

Package ‘mDAG’

March 13, 2019

Type Package

Title Inferring Causal Network from Mixed Observational Data Using a Directed Acyclic Graph

Version 1.2

Date 2019-3-13

Author Wujuan Zhong, Li Dong

Maintainer Wujuan Zhong <zhongwujuan@gmail.com>

Description Learning a mixed directed acyclic graph based on both continuous and categorical data.

LazyData true

Encoding UTF-8

License GPL (>= 2)

Imports Rcpp (>= 0.12.14), pcalg, mgm, bnlearn, methods, nnet

Depends R (>= 2.10), logistf

LinkingTo Rcpp, RcppArmadillo

RoxygenNote 6.1.1

NeedsCompilation yes

Repository CRAN

Date/Publication 2019-03-13 18:20:03 UTC

R topics documented:

example_data	2
mDAG	2
Index	5

 example_data

Example data

Description

An example data, including 5 variables (4 continuous variables and 1 binary variable) and 100 samples.

Usage

```
data(example_data)
```

mDAG

Inferring Causal Network from Mixed Observational Data Using a Directed Acyclic Graph

Description

This function learns a mixed directed acyclic graph based on both continuous and categorical data.

Usage

```
mDAG(data, type, level, lambdaGam = 0.25, ruleReg = "OR",
      threshold = "LW", weights = rep(1, nrow(data)), alpha = 0.05,
      nperm = 10000)
```

Arguments

data	A $n \times p$ matrix. Each row is a sample; each column is a variable.
type	A string vector of length p , indicating the type of variable for each column in data. 'g' for Gaussian, 'c' for categorical.
level	A vector of length p , indicating the number of categories of each variable. For continuous variables, set it to 1.
lambdaGam	Hyperparameter γ in the EBIC if <code>lambdaSel = 'EBIC'</code> . Defaults is <code>lambdaGam = 0.25</code> .
ruleReg	Default is 'OR'. Rule used to combine two estimates from nodewise regression (one from regressing A on B and the other from B on A). <code>ruleReg = 'AND'</code> requires both estimates to be nonzero in order to set the edge to be present. <code>ruleReg = 'OR'</code> requires at least one estimate to be nonzero in order to set the edge to be present.
threshold	Default is 'LW'. A threshold below which the combined estimates from node-wise regression are put to zero. <code>threshold = 'LW'</code> refers to the threshold in Loh and Wainwright (2012). <code>threshold = 'HW'</code> refers to the threshold in Haslbeck and Waldorp (2016). If <code>threshold = 'none'</code> no thresholding is applied.

weights	A vector of length n , indicating weights for observations.
alpha	Significance level for permutation test of conditional independence. Default is 0.05.
nperm	The number of permutations in the permutation test of conditional independence. Default is 10000.

Value

A list of the following components:

- arcs: A two-column matrix, indicating arcs of the DAG.
- nodes: A list. Each element is named after a node and contains the following elements.
 - nbr: a string vector indicating the neighbourhood of the node.
 - parents: a string vector indicating the parents of the node.
 - children: a string vector indicating the children of the node.
- skeleton: A $p \times p$ adjacency matrix. If there is an edge from node i to node j , its (i, j) th entry = 1; otherwise = 0.

Author(s)

Wujuan Zhong, Li Dong

References

- Jonas M. B. Haslbeck, Lourens J. Waldorp (2016). mgm: Structure Estimation for Time-Varying Mixed Graphical Models in high-dimensional Data arXiv preprint:1510.06871v2
- Markus Kalisch, Martin Maechler, Diego Colombo, Marloes H. Maathuis, Peter Buehlmann (2012). Causal Inference Using Graphical Models with the R Package pcalg. *Journal of Statistical Software*, 47(11), 1-26.
- Loh, P. L., & Wainwright, M. J. (2012, December). Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses. In *NIPS* (pp. 2096-2104).
- Haslbeck, J., & Waldorp, L. J. (2016). mgm: Structure Estimation for time-varying Mixed Graphical Models in high-dimensional Data. arXiv preprint arXiv:1510.06871.
- Marco Scutari (2010). Learning Bayesian Networks with the bnlearn R Package. *Journal of Statistical Software*, 35(3), 1-22.
- Venables, W. N. & Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
- Georg Heinze and Meinhard Ploner (2018). logistf: Firth's Bias-Reduced Logistic Regression. R package version 1.23.
- Min Jin Ha (2013). PenPC: A Two-step Approach to Estimate the Skeletons of High Dimensional Directed Acyclic Graphs. R package version 0.99.1.

Examples

```
# load package
library(mDAG)
type=c("g","g","g","g","c")
level=c(1,1,1,1,2)
# To save time for running example, we set nperm as 150.
# Use default nperm=10000 to generate a more reliable DAG for your own data.
dag=mDAG(data=example_data, type=type, level=level, nperm=150)
print(dag$skeleton)
# draw the DAG
# library(bnlearn)
# bnlearn:::graphviz.backend(nodes=names(dag$nodes),arcs=dag$arcs,shape="rectangle")
```

Index

`example_data`, [2](#)

`mDAG`, [2](#)