Package 'equSA'

January 20, 2018

Type Package

Title Estimate Directed and Undirected Graphical Models and Construct Networks

Version 1.1.5

Date 2018-01-16

Author Bochao Jia, Faming Liang, Runmin Shi, Suwa Xu

Maintainer Bochao Jia <jbc409@ufl.edu>

Depends R (>= 3.0.2)

Imports igraph, huge, XMRF, ZIM, mvtnorm, speedglm

Description Provides an equivalent measure of partial correlation coefficients for high-dimensional Gaussian Graphical Models to learn and visualize the underlying relationships between variables from single or multiple datasets. You can refer to Liang, F., Song, Q. and Qiu, P. (2015) <doi:10.1080/01621459.2015.1012391> for more detail. Based on this method, the package also provides the method for constructing networks for Next Generation Sequencing Data, for jointly estimating multiple Gaussian Graphical Models and constructing directed acyclic graph (Bayesian Network).

License GPL-2

LazyLoad true

NeedsCompilation yes

Repository CRAN

Date/Publication 2018-01-20 15:44:41 UTC

RoxygenNote 6.0.1

R topics documented:

`equSA-package` *Graphical model has been widely used in may scientific fileds to describe the conditional independent relationships for a large set of random variables. Through this package, we provide tools to learn both undirected graph (Markov Random Field) and directed acyclic graph (Bayesian Network). p*

Description

The package contains two parts, learning undirected graph and directed acyclic graph.

In the first part, the package provides an equivalent measure of partial correlation coefficients for high-dimensional Gaussian Graphical Models to learn and visualize the underlying relationships between variables from single or multiple datasets. The package also provides the method for constructing networks for Next Generation Sequencing Data. Besides, it includes the method for jointly estimating Gaussian Graphical Models of multiple datasets.

In the second part, the package implements the p-learning algorithm which is used to learn Bayesian networks for general types of random variables.

Details

Package:	equSA
Type:	Package
Version:	1.1.5
Date:	2018-01-16
License:	GPL-2

We propose an equvalent mearsure of partial correlation coeffient estimator called ψ estimators which enable us to estimate these networks via sparse, high-dimensional undirected graphical models. (Liang, F et al, 2015)

Here, we provide the community a convenient and useful tool to learn a Gaussian Graphical Models.

To estimate the network structures from Gaussian distributed data with this package, users simply need to specify the `"method"` in the main function, for example `equSAR(data,...)` to fit GGM to get the estimated adjacency matrix.

In this package, we also provide the code for combining Networks from two different dataset `combineR(data1,data2,...)` and the code for detecting difference between two Networks, for example `diffR(data1,data2,...)`. data1 and data2 should share the same dimension of variables (p) but allow have different samples (n).

This package also implement the Algorithm 17.1 of Friedman et al(2001), i.e estimate the covariance and precision matrix of the data given its structure. `solcov(data,struct,...)`

Besides estimating single GGM, we also propose a joint estimation method for multiple GGM. This is achieved by ψ- learning algorithm for graphical model at each time point combined with an Bayesian data integration method to estimate integrative ψ scores. Then multiple hypothesis tests were applied to identify the edges for each pair of variables. `JGGM(data,...)`.

If the data are not Normalized, for example, the count data, we propose a random effect model-based transformation to continuized data `ContTran(data,...)`, and then we transform the continuized data to Gaussian via a semiparametric transformation and then apply ψ- learning algorithm to reconstruct networks. The proposed method is consistent, and the resulting network satisfies the faithfulness and global Markov properties.The most common application is to estimate Gene Regulatory Networks from Next Generation Sequencing Data (Jia, B et al, 2017)

For learning Bayesian network, the package currently supports for Gaussian, binary and poisson data and also mixed type of data. `p_learning(data,...)`. The proposed algorithm provides a feasible way to describe conditional dependence relationships. A straightforward application of the Bayesian network is selection of causal features for high-dimensional generalized linear model.

Author(s)

Bochao Jia, Faming Liang, Runmin Shi, Suwa Xu Maintainer: Bochao Jia<jbc409@ufl.edu>

References

Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning (Vol. 1). Springer, Berlin: Springer series in statistics.

Liang, F., Song, Q. and Qiu, P. (2015). An Equivalent Measure of Partial Correlation Coefficients for High Dimensional Gaussian Graphical Models. J. Amer. Statist. Assoc., 110, 1248-1265.<doi:10.1080/01621459.2015.1012391>

Liang, F. and Zhang, J. (2008) Estimating FDR under general dependence using stochastic approximation. Biometrika, 95(4), 961-977.<doi:10.1093/biomet/asn036>

Liu, H., Lafferty, J. and Wasserman, L. (2009). The Nonparanormal: Semiparametric Estimation of High Dimensional Undirected Graphs. Journal of Machine Learning Research , 10, 2295-2328.

Jia, B., Xu, S., Xiao, G., Lamba, V., Liang, F. (2017) Inference of Genetic Networks from Next Generation Sequencing Data. Biometrics.

Jia, B. and Liang, F. (2017) Joint Estimation of Multiple Gaussian Graphical Models via Multiple Hypothesis Tests (preparing)

Jean-Philippe, Pellet and Andr\'e,Elisseeff (2008). Using Markov blankets for causal structure learning. Journal of Machine Learning Research, 9, 1295-1342.

Suwa, Xu and Faming, Liang (2017). Learning High-Dimensional Bayesian Networks for General Types of Random Variables. Submitted to Journal of Machine Learning.

Kalisch, Markus and B\"uhlmann, Peter (2007). Estimating high-dimensional directed acyclic graphs with the PC-algorithm. Journal of Machine Learning Research, 8, 613-636.

Examples

```
library(equSA)
data(TR0)
subset <- TR0
equSAR(subset)
```

`combineR` *Combine two networks.*

Description

Combine two networks to a single one from datasets of two groups by our calculated ψ scores.

Usage

```
combineR(Data1,Data2,ALPHA1=0.05,ALPHA2=0.05)
```

Arguments

`Data1`	a n_1xp data matrix.
`Data2`	a n_2xp data matrix.
`ALPHA1`	The significance level of correlation screening for each dataset. In general, a high significance level of correlation screening will lead to a slightly large separator set S_ij, which reduces the risk of missing some important variables in the conditioning set. Including a few false variables in the conditioning set will not hurt much the accuracy of the ψ-partial correlation coefficient.
`ALPHA2`	The significance level of ψ screening for integrative estimation of ψ scores.

Value

`A`	pxp Adjacency matrix of the combined graph.

Author(s)

Bochao Jia<jbc409@ufl.edu> and Faming Liang

References

Liang, F., Song, Q. and Qiu, P. (2015). An Equivalent Measure of Partial Correlation Coefficients for High Dimensional Gaussian Graphical Models. J. Amer. Statist. Assoc., 110, 1248-1265.

Liang, F. and Zhang, J. (2008) Estimating FDR under general dependence using stochastic approximation. Biometrika, 95(4), 961-977.

Examples

```
#library(equSA)
#data(SR0)
#data(TR0)
#combineR(SR0,TR0)
```

`Cont2Gaus` *A transfomation from count data into Gaussian data*

Description

To transform count data into Gaussian distributed and also keep the consistency for contructing networks.

Usage

```
Cont2Gaus(iData,total_iteration=5000,stepsize=0.05)
```

Arguments

`iData`	a *nxp* count data matrix.
`total_iteration`	Total iteration number for Baysian random effect model-based transformation, default of 5000.
`stepsize`	The stepsize of updating parameters in transformation, default of 0.05.

Details

This is the function that transform the count data into Gaussian data which include two steps. First, we do data continuized transformation `ContTran(data,...)` and then we apply the semiparametric transformation (Liu, H et al, 2009) provided in *"huge"* packages to tranform continuized data into Gaussian distributed.

Value

`Gaus`	*nxp* matrix of Normalized data with Gaussian distribution.

Author(s)

Bochao Jia<jbc409@ufl.edu> and Faming Liang

References

Jia, B., Xu, S., Xiao, G., Lamba, V., Liang, F. (2017) Inference of Genetic Networks from Next Generation Sequencing Data. Biometrics, in press.

Liu, H., Lafferty, J. and Wasserman, L. (2009). The Nonparanormal: Semiparametric Estimation of High Dimensional Undirected Graphs. Journal of Machine Learning Research , 10, 2295-2328.

Examples

```
library(equSA)
data(count)
Cont2Gaus(count,total_iteration=1000)
```

`ContSim` *A simulation method for generating count data from multivariate Zero-Inflated Negative Binomial distributions*

Description

Implements the data generation from multivariate Zero-Inflated Negative Binomial (ZINB) distributions with different graph structures, including `"random"`, `"hub"`, `"cluster"`, `"AR(2)"` and `"scale-free"`.

Usage

```
ContSim(n, p,  v = NULL, u = NULL, g = NULL,
prob = NULL, vis = FALSE, verbose = TRUE,
graph.type="AR(2)", k=3.30, lambda=515, omega=0.003,
lower.tail = TRUE, log.p = FALSE)
```

Arguments

`n`	The number of observations (sample size).
`p`	The number of variables (dimension).
`graph.type`	The graph structure with 4 options: `"random"`, `"hub"`, `"cluster"`, `"AR(2)"` and `"scale-free"`.
`v`	The off-diagonal elements of the precision matrix, controlling the magnitude of partial correlations with `u`. The default value is `0.3`.
`u`	A positive number being added to the diagonal elements of the precision matrix, to control the magnitude of partial correlations. The default value is `0.1`.

g	For "cluster" or "hub" graph, g is the number of hubs or clusters in the graph. The default value is about d/20 if d >= 40 and 2 if d < 40. NOT applicable to "random" and "AR(2)" graph.
prob	For "random" graph, it is the probability that a pair of nodes has an edge. The default value is 3/d. For "cluster" graph, it is the probability that a pair of nodes has an edge in each cluster. The default value is 6*g/d if d/g <= 30 and 0.3 if d/g > 30. NOT applicable to "hub" or "AR(2)" graphs.
vis	Visualize the adjacency matrix of the true graph structure, the graph pattern, the covariance matrix and the empirical covariance matrix. The default value is FALSE
verbose	If verbose = FALSE, tracing information printing is disabled. The default value is TRUE.
k	dispersion parameter of ZINB distribution, default of 3.30.
lambda	vector of (non-negative) means of ZINB distribution, default of 515.
omega	zero-inflation parameter of ZINB distribution, default of 0.003.
lower.tail	logical; if TRUE (default), probabilities are P[X <= x], otherwise, P[X> x].
log.p	logical; if TRUE, probabilities p are given as log(p).

Details

This is the function that can generate dataset from multivariate Zero-Inflated Negative Binomial distributions with different graph structures, including "random", "hub", "cluster", "AR(2)" and "scale-free".

Given the adjacency matrix theta, the graph patterns are generated as below:

(I) "random": Each pair of off-diagonal elements are randomly set theta[i,j]=theta[j,i]=1 for i!=j with probability prob, and 0 other wise. It results in about d*(d-1)*prob/2 edges in the graph.

(II)"hub":The row/columns are evenly partitioned into g disjoint groups. Each group is associated with a "center" row i in that group. Each pair of off-diagonal elements are set theta[i,j]=theta[j,i]=1 for i!=j if j also belongs to the same group as i and 0 otherwise. It results in d - g edges in the graph.

(III)"cluster":The row/columns are evenly partitioned into g disjoint groups. Each pair of off-diagonal elements are set theta[i,j]=theta[j,i]=1 for i!=j with the probability probif both i and j belong to the same group, and 0 other wise. It results in about g*(d/g)*(d/g-1)*prob/2 edges in the graph.

(IV)"AR(2)": The off-diagonal elements are set to be theta[i,j]=1 if 1<=|i-j|<=g and 0 other wise. It results in (2d-1-g)*g/2 edges in the graph.

(V) "scale-free": The graph is generated using B-A algorithm. The initial graph has two connected nodes and each new node is connected to only one node in the existing graph with the probability proportional to the degree of the each node in the existing graph. It results in d edges in the graph.

The adjacency matrix `theta` has all diagonal elements equal to `0`. To obtain a positive definite precision matrix, the smallest eigenvalue of `theta*v` (denoted by `e`) is computed. Then we set the precision matrix equal to `theta*v+(|e|+0.1+u)I`. The covariance matrix is then computed for generating multivariate ZINB dataset.

The default values for parameters `k`, `lambda` and `omega` of ZINB distribution are estimated from a real TCGA dataset. See Jia.B et al(2017) for more detail.

Value

A list of two elements:

`data`	The simulated count dataset in a *nxp* matrix.
`Adj`	*pxp* The adjacency matrix of true graph structure (in sparse matrix representation) for the generated data

Author(s)

Bochao Jia<jbc409@ufl.edu>

References

Jia, B., Xu, S., Xiao, G., Lamba, V., Liang, F. (2017) Inference of Genetic Networks from Next Generation Sequencing Data. Biometrics, in press.

T. Zhao and H. Liu.(2012) The huge Package for High-dimensional Undirected Graph Estimation in R. Journal of Machine Learning Research.

Yahav, I., and Shmueli, G. (2012). On generating multivariate Poisson data in management science applications. Applied Stochastic Models in Business and Industry, 28(1), 91-102.

Examples

```
library(equSA)
ContSim(100,200)
```

`ContTran` *A data continuized transformation*

Description

To transform count data into continuous data.

Usage

```
ContTran(iData,total_iteration=5000,stepsize=0.05)
```

Arguments

iData	a *nxp* count data matrix.
total_iteration	total iteration number for Baysian random effect model-based transformation, default of 5000.
stepsize	The stepsize of updating parameters in transformation, default of 0.05.

Details

This is the function that transform the count data into continuized data.

Value

continuz	*nxp* matrix of continuized data.

Author(s)

Bochao Jia<jbc409@ufl.edu>, Suwa Xu and Faming Liang

References

Jia, B., Xu, S., Xiao, G., Lamba, V., Liang, F. (2017) Inference of Genetic Networks from Next Generation Sequencing Data. Biometrics, in press.

Examples

```
library(equSA)
data(count)
ContTran(count,total_iteration=1000)
```

---

count *An example of count dataset for constructing networks*

Description

count is a simulated dataset for illustrating our proposed method for inferening networks from next generation sequencing data.

Usage

```
data(count)
```

Format

count dataset is a 100x200 matrix. Each row represents a observation and each column represents a variable. It is generated from an overdispersion and zero-inflated Possion distribution.

References

Jia, B., Xu, S., Xiao, G., Lamba, V., Liang, F. (2017) Inference of Genetic Networks from Next Generation Sequencing Data. Biometrics, in press.

`DAGsim` *Simulate a directed acyclic graph with mixed data (continuous and binary)*

Description

Simulate a directed acyclic graph with mixed data (continuous and binary).

Usage

```
DAGsim(n, p, sparsity = 0.02, p.binary)
```

Arguments

`n`	number of observations.
`p`	number of variables.
`sparsity`	sparsity of the graph.
`p.binary`	number of binary variables.

Details

The default value of sparsity is 0.02.

Value

A list of four objects.

`Adjacency.matrix`
: *pxp* The simulated adjacency matrix which indicates the true structure of directed acyclic graph. If the (i,j)th element is equal to 1, there exists a directed edge from X_i to eqnX_j.

`Data`
: The simulated dataset in a *nxp* matrix.

`gaussian.index`
: The index of continuous variables.

`binary.index`
: The index of binary variables.

Author(s)

Suwa Xu, Faming Liang

References

Kalisch, Markus and B\"uhlmann, Peter (2007). Estimating high-dimensional directed acyclic graphs with the PC-algorithm. Journal of Machine Learning Research, 8, 613-636.

Suwa, Xu and Faming, Liang (2017). Learning High-Dimensional Bayesian Networks for General Types of Random Variables. Submitted to Biometrika.

Examples

```
# library(equSA)
# set.seed(3)
# dagsim <- DAGsim(n = 3000, p = 100, sparsity = 0.02, p.binary = 50)
# data3000 <- dagsim$data
# cont_index  <- dagsim$gaussian.index
# binary_index<- dagsim$binary.index
# truegraph <- dagsim$Adjacency.matrix
```

`diffR` *Detect difference between two networks.*

Description

Detecting significant different edges between two networks using our calculated ψ scores.

Usage

```
diffR(Data1,Data2,ALPHA1=0.05,ALPHA2=0.05)
```

Arguments

`Data1`	a *n_1*x*p* data matrix.
`Data2`	a *n_2*x*p* data matrix.
`ALPHA1`	The significance level of correlation screening for each dataset. In general, a high significance level of correlation screening will lead to a slightly large separator set S_ij, which reduces the risk of missing some important variables in the conditioning set. Including a few false variables in the conditioning set will not hurt much the accuracy of the ψ-partial correlation coefficient.
`ALPHA2`	The significance level of ψ screening for integrative estimation of ψ scores.

Value

`A`	*p*x*p* adjacency matrix of the combined graph.

Author(s)

Bochao Jia<jbc409@ufl.edu> and Faming Liang

References

Liang, F., Song, Q. and Qiu, P. (2015). An Equivalent Measure of Partial Correlation Coefficients for High Dimensional Gaussian Graphical Models. J. Amer. Statist. Assoc., 110, 1248-1265.

Liang, F. and Zhang, J. (2008) Estimating FDR under general dependence using stochastic approximation. Biometrika, 95(4), 961-977.

Examples

```
#library(equSA)
#data(SR0)
#data(TR0)
#diffR(SR0,TR0,ALPHA2=0.2)
```

---

equSAR *An equvalent mearsure of partial correlation coeffients*

Description

Infer networks from Gaussian data using our proposed ψ -learning algorithm.

Usage

```
equSAR(iData,iMaxNei=as.integer(iDataNum/log(iDataNum)),
ALPHA1=0.05,ALPHA2=0.05,GRID=2,iteration=100)
```

Arguments

iData	a *nxp* data matrix.
iMaxNei	Neiborhood size in correlation screening step, default to n/log(n).
ALPHA1	The significance level of correlation screening. In general, a high significance level of correlation screening will lead to a slightly large separator set S_ij, which reduces the risk of missing some important variables in the conditioning set. Including a few false variables in the conditioning set will not hurt much the accuracy of the ψ-partial correlation coefficient.
ALPHA2	The significance level of ψ screening.
GRID	The number of components for the ψ-scores. The default value is 2.
iteration	Number of iterations for screening. The default value is 100.

Details

This is the main function of the package that fit the Gaussian Graphical Models and obtain the ψ scores and adjacency matrix.

Value

A list of two elements:

Adj	*pxp* adjacency matrix of the generated graph.
score	Estimated ψ score matrix which has 3 columns. The first two columns denote the pair indices of variables i and j and the last column denote the calculated ψ scores for this pair.

Author(s)

Bochao Jia and Faming Liang<faliang@ufl.edu>

References

Liang, F., Song, Q. and Qiu, P. (2015). An Equivalent Measure of Partial Correlation Coefficients for High Dimensional Gaussian Graphical Models. J. Amer. Statist. Assoc., 110, 1248-1265.

Liang, F. and Zhang, J. (2008) Estimating FDR under general dependence using stochastic approximation. Biometrika, 95(4), 961-977.

Examples

```
library(equSA)
data(SR0)
subset <- SR0
equSAR(subset)
```

JGGM *Joint estimation of Multiple Gaussian Graphical Models*

Description

Infer networks from Multiple Gaussian data from differnt groups using our proposed algorithm.

Usage

```
JGGM(data,ALPHA1=0.05,ALPHA2=0.01)
```

Arguments

data	a list of *n*x*p* data matrices. n can be different for each dataset but p should be the same.
ALPHA1	The significance level of correlation screening. In general, a high significance level of correlation screening will lead to a slightly large separator set S_ij, which reduces the risk of missing some important variables in the conditioning set. Including a few false variables in the conditioning set will not hurt much the accuracy of the ψ-partial correlation coefficient.
ALPHA2	The significance level of ψ screening.

Details

This is the function that can jointly estimate multiple GGMs which can integrate the information throughtout all datasets. The method mainly consists three steps: (i) separate estimation of ψ-scores for each dataset, (ii) identifies possible changes of each edge across different groups and integrate the ψ scores across different groups simultaneously and (iii) apply multiple hypothesis test to identify edges using integrated ψ scores.

Value

A list of two elements:

A	An array of multiple adjacency matrices of networks which is a *M*x*p*x*p* array. *M* is the number of dataset groups, *p* is the dimension of variables in each group.
score	Estimated integrative ψ scores matrix for all pairs of different datasets. The first two columns denote the pair indices of variables i and j and the rest columns denote the Estimated integrative ψ scores for this pair in different groups.

Author(s)

Bochao Jia<jbc409@ufl.edu> and Faming Liang

References

Liang, F., Song, Q. and Qiu, P. (2015). An Equivalent Measure of Partial Correlation Coefficients for High Dimensional Gaussian Graphical Models. J. Amer. Statist. Assoc., 110, 1248-1265.

Liang, F. and Zhang, J. (2008) Estimating FDR under general dependence using stochastic approximation. Biometrika, 95(4), 961-977.

Jia, B. and Liang, F. (2017) Joint Estimation of Multiple Gaussian Graphical Models via Multiple Hypothesis Tests (preparing)

Examples

```
#library(equSA)
#data(SR0)
#data(TR0)
#data_all <- vector("list",2)
#data_all[[1]] <- SR0
```

```
#data_all[[2]] <- TR0
#JGGM(data_all,ALPHA1=0.05,ALPHA2=0.01)
```

---

mixed3000 *One example dataset for p_learning*

Description

mixed3000 is a simulated dataset for illustration our p_learning alogorithm.

Usage

```
data(mixed3000)
```

Format

mixed3000 dataset is a 3000x100 matrix. Each row represents a observation and each column represents a variable. It contains 50 continuous variables and 50 binary variables.

References

Suwa, Xu and Faming, Liang (2017). Learning High-Dimensional Bayesian Networks for General Types of Random Variables. Submitted to Biometrika.

---

pcorselR *Multiple hypothesis test*

Description

Infer networks from ψ scores using multiple hypothesis test in ψ screening procedure.

Usage

```
pcorselR(score, ALPHA2=0.05,GRID=2,iteration=100)
```

Arguments

score	ψ score matrix which has 3 columns. The first two columns denote the pair of variables i and j and the last column denote the calculated ψ scores for this pair.
ALPHA2	The significance level of ψ screening, default of 0.05.
GRID	The number of components for the ψ-scores. The default value is 2.
iteration	Number of iterations for screening. The default value is 100.

Details

This is the function that conduct multiple hypothesis test for ψ scores, thus we called it ψ screening procedure.

Value

qqqscore — The threshold value of ψ scores which indicates that if one pair of variables has larger ψ scores than this threshold value in the ψ score matrix, this pair is considered as connected, i.e there is an edge between this pair of variables.

Author(s)

Bochao Jia, Faming liang<faliang@ufl.edu>

References

Liang, F., Song, Q. and Qiu, P. (2015). An Equivalent Measure of Partial Correlation Coefficients for High Dimensional Gaussian Graphical Models. J. Amer. Statist. Assoc., 110, 1248-1265.

Liang, F. and Zhang, J. (2008) Estimating FDR under general dependence using stochastic approximation. Biometrika, 95(4), 961-977.

Examples

```
library(equSA)
data(SR0)
U <- psical(SR0, ALPHA1=0.05,iteration=50)
##   probit transformation for psi scores ###
z<-U[,3]
q<-pnorm(-abs(z), log.p=TRUE)
q<-q+log(2.0)
s<-qnorm(q,log.p=TRUE)
s<-(-1)*s
U<-cbind(U[,1:2],s)
## subsampling for psi scores ###
N <- length(U[,1])
ratio<-ceiling(N/100000)
U<-U[order(U[,3]), 1:3]
m<-floor(N/ratio)
m0<-N-m*ratio
s<-sample.int(ratio,m,replace=TRUE)
for(i in 1:length(s)) s[i]<-s[i]+(i-1)*ratio
if(m0>0){
  s0<-sample.int(m0,1)+length(s)*ratio
  s<-c(s,s0)
}
Us<-U[s,]
y <- round(Us,6)
##  multiple hypothesis tests ###
pcorselR(y,ALPHA2=0.05)
```

plotGraph *Plot Single Network*

Description

Plot a network with specific layout.

Usage

```
plotGraph(net, fn = "", th = 1e-06, mylayout = NULL)
```

Arguments

net	a square adjacency matrix of the network to be plotted.
fn	file name to save the network plot. Default to be an empty string, so the network is plotted to the standard output (screen). NOTE: if a file name is specified, it should be file name for PDF file.
th	numeric value, default to 1e-06. To specify the threshold if the estimated coefficient between two variables is to be considered connected.
mylayout	graph layout to draw the network, default to NULL.

Details

This function serves as the alternative plotting function to allow users to plot a specific network with specific layout, such as plotting the simulated network.

Value

Returns the layout object from igraph package - numeric matrix of two columns and the rows with the same number as the number of vertices.

Examples

```
library(equSA)
data(SR0_mat)
plotGraph(as.matrix(SR0_mat))
```

plotJGraph *Plot Networks*

Description

Plot multiple networks with specific layout.

Usage

```
plotJGraph(A,fn="Net",th = 1e-06, mylayout = NULL)
```

Arguments

A	An array of multiple adjacency matrices of networks to be plotted which is a *M*x*p*x*p* array. *M* is the number of dataset groups, *p* is the dimension of variables in each group.
fn	file name to save the network plots. Default to be an string called "Net". NOTE: It should be file name for PDF file.
th	numeric value, default to 1e-06. To specify the threshold if the estimated coefficient between two variables is to be considered connected.
mylayout	graph layout to draw networks, default to NULL.

Details

This function serves as the alternative plotting function to allow users to plot multiple networks with specific layout, such as plotting the simulated networks.

Value

Returns the multiple layout objects from igraph package - numeric matrix of two columns and the rows with the same number as the number of vertices.

Author(s)

Bochao Jia, Faming liang<faliang@ufl.edu>

References

Allen, G.I., and Liu, Z. (2012). A Log-Linear graphical model for inferring genetic networks from high-throughput sequencing data. *The IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2012)*.

Jia, B. and Liang, F. (2017) Joint Estimation of Multiple Gaussian Graphical Models via Multiple Hypothesis Tests (preparing)

Examples

```
#library(equSA)
#data(SR0)
#data(TR0)
#data_all <- vector("list",2)
#data_all[[1]] <- SR0
#data_all[[2]] <- TR0
#A <- JGGM(data_all,ALPHA1=0.05,ALPHA2=0.01)$Array
#plotJGraph(A)
```

---

psical *An calculation of ψ scores.*

Description

To compute an equvalent mearsure of partial correlation coeffients called ψ scores.

Usage

```
psical(iData,iMaxNei=as.integer(iDataNum/log(iDataNum)),
ALPHA1=0.05,GRID=2,iteration=100)
```

Arguments

iData	a *nxp* data matrix.
iMaxNei	Neiborhood size in correlation screening step, default to n/log(n).
ALPHA1	The significance level of correlation screening. In general, a high significance level of correlation screening will lead to a slightly large separator set S_{ij}, which reduces the risk of missing some important variables in the conditioning set. Including a few false variables in the conditioning set will not hurt much the accuracy of the ψ-partial correlation coefficient.
GRID	The number of components for the corrlation scores. The default value is 2.
iteration	Number of iterations for screening. The default value is 100.

Details

This is the function to calculate ψ scores and can be used in combining or detecting difference of two networks.

Value

score	Estimated ψ score matrix which has 3 columns. The first two columns denote the pair indices of variables i and j and the last column denote the calculated ψ scores for this pair.

Author(s)

Bochao Jia, Faming liang<faliang@ufl.edu>

References

Liang, F., Song, Q. and Qiu, P. (2015). An Equivalent Measure of Partial Correlation Coefficients for High Dimensional Gaussian Graphical Models. J. Amer. Statist. Assoc., 110, 1248-1265.

Liang, F. and Zhang, J. (2008) Estimating FDR under general dependence using stochastic approximation. Biometrika, 95(4), 961-977.

Examples

```
library(equSA)
data(SR0)
subset <- SR0
psical(subset)
```

---

p_learning *Construct Bayesian Network based on p-learning algorithm.*

Description

Construct Bayeisan network for general types of random varaibles based on *p*-learning algorithm.

Usage

```
p_learning(data, gaussian.index, binary.index, poisson.index,
alpha1 = 0.1, alpha2 = 0.02, alpha3 = 0.02)
```

Arguments

data	The data matrix, of dimensions *n*x*p*. Each row is an observation vector.
gaussian.index	The index vector of continuous nodes. The default value is NULL.
binary.index	The index vector of binary nodes. The default value is NULL.
poisson.index	The index vector of poisson nodes. The default value is NULL.
alpha1	The significant level of step(a) of *p*-screening method. The default value is 0.1.
alpha2	The significant level of step(c) of *p*-screening method. The dafault value is 0.02.
alpha3	The significant level of solving Markov Blankets. The default value is 0.02.

Details

This is the function that implements the p-learning algorithm.

Value

A list of one object.

PDAG The derived partial directed acyclic graph.

Author(s)

Suwa Xu and Faming Liang

References

Suwa, Xu and Faming, Liang (2017). Learning High-Dimensional Bayesian Networks for General Types of Random Variables. Submitted to Biometrika.

Examples

```
#library(equSA)
#data(mixed3000)
#pdag3000 <- p_learning(data =mixed3000$data, gaussian.index =
#mixed3000$gaussian.index,binary.index <- mixed3000$binary.index)$PDAG
```

---

simtoequiv *Transform a directed acyclic graph into an equivalent correct graph.*

Description

A correct graph is specified by its adjacencies and V-structures only.

Usage

```
simtoequiv(edgematrix)
```

Arguments

edgematrix The simulated true graph.

Value

A list of one object.

PDAG The equivalent correct graph.

Author(s)

Suwa Xu, Faming Liang

References

Jean-Philippe, Pellet and Andr\'e,Elisseeff (2008). Using Markov blankets for causal structure learning. Journal of Machine Learning Research, 9, 1295-1342.

Suwa, Xu and Faming, Liang (2017). Learning High-Dimensional Bayesian Networks for General Types of Random Variables. Submitted to Biometrika.

Examples

```
# library(equSA)
# load("mixed3000.rda")
# equiv_graph <- simtoequiv(mixed3000$Adjacency.matrix)$PDAG
```

`solcov` *Calculate covariance matrix and precision matrix*

Description

Calculate the adjusted covriance matrix and precision matrix given the network structure from high dimesional dataset.

Usage

```
solcov(data, struct, tol=10^-5)
```

Arguments

`data`	A *nxp* data matrix.
`struct`	A preacquired adjacency matrix
`tol`	Tolerant value, default is 10^-5

Value

A list of two elements:

`COV`	Adjusted covriance matrix
`PRE`	Precision matrix

Author(s)

Bochao Jia& Runmin Shi <jbc409@ufl.edu>

References

Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning (Vol. 1). Springer, Berlin: Springer series in statistics.

Examples

```
library(equSA)
data(SR0)
data(SR0_mat)
subSR0 <- SR0[1:10,1:10]
subSR0_mat <- SR0_mat[1:10,1:10]
solcov(subSR0, subSR0_mat)


          # library(equSA)
          # data(SR0)
          #  data(SR0_mat)
          # solcov(SR0, SR0_mat)
```

SR0 *One example dataset for equSA*

Description

SR0 is a simulated dataset for illustration our equSA alogorithm.

Usage

```
data(SR0)
```

Format

SR0 dataset is a 100x200 matrix. Each row represents a observation and each column represents a variable.

References

Liang, F., Song, Q. and Qiu, P. (2015). An Equivalent Measure of Partial Correlation Coefficients for High Dimensional Gaussian Graphical Models. J. Amer. Statist. Assoc., 110, 1248-1265.

SR0_mat *The adjacency matrix for SR0 dataset.*

Description

SR0_mat is an estimated adjacency matrix by ψ- learning alogorithm.

Usage

```
data(SR0_mat)
```

Format

SR0_mat a 200x200 matrix with binary values. When its element (i,j) equals to 1, there exists an edge between variable i and j. Otherwise, it equals to 0.

References

Liang, F., Song, Q. and Qiu, P. (2015). An Equivalent Measure of Partial Correlation Coefficients for High Dimensional Gaussian Graphical Models. J. Amer. Statist. Assoc., 110, 1248-1265.

`TR0` *One example dataset for equSA*

Description

TR0 is a simulated dataset for illustration our equSA alogorithm.

Usage

```
data(TR0)
```

Format

TR0 dataset is a 100x200 matrix. Each row represents a observation and each column represents a variable.

References

Liang, F., Song, Q. and Qiu, P. (2015). An Equivalent Measure of Partial Correlation Coefficients for High Dimensional Gaussian Graphical Models. J. Amer. Statist. Assoc., 110, 1248-1265.

`TR0_mat` *The adjacency matrix for TR0 dataset.*

Description

TR0_mat is an estimated adjacency matrix by ψ- learning alogorithm.

Usage

```
data(TR0_mat)
```

Format

TR0_mat a 200x200 matrix with binary values. When its element (i,j) equals to 1, there exists an edge between variable i and j. Otherwise, it equals to 0.

References

Liang, F., Song, Q. and Qiu, P. (2015). An Equivalent Measure of Partial Correlation Coefficients for High Dimensional Gaussian Graphical Models. J. Amer. Statist. Assoc., 110, 1248-1265.

Index