CEP（**c**ause-**e**ffect **p**airs） ベンチマーク

scatter_0001.png	scatter_0002.png	scatter_0003.png	scatter_0004.png	scatter_0005.png	scatter_0006.png	scatter_0007.png	scatter_0008.png	scatter_0009.png
scatter_0010.png	scatter_0011.png	scatter_0012.png	scatter_0013.png	scatter_0014.png	scatter_0015.png	scatter_0016.png	scatter_0017.png	scatter_0018.png
scatter_0019.png	scatter_0020.png	scatter_0021.png	scatter_0022.png	scatter_0023.png	scatter_0024.png	scatter_0025.png	scatter_0026.png	scatter_0027.png
scatter_0028.png	scatter_0029.png	scatter_0030.png	scatter_0031.png	scatter_0032.png	scatter_0033.png	scatter_0034.png	scatter_0035.png	scatter_0036.png
scatter_0037.png	scatter_0038.png	scatter_0041.png	scatter_0042.png	scatter_0043.png	scatter_0044.png	scatter_0045.png	scatter_0046.png	

Database with cause-effect pairs

https://webdav.tuebingen.mpg.de/cause-effect/README
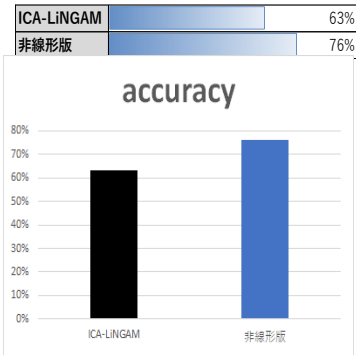
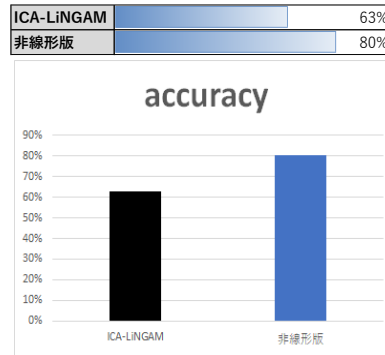J. M. Mooij, J. Peters, D. Janzing, J. Zscheischler, B. Schoelkopf

"Distinguishing cause from effect using observational data: methods and benchmarks"

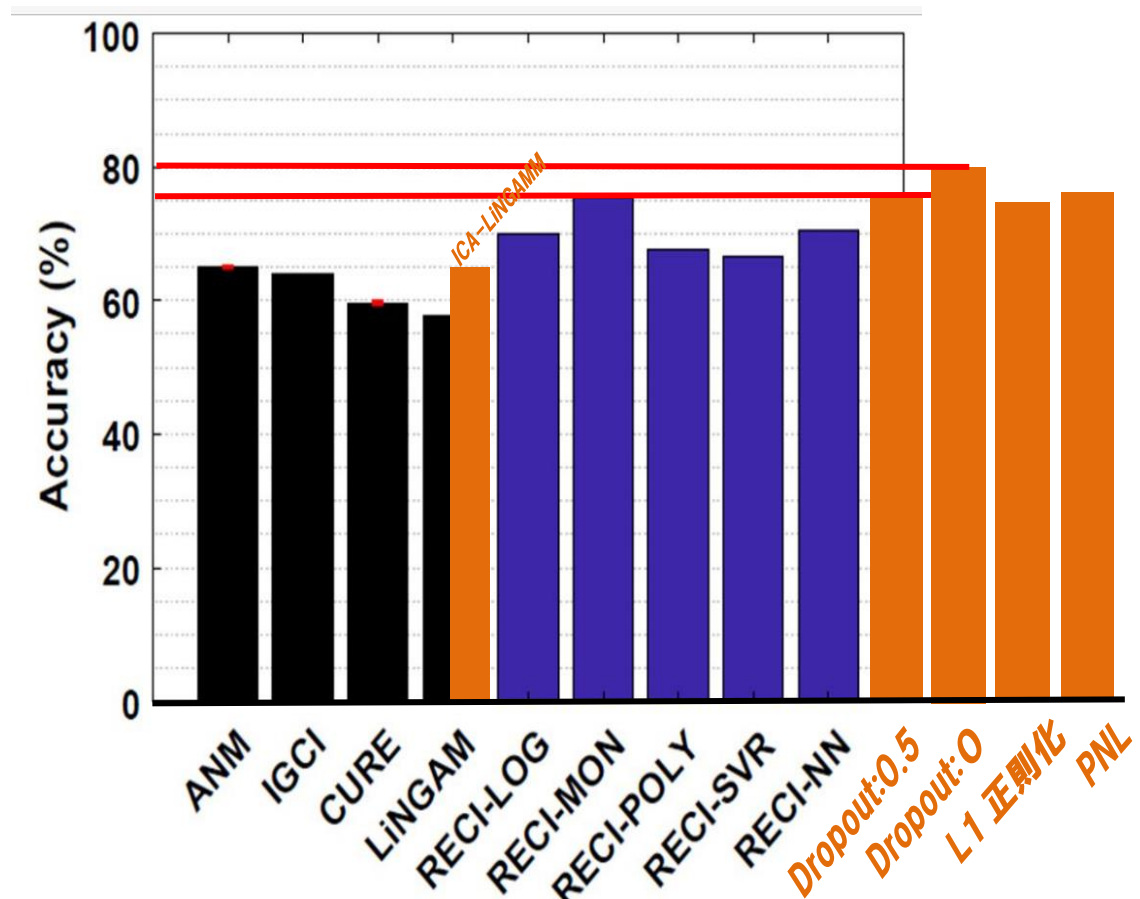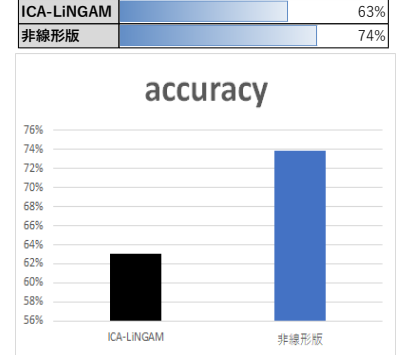Journal of Machine Learning Research 17(32):1-102, 2016

Dropout:0.5, PNL

ICA-LiNGAM		63%
非線形版		76%

accuracy

Dropout:0

ICA-LiNGAM		63%
非線形版		80%

accuracy

L1 正則化

ICA-LiNGAM		63%
非線形版		74%

accuracy

元グラフ引用）非線形性に基づく大規模因果推論原理・手法の研究

https://kaken.nii.ac.jp/ja/grant/KAKENHI-PROJECT-17K00305/

Dropout:0.5

	X	Y	ground truth	ICA-LiNGAM		非線形	
pair0001	Altitude	Temperature	->	❌	0	✅	1
pair0002	Altitude	Precipitation	->	❌	0	✅	1
pair0003	Longitude	Temperature	->	✅	1	✅	1
pair0004	Altitude	Sunshine hours	->	✅	1	✅	1
pair0005	Age	Length	->	❌	0	✅	1
pair0006	Age	Shell weight	->	❌	0	❌	0
pair0007	Age	Diameter	->	❌	0	✅	1
pair0008	Age	Height	->	❌	0	❌	0
pair0009	Age	Whole weight	->	❌	0	❌	0
pair0010	Age	Shucked weight	->	❌	0	❌	0
pair0011	Age	Viscera weight	->	❌	0	❌	0
pair0012	Age	Wage per hour	->	✅	1	✅	1
pair0013	Displacem	Fuel consumption	->	✅	1	✅	1
pair0014	Horse pow	Fuel consumption	->	❌	0	❌	0
pair0015	Weight	Fuel consumption	->	✅	1	✅	1
pair0016	Horsepow	Acceleration	->	✅	1	✅	1
pair0017	Age	Dividends from stocks	->	✅	1	✅	1
pair0018	Age	Concentration GAG	->	✅	1	✅	1
pair0019	Current du	Next interval	->	✅	1	✅	1
pair0020	Latitude	Temperature	->	✅	1	✅	1
pair0021	Longitude	Precipitation	->	✅	1	✅	1
pair0022	Age	Height	->	❌	0	✅	1
pair0023	Age	Weight	->	❌	0	✅	1
pair0024	Age	Heart rate	->	❌	0	✅	1
pair0025	Cement	Compressive strength	->	✅	1	✅	1
pair0026	Blast furna	Compressive strength	->	✅	1	✅	1
pair0027	Fly ash	Compressive strength	->	✅	1	✅	1
pair0028	Water	Compressive strength	->	❌	0	✅	1
pair0029	Superplas	Compressive strength	->	✅	1	✅	1
pair0030	Coarse ag	Compressive strength	->	✅	1	✅	1
pair0031	Fine aggre	Compressive strength	->	✅	1	✅	1
pair0032	Age	Compressive strength	->	❌	0	❌	0
pair0033	Alcohol co	Mean corpuscular volume	->	✅	1	❌	0
pair0034	Alcohol co	Alkaline phosphotase	->	✅	1	✅	1
pair0035	Alcohol co	Alanine aminotransferase	->	✅	1	✅	1
pair0036	Alcohol co	Aspartate aminotransferase	->	✅	1	❌	0
pair0037	Alcohol co	Gamma-glutamyl transpeptdase	->	✅	1	❌	0
pair0038	Age	Body mass index	->	❌	0	✅	1
pair0039	Age	Serum insulin	->	✅	1	✅	1
pair0040	Age	Diastolic blood pressure	->	✅	1	✅	1
pair0041	Age	Plasma glucose concentration	->	❌	0	❌	0
pair0042	Day of the	Temperature	->	✅	1	✅	1
pair0043	Temperatu	Temperature at t+1	->	✅	1	✅	1
pair0044	Pressure a	Pressure at t+1	->	✅	1	✅	1
pair0045	Sea level	Sea level pressure at t+1	->	✅	1	✅	1
pair0046	Relative h	Relative humidity at t+1	->	✅	1	✅	1
				63%		**76%**	

Dropout:無し

	X	Y	ground truth	ICA-LiNGAM		非線形	
pair0001	Altitude	Temperature	->	❌	0	✅	1
pair0002	Altitude	Precipitation	->	❌	0	✅	1
pair0003	Longitude	Temperature	->	✅	1	✅	1
pair0004	Altitude	Sunshine hours	->	✅	1	✅	1
pair0005	Age	Length	->	❌	0	✅	1
pair0006	Age	Shell weight	->	❌	0	✅	1
pair0007	Age	Diameter	->	❌	0	✅	1
pair0008	Age	Height	->	❌	0	❌	0
pair0009	Age	Whole weight	->	❌	0	❌	0
pair0010	Age	Shucked weight	->	❌	0	❌	0
pair0011	Age	Viscera weight	->	❌	0	❌	0
pair0012	Age	Wage per hour	->	✅	1	✅	1
pair0013	Displacem	Fuel consumption	->	✅	1	✅	1
pair0014	Horse pow	Fuel consumption	->	❌	0	❌	0
pair0015	Weight	Fuel consumption	->	✅	1	✅	1
pair0016	Horsepow	Acceleration	->	✅	1	❌	0
pair0017	Age	Dividends from stocks	->	✅	1	✅	1
pair0018	Age	Concentration GAG	->	✅	1	✅	1
pair0019	Current du	Next interval	->	✅	1	❌	0
pair0020	Latitude	Temperature	->	✅	1	✅	1
pair0021	Longitude	Precipitation	->	✅	1	✅	1
pair0022	Age	Height	->	❌	0	✅	1
pair0023	Age	Weight	->	❌	0	✅	1
pair0024	Age	Heart rate	->	❌	0	✅	1
pair0025	Cement	Compressive strength	->	✅	1	✅	1
pair0026	Blast furna	Compressive strength	->	✅	1	✅	1
pair0027	Fly ash	Compressive strength	->	✅	1	✅	1
pair0028	Water	Compressive strength	->	❌	0	✅	1
pair0029	Superplas	Compressive strength	->	✅	1	✅	1
pair0030	Coarse ag	Compressive strength	->	✅	1	✅	1
pair0031	Fine aggre	Compressive strength	->	✅	1	✅	1
pair0032	Age	Compressive strength	->	❌	0	✅	1
pair0033	Alcohol co	Mean corpuscular volume	->	✅	1	✅	1
pair0034	Alcohol co	Alkaline phosphotase	->	✅	1	✅	1
pair0035	Alcohol co	Alanine aminotransferase	->	✅	1	✅	1
pair0036	Alcohol co	Aspartate aminotransferase	->	✅	1	❌	0
pair0037	Alcohol co	Gamma-glutamyl transpeptdase	->	✅	1	❌	0
pair0038	Age	Body mass index	->	❌	0	✅	1
pair0039	Age	Serum insulin	->	✅	1	✅	1
pair0040	Age	Diastolic blood pressure	->	✅	1	✅	1
pair0041	Age	Plasma glucose concentration	->	❌	0	✅	1
pair0042	Day of the	Temperature	->	✅	1	✅	1
pair0043	Temperatu	Temperature at t+1	->	✅	1	✅	1
pair0044	Pressure a	Pressure at t+1	->	✅	1	✅	1
pair0045	Sea level p	Sea level pressure at t+1	->	✅	1	✅	1
pair0046	Relative h	Relative humidity at t+1	->	✅	1	✅	1
				63%		**80%**	

L1 正則化

	X	Y	ground truth	ICA-LiNGAM		非線形	
pair0001	Altitude	Temperature	->	❌	0	✅	1
pair0002	Altitude	Precipitation	->	❌	0	✅	1
pair0003	Longitude	Temperature	->	✅	1	✅	1
pair0004	Altitude	Sunshine hours	->	✅	1	✅	1
pair0005	Age	Length	->	❌	0	✅	1
pair0006	Age	Shell weight	->	❌	0	❌	0
pair0007	Age	Diameter	->	❌	0	✅	1
pair0008	Age	Height	->	❌	0	❌	0
pair0009	Age	Whole weight	->	❌	0	❌	0
pair0010	Age	Shucked weight	->	❌	0	❌	0
pair0011	Age	Viscera weight	->	❌	0	❌	0
pair0012	Age	Wage per hour	->	✅	1	✅	1
pair0013	Displacem	Fuel consumption	->	✅	1	✅	1
pair0014	Horse pow	Fuel consumption	->	❌	0	❌	0
pair0015	Weight	Fuel consumption	->	✅	1	✅	1
pair0016	Horsepow	Acceleration	->	✅	1	❌	0
pair0017	Age	Dividends from stocks	->	✅	1	✅	1
pair0018	Age	Concentration GAG	->	✅	1	❌	0
pair0019	Current du	Next interval	->	✅	1	✅	1
pair0020	Latitude	Temperature	->	✅	1	✅	1
pair0021	Longitude	Precipitation	->	✅	1	✅	1
pair0022	Age	Height	->	❌	0	✅	1
pair0023	Age	Weight	->	❌	0	✅	1
pair0024	Age	Heart rate	->	❌	0	✅	1
pair0025	Cement	Compressive strength	->	✅	1	✅	1
pair0026	Blast furna	Compressive strength	->	✅	1	✅	1
pair0027	Fly ash	Compressive strength	->	✅	1	✅	1
pair0028	Water	Compressive strength	->	❌	0	✅	1
pair0029	Superplas	Compressive strength	->	✅	1	✅	1
pair0030	Coarse ag	Compressive strength	->	✅	1	✅	1
pair0031	Fine aggre	Compressive strength	->	✅	1	✅	1
pair0032	Age	Compressive strength	->	❌	0	✅	1
pair0033	Alcohol co	Mean corpuscular volume	->	✅	1	❌	0
pair0034	Alcohol co	Alkaline phosphotase	->	✅	1	✅	1
pair0035	Alcohol co	Alanine aminotransferase	->	✅	1	✅	1
pair0036	Alcohol co	Aspartate aminotransferase	->	✅	1	❌	0
pair0037	Alcohol co	Gamma-glutamyl transpeptdase	->	✅	1	❌	0
pair0038	Age	Body mass index	->	❌	0	✅	1
pair0039	Age	Serum insulin	->	✅	1	✅	1
pair0040	Age	Diastolic blood pressure	->	✅	1	✅	1
pair0041	Age	Plasma glucose concentration	->	❌	0	❌	0
pair0042	Day of the	Temperature	->	✅	1	✅	1
pair0043	Temperatu	Temperature at t+1	->	✅	1	✅	1
pair0044	Pressure a	Pressure at t+1	->	✅	1	✅	1
pair0045	Sea level	Sea level pressure at t+1	->	✅	1	✅	1
pair0046	Relative h	Relative humidity at t+1	->	✅	1	✅	1
				63%		**74%**	

PNL モデル

$$g(\mathrm{x}_i) = f_i\big(\boldsymbol{pa}(\mathrm{x}_i)\big) + \epsilon_i$$

$\boldsymbol{pa}(\mathrm{x}_i)$は$\mathrm{x}_i$の親（原因）となる変数、$g_i, f_i$は非線形関数$f_i$が線形、$g_i$が恒等関数の場合は通常の LiNGAM（線形モデル）になる。親（原因）となる変数を見つければ良いが実験モデルとして以下の B 行列を想定する。B 行列は非巡回の因果関係であれば下半三角行列になるため変数の置換を与えれば任意の非巡回の因果関係を与える事が出来る。

非巡回の因果関係であれば以下の行列を定義できる。$B_{ij} = 1$。

$$\begin{pmatrix} 0 & 0 & \cdots & 0 \\ B_{21} & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ B_{n1} & B_{n2} & \cdots & 0 \end{pmatrix}$$

$$B_{ij} \neq 0$$

行方向が子（結果）で列方向が親（原因）として与える事が出来る。

未観測の潜在共通変数が存在する可能性が考えられる場合、変数x_lに対して$\mu(\mathrm{x}_i)$を指定した以下のモデルを定義する。

$$g(\mathrm{x}_i) = f_i\big(\boldsymbol{pa}(\mathrm{x}_i), \boldsymbol{\mu}(\mathrm{x}_i)\big) + \epsilon_i$$

一般化ガウス分布

$$\boldsymbol{\mu} \sim Generalized\_Gaussian(\beta, \rho, \tilde{x}) \equiv \frac{\beta^{\frac{1}{2}}}{2\Gamma(1 + \frac{1}{\rho})} \boldsymbol{exp}\left(-\beta^{\frac{1}{2}}|x - \tilde{x}|^{\rho}\right)$$

β, ρは未定のパラメータで以下の最適化によって定める。

$$mse_i = MSE(\mathrm{x}_i) = \frac{1}{n}\sum (g(\mathrm{x}_i) - f_i\big(\boldsymbol{pa}(\mathrm{x}_i), \mu(\mathrm{x}_i)\big))^2$$

$$MI(\boldsymbol{e_i}, \boldsymbol{e_j}) \equiv \iint p(\boldsymbol{e_i}, \boldsymbol{e_j}) log\left(\frac{p(\boldsymbol{e_i}, \boldsymbol{e_j})}{p(\boldsymbol{e_i})p(\boldsymbol{e_j})}\right) d\boldsymbol{e_i} d\boldsymbol{e_j}$$

MIは相互情報量で 0 であれば独立変数になる。つまり、μは残差と相互情報量がともに最小になるように未定のパラメータβ, ρを求める。