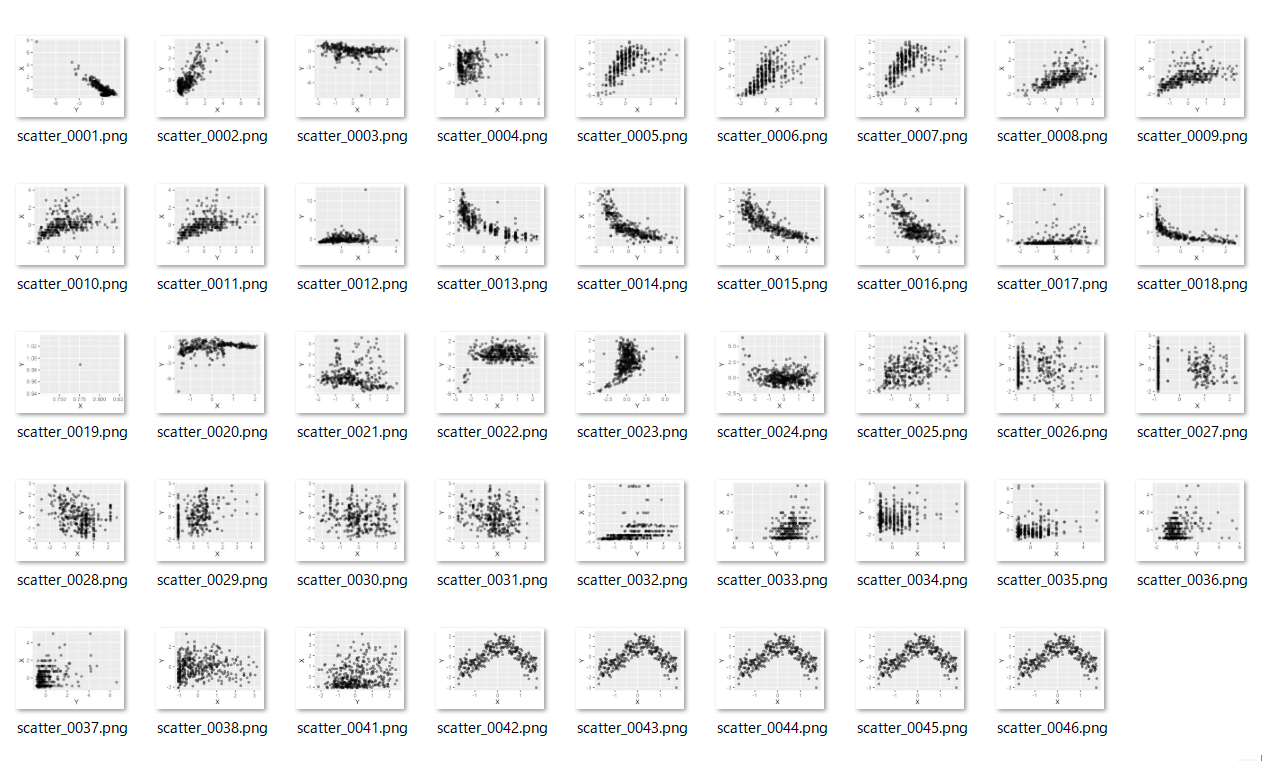
CEP（**c**ause-**e**ffect **p**airs） ベンチマーク



Database with cause-effect pairs

<https://webdav.tuebingen.mpg.de/cause-effect/README>

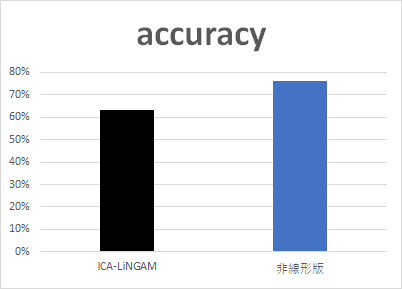
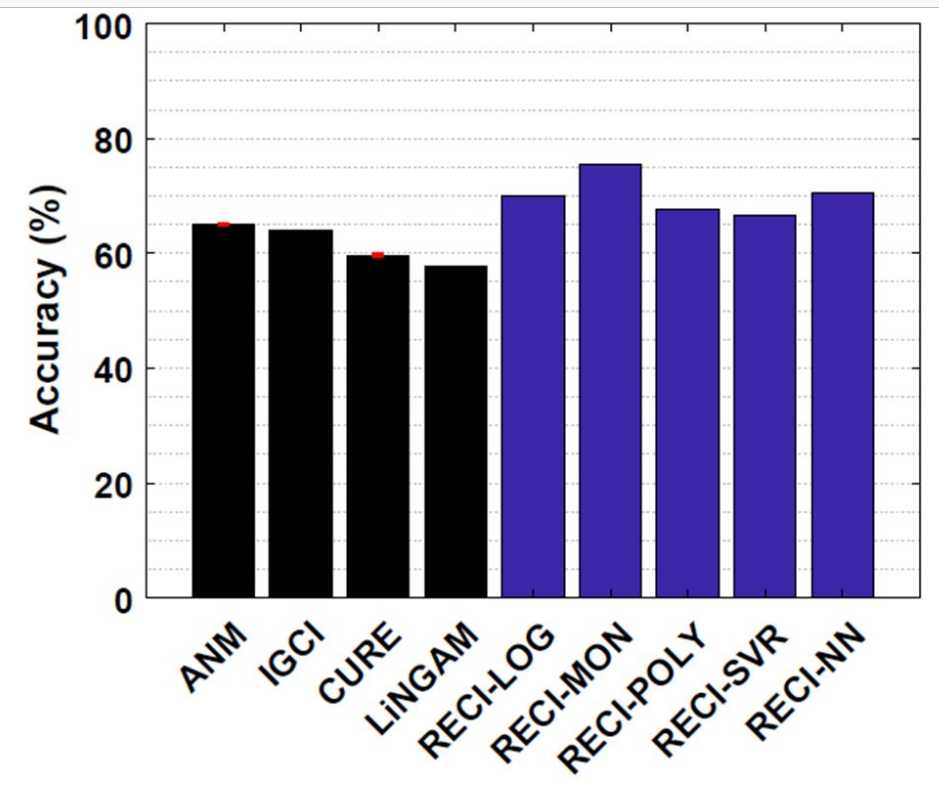
J. M. Mooij, J. Peters, D. Janzing, J. Zscheischler, B. Schoelkopf

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

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

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

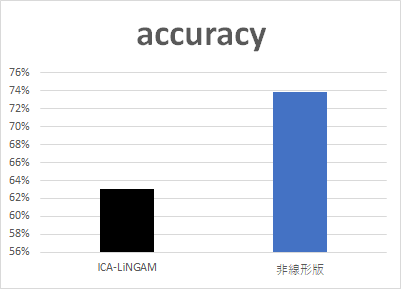
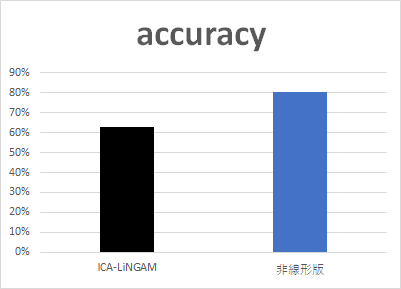
<https://kaken.nii.ac.jp/ja/grant/KAKENHI-PROJECT-17K00305/>



Dropout:0.5, PNL



Dropout:0



L1正則化

***Dropout:0.5***

***Dropout:０***

***L1正則化***

***PNL***

***ICA-LiNGAMM***

Dropout:0.5

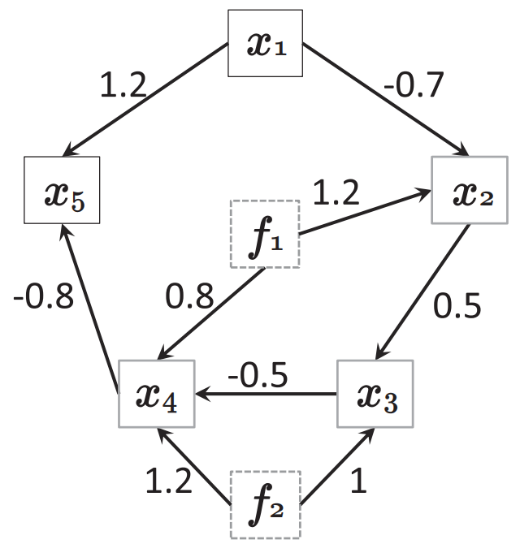


Dropout:無し

L1正則化



LiNGAM\_latest3c.csv



f1,f2が交絡変数

テストはf1,f2もデータにある場合とf1,f2がデータに無い場合をテストする。

※Pairwise LiNGAM.で検証に使われたデータ

**点数=**

**accuracy=**

線形因果探索はICA-LiNGAMを適用し、因果効果が0.01未満は削除

非線形回帰のため係数はなく代わりにfeature importanceを表記

※feature importanceが０．７以下のエッジは削除





ICA-LiNGAM

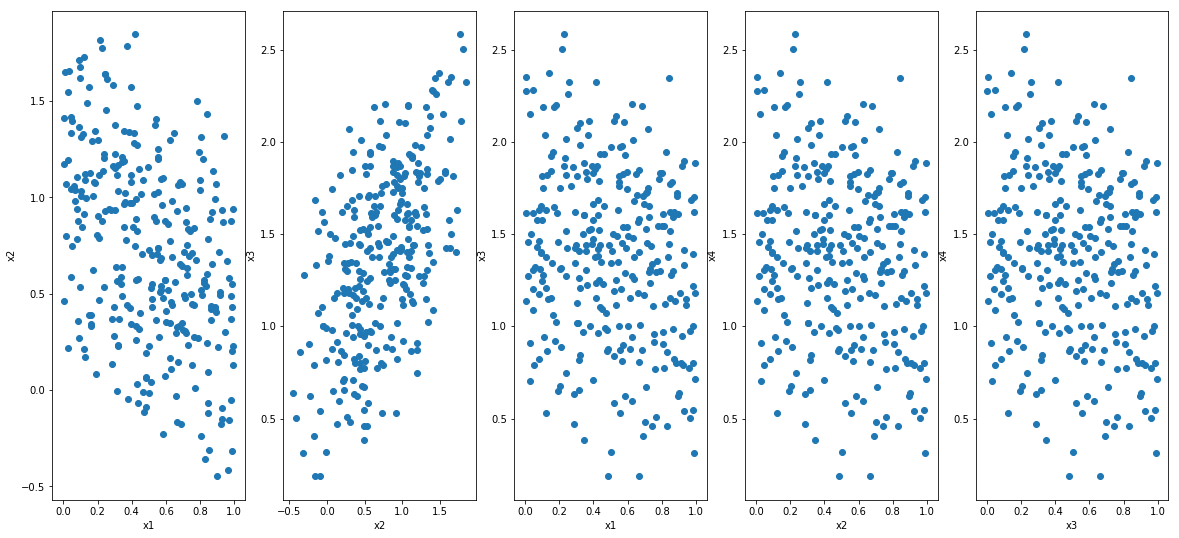


ICA-LiNGAM

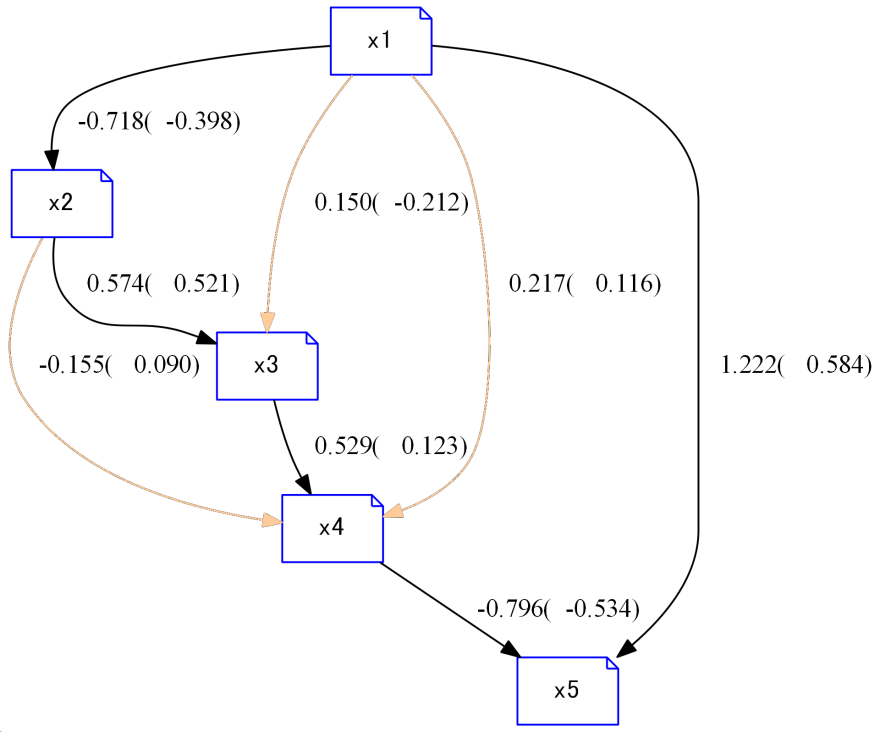
データは線形、アルゴリズムは線形を仮定

線形因果構造

LiNGAM\_latest3.csv



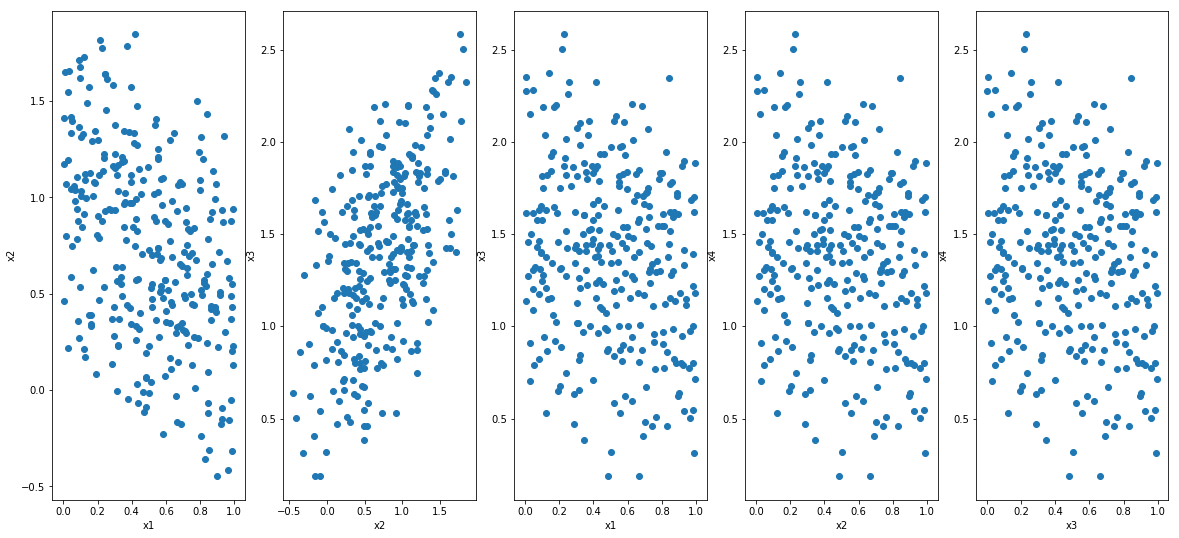




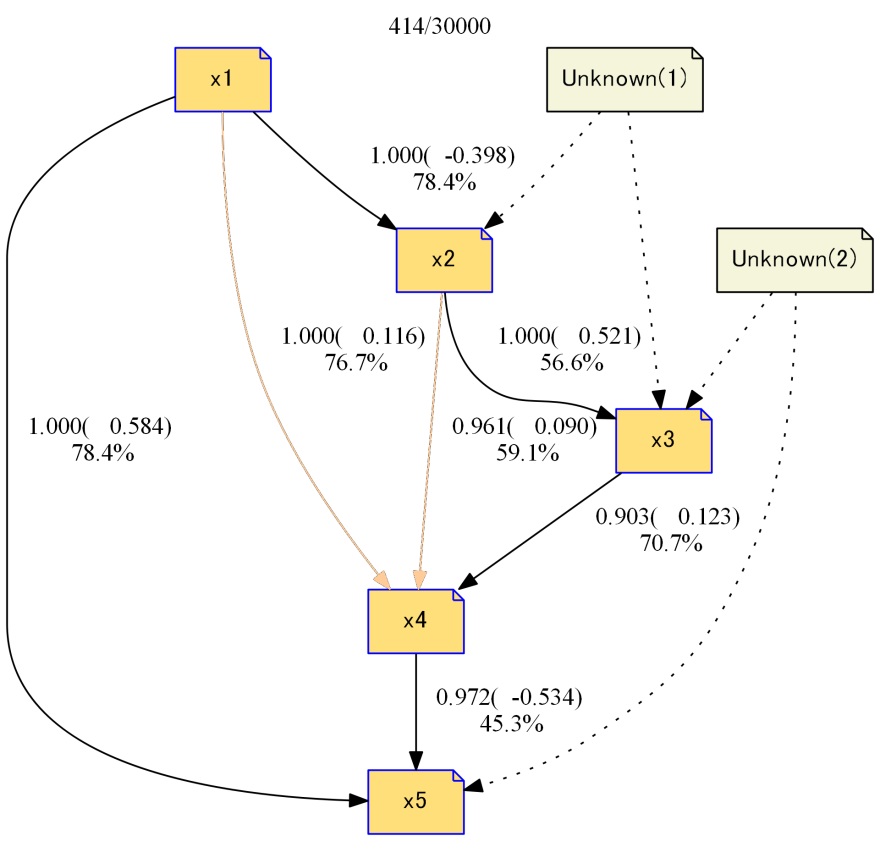
データは線形、アルゴリズムは線形を仮定しない

線形因果構造

LiNGAM\_latest3.csv



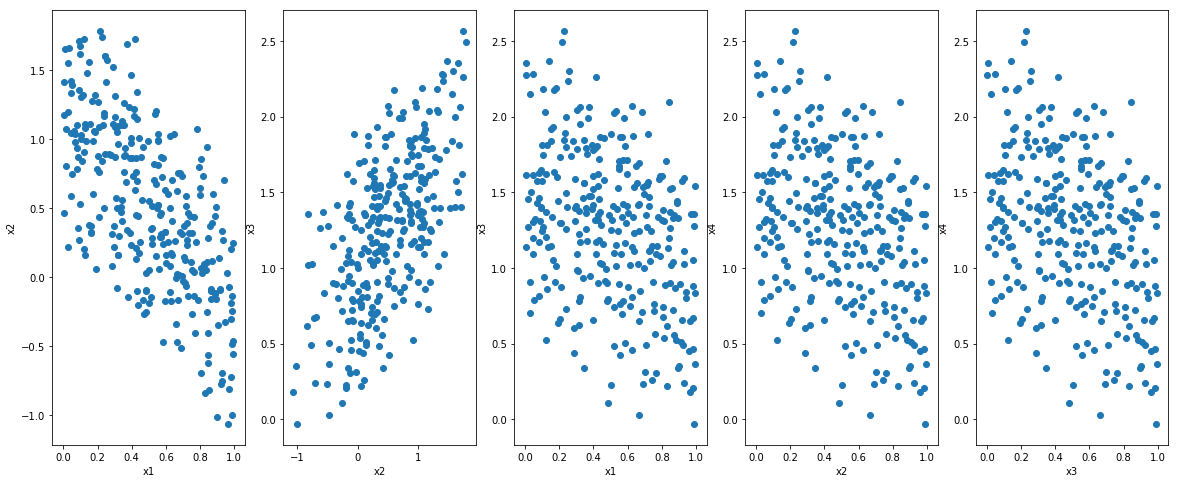




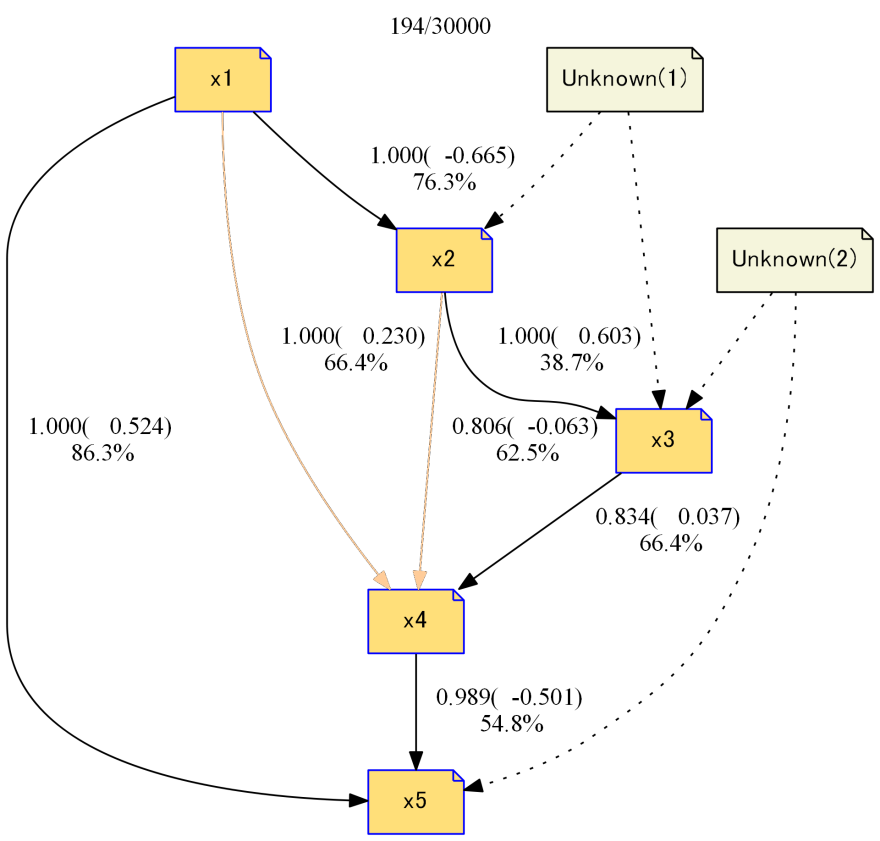
データは一部非線形、アルゴリズムは線形を仮定しない

非線形因果構造

nonlinear\_LiNGAM\_latest3c.csv



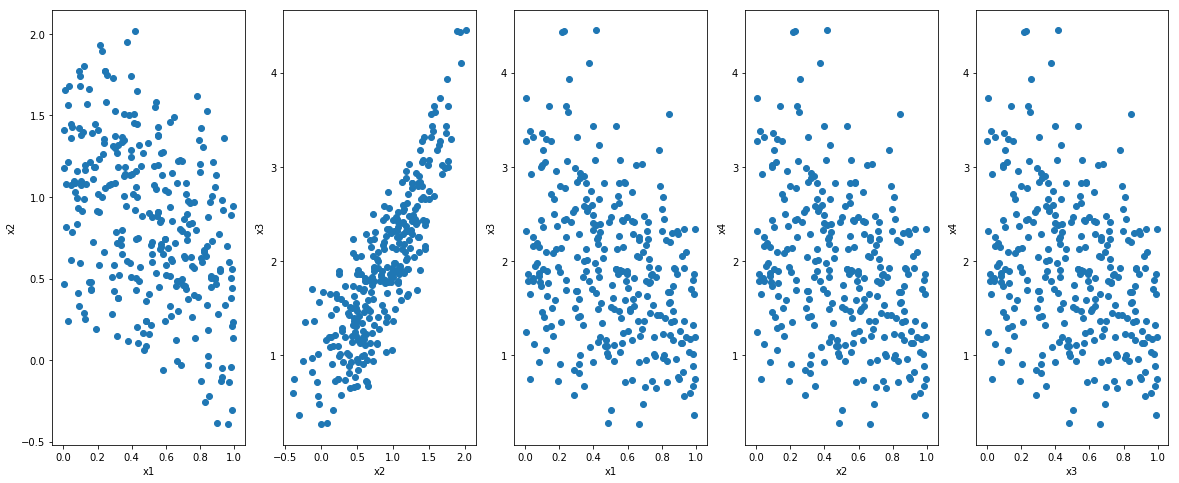




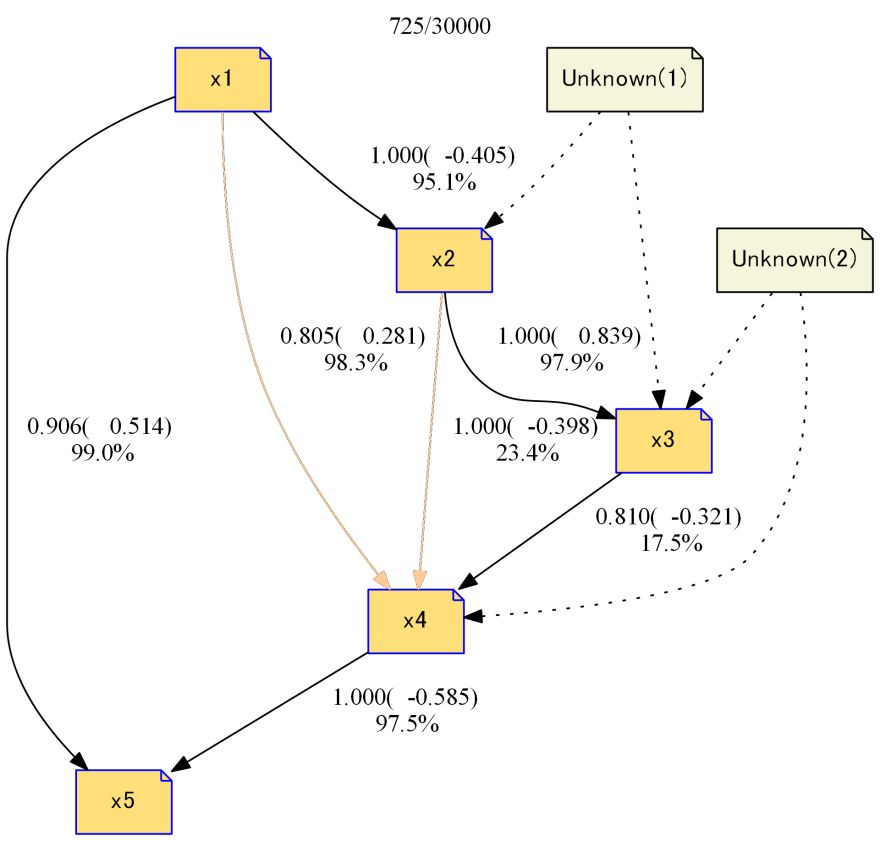
データは一部非線形、アルゴリズムは線形を仮定しない

非線形因果構造

nonlinear\_LiNGAM\_latest3b.csv



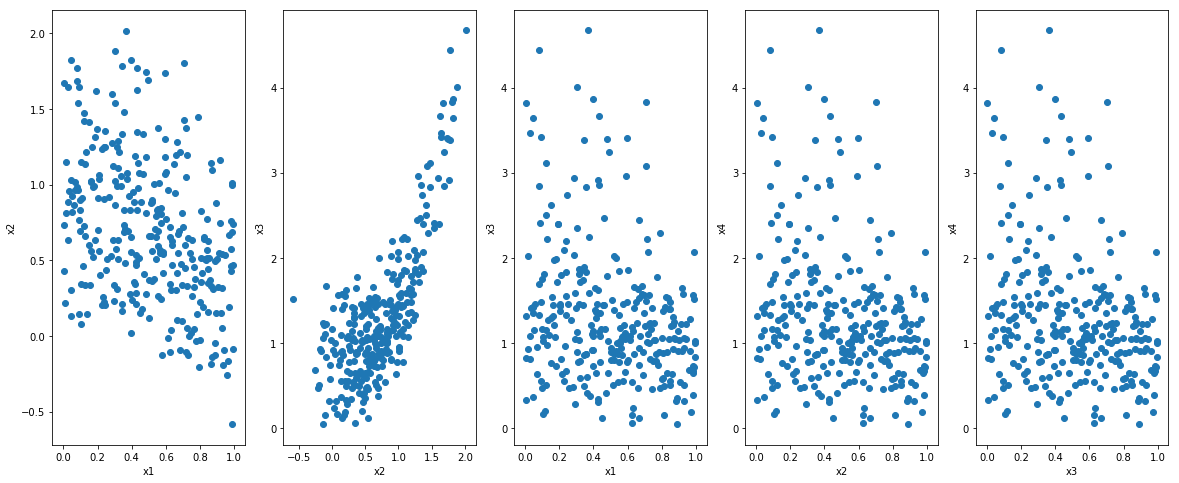




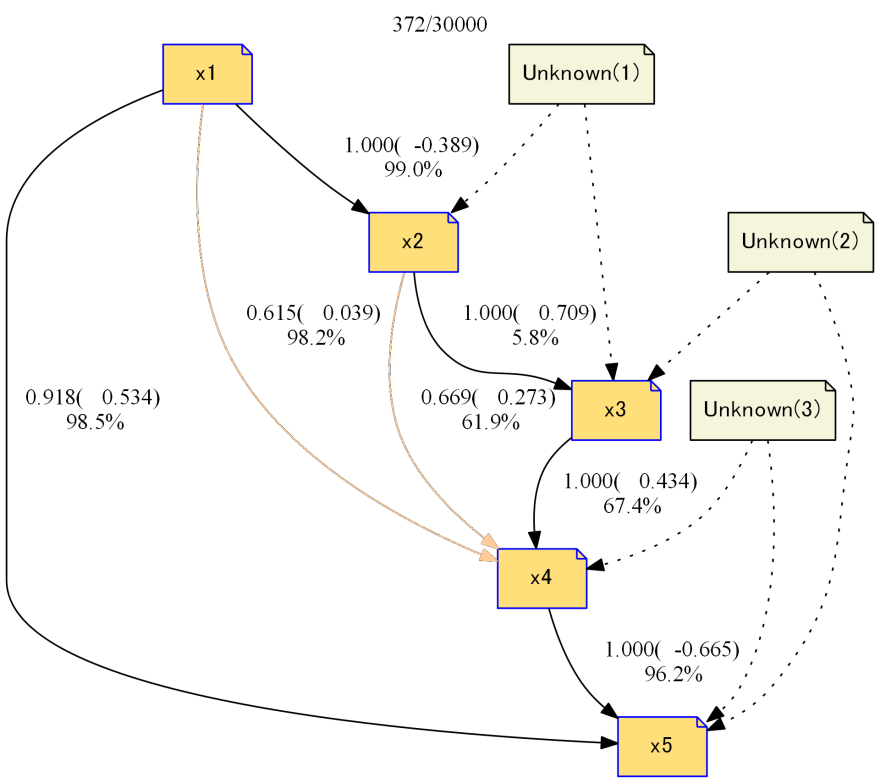
データは全部非線形、アルゴリズムは線形を仮定しない

非線形因果構造

nonlinear\_LiNGAM\_latest3a.csv



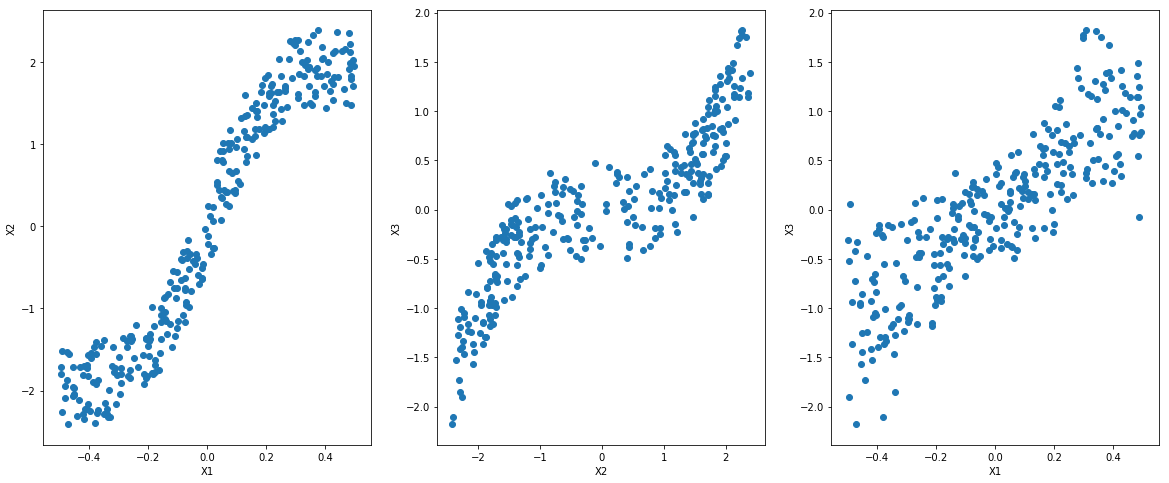




データは全部非線形、アルゴリズムは線形を仮定しない

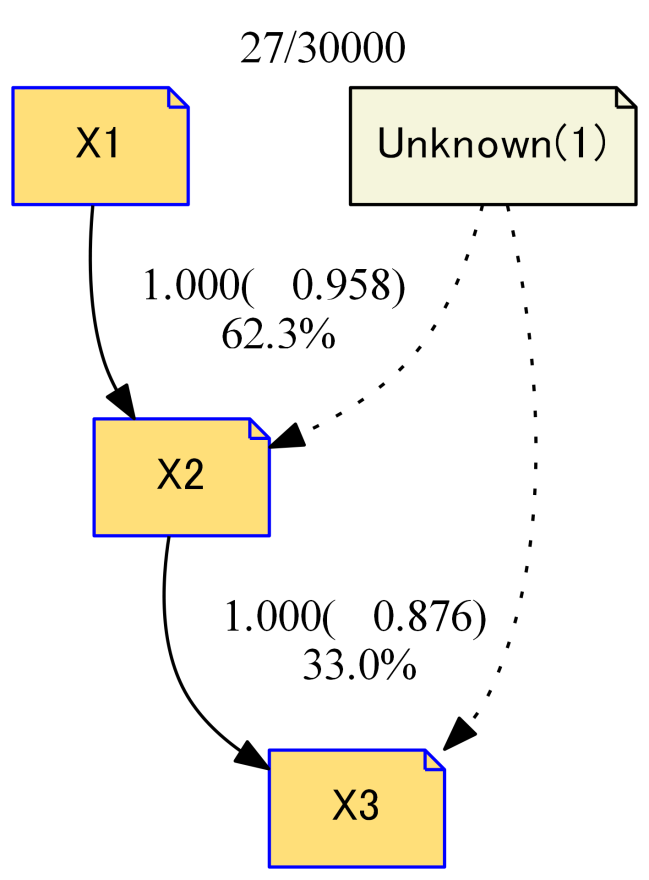
非線形因果構造

nonlinear.csv



Causal Discovery with Cascade Nonlinear Additive Noise

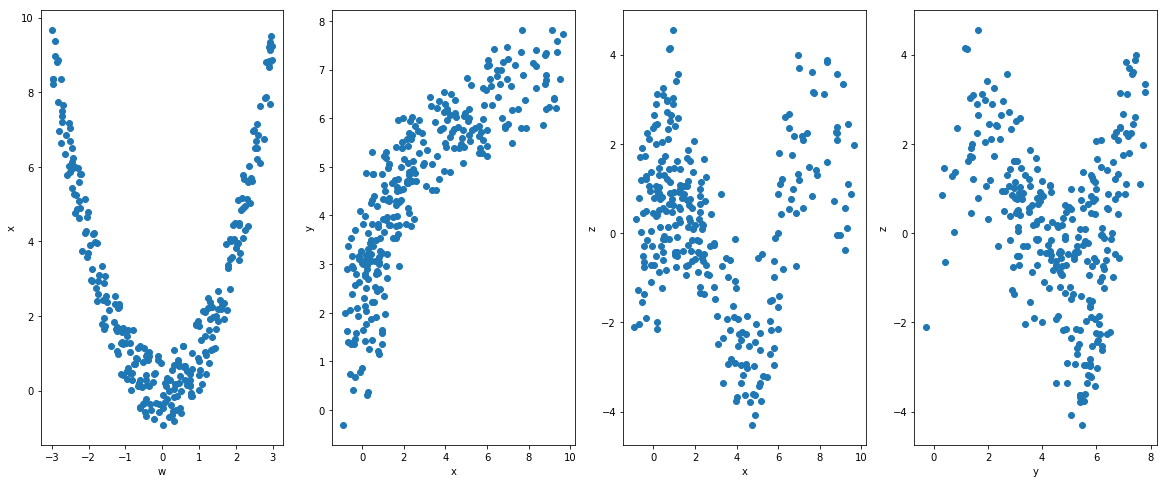


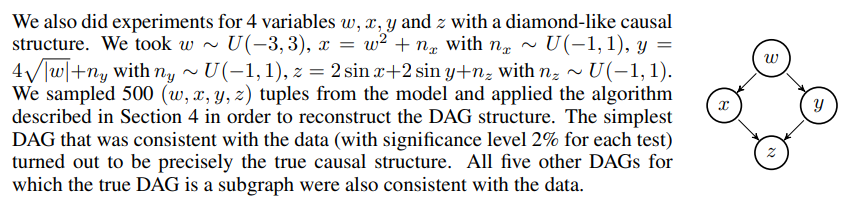


データは全部非線形、アルゴリズムは線形を仮定しない

非線形因果構造

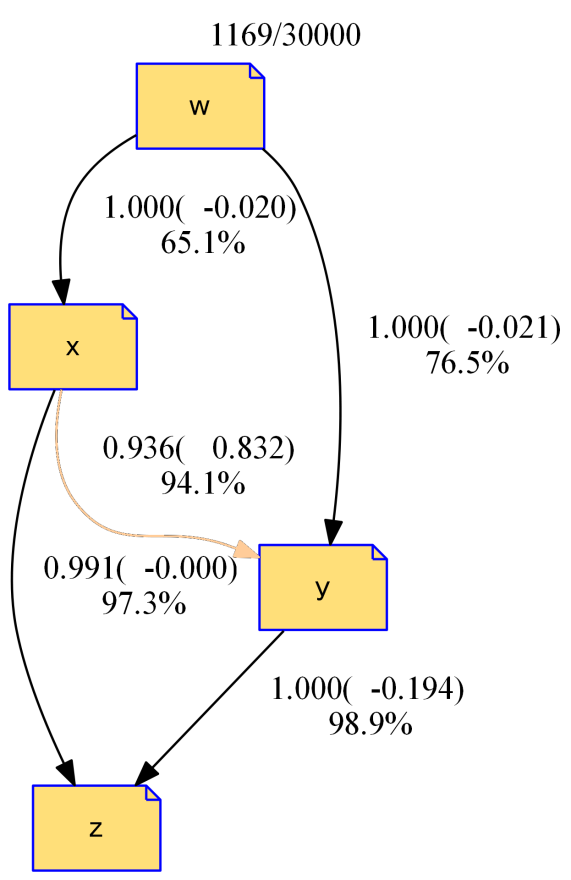
Nonlinear2.csv



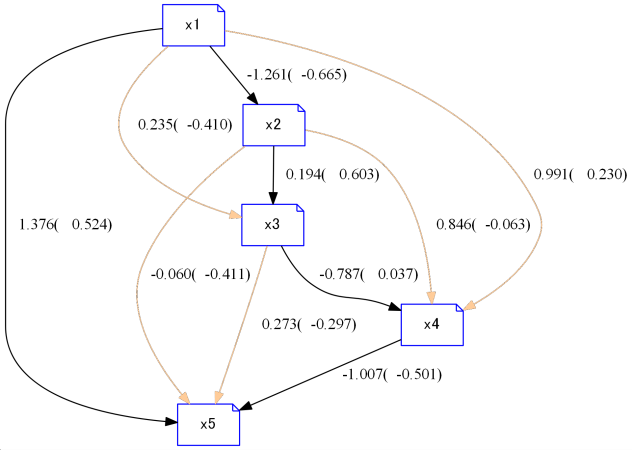


<https://webdav.tuebingen.mpg.de/causality/NIPS2008-Hoyer.pdf>





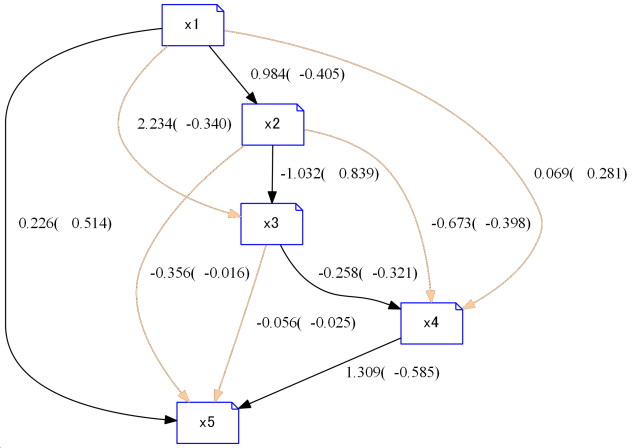
ICA-LiNGAM



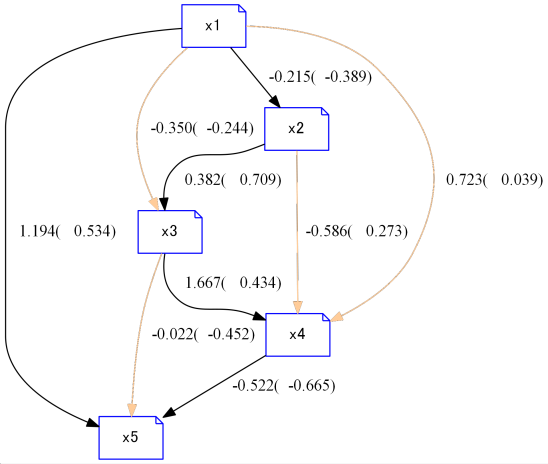


nonlinear\_LiNGAM\_latest3c.csv

nonlinear\_LiNGAM\_latest3b.csv

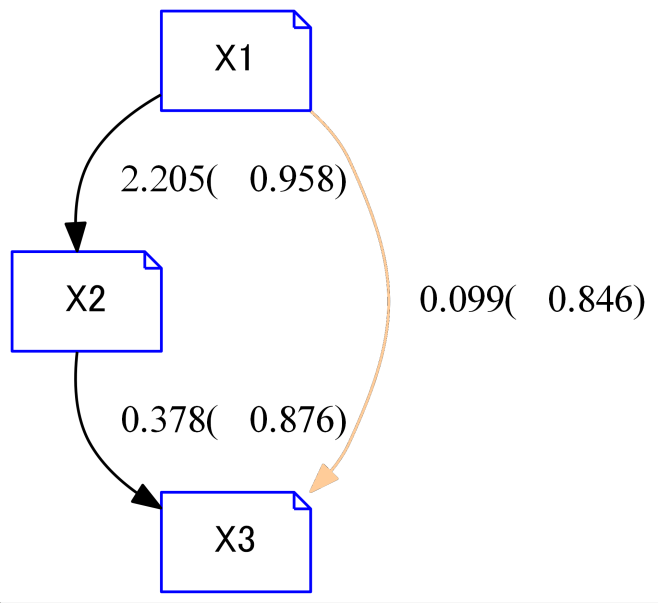






nonlinear\_LiNGAM\_latest3a.csv

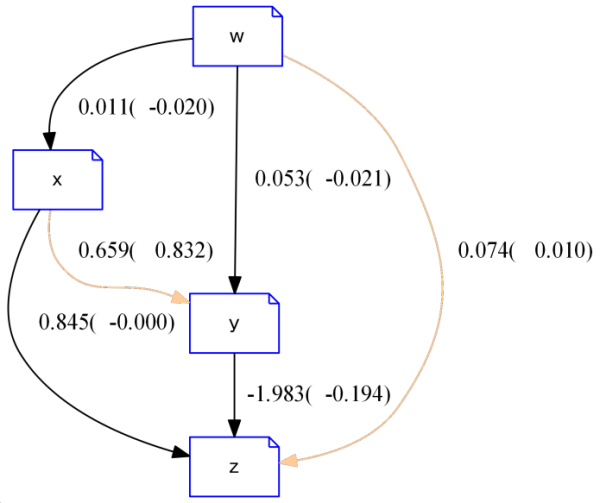




nonlinear.csv



nonlinear2.csv



**統計的因果**（データの分布からA→Bなのか、B→Aなのかをデータから見抜くこと）

LiNGAMモデル（Linear Non-Gaussian Acyclic Model： 線形非ガウシアン非巡回モデル）

n個の観測変数 **x**=があったとき、どの変数がどの変数にどれくらい影響を与えているか（因果関係）を推定する。

**ε**=は未観測変数で観測変数に影響を与えているであろう観測されなかった変数で例えばノイズなどの直接観測できない要素。

**仮定１**：**ε**=は互いに独立で非ガウス連続分布に従う。

注：独立性を仮定することは、各観測変数に対して、未観測共通原因がないことを意味する。

**仮定２**：線形性

以下、上記マトリクスBを非巡回有向グラフ行列と呼ぶことにする。

**仮定３**：各観測変数の因果関係に非巡回性を仮定します。これは因果グラフｓい考えた時、どの変数から出発して関係性をたどっても、元の変数に戻って来られないことを意味するため

Nonlinear causal discovery with additive noise models

はの親（原因）となる変数、は非線形関数が線形の場合は通常の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。。

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。

また、がと独立にならないペアがある場合は未観測の潜在共通変数が存在する可能性が考えられる。そこでと独立にならないの全ての変数に対してを指定した以下のモデルを定義する。

一般化ガウス分布

は未定のパラメータで以下の最適化によって定める。

は相互情報量で０であれば独立変数になる。つまり、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を求める。この最適化問題は多目的最適化となるため以下の損失関数を最小化する。

拡大Tchebyshev・スカラー化関数