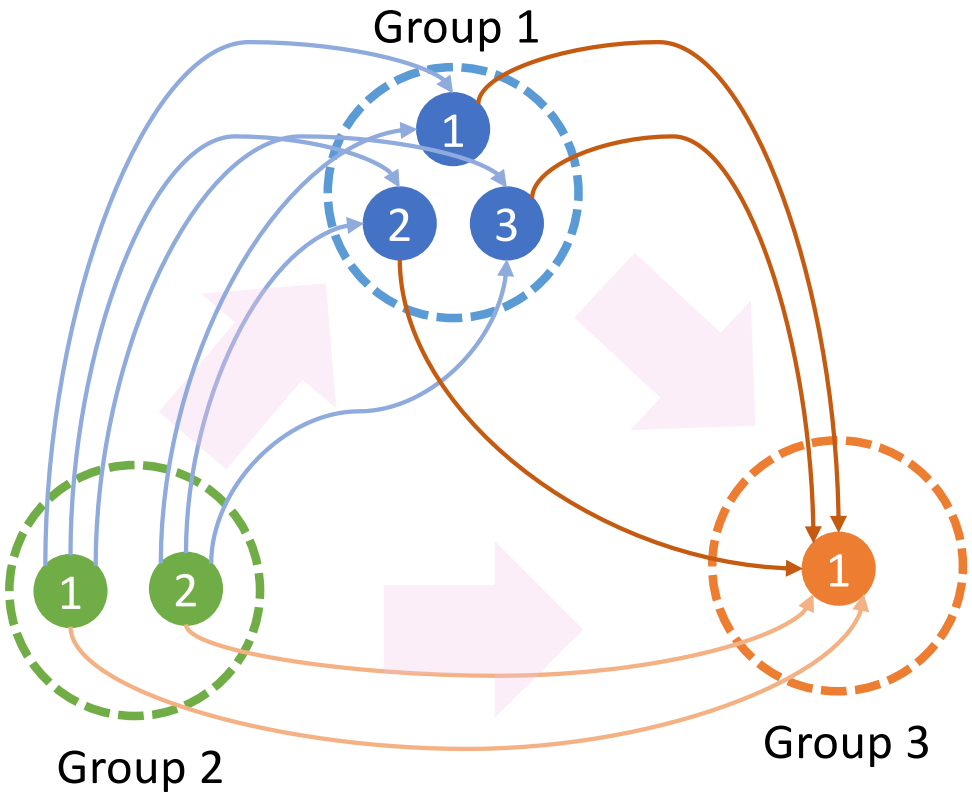
$$argmin_{\mathbf{W}} \; \ell(\mathbf{W}) + \lambda ||\mathbf{W}||_1$$

elementwise product

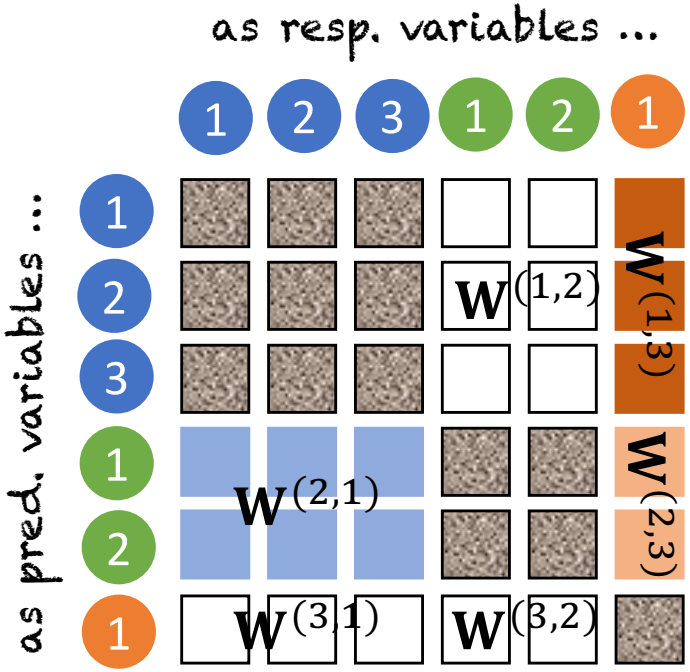
$$subject\ to \; tr\left(e^{\mathbf{W} \circ \mathbf{W}}\right) - p = 0$$

# of nodes

Acyclicity restriction



$$\mathbf{W} =$$



$$\mathbf{A}(\mathbf{W} \circ \mathbf{W})\mathbf{B} = \begin{bmatrix} ||\mathbf{W}^{(1,1)}||_F & ||\mathbf{W}^{(1,2)}||_F & ||\mathbf{W}^{(1,3)}||_F \\ ||\mathbf{W}^{(2,1)}||_F & ||\mathbf{W}^{(2,2)}||_F & ||\mathbf{W}^{(2,3)}||_F \\ ||\mathbf{W}^{(3,1)}||_F & ||\mathbf{W}^{(3,2)}||_F & ||\mathbf{W}^{(3,3)}||_F \end{bmatrix}$$

Index matrices with proper dimensions

Newly proposed NOTEARGIS method is ….

Loss function — Group Lasso

$$argmin_{\mathbf{W}} \; \ell(\mathbf{W}) + \lambda \Sigma_{i \neq j} ||\mathbf{W}^{(i,j)}||_F$$

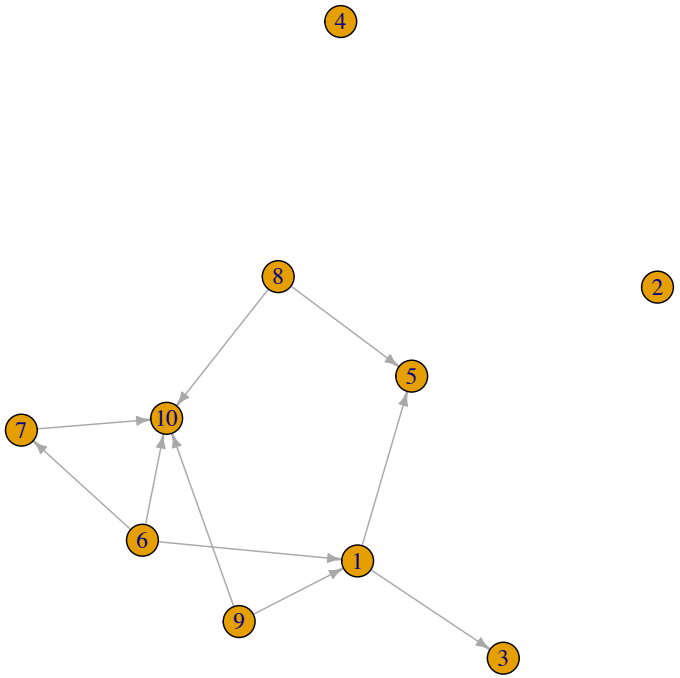
$$subject\ to \; \underbrace{tr\left(e^{\mathbf{A}(\mathbf{W} \circ \mathbf{W})\mathbf{B}}\right) - d = 0}_{\text{acyclicity restriction among groups}}$$

# of groups

Why NOTEARGIS ?

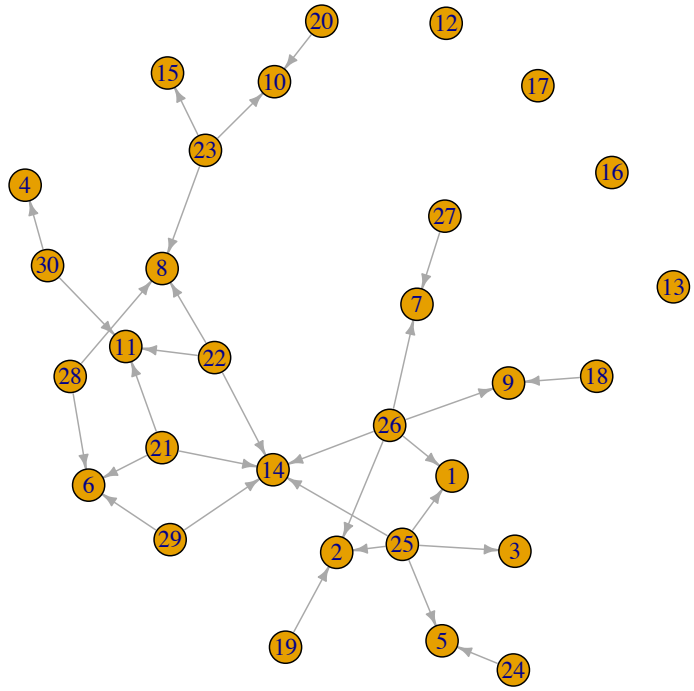
- Easily applicable to any type of variables
- Can be extended to a mixed-type DAG network learning.
- More precise result for structure learning.

Grouped Gaussian Data
10 groups(nodes) – 2 elements for each
50 observations; 50 simulation iterations



Method	E	TPR	FDR	SHD
NOTEARGIS	7.04	0.7	0.3	5.34
NOTEARS	3.06	0.31	0.4	8.48

Multi-level Data
30 nodes – 2, 3, or 4 levels
50 observations; 50 simulation iterations



Method	E	TPR	FDR	SHD
NOTEARGIS	7.88	0.26	0.68	45.52
CD	4.46	0.15	0.9	70.94