

Package ‘BigSEM’

September 29, 2016

Type Package

Title Constructing Large Systems of Structural Equations

Version 0.2

Date 2016-09-29

Author Chen Chen, Dabao Zhang

Maintainer Dabao Zhang <zhangdb@stat.purdue.edu>

Description Construct large systems of structural equations using the two-stage penalized least squares (2SPLS) method proposed by Chen, Zhang and Zhang (2016).

Imports MASS, parcor, stats

Depends R(>= 2.10.0)

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2016-09-29 23:46:31

R topics documented:

BigSEM-package	2
bootsem	2
fitsem	3
getedges	5
simdata	6
yeast	6
Index	8

BigSEM-package

*Constructing Large Systems of Structural Equations***Description**

This package implements the two-stage penalized least squares (2SPLS) method proposed by Chen, Zhang and Zhang (2016), which can construct a large system of structural equations, with the k -th structural equation as $Y_k = Y_{-k}\gamma_k + X_{S_k}\psi_{S_k} + \epsilon_k$.

Author(s)

Chen Chen <chen1167@stat.purdue.edu>, Dabao Zhang <zhangdb@stat.purdue.edu>.

References

Chen, C., Zhang, M., and Zhang, D. (2016) *A Two-Stage Penalized Least Squares Method for Constructing Large Systems of Structural Equations*. (Submitted)

bootsem

*Bootstrap Analysis of the Fitted System of Structural Equations***Description**

Use the bootstrap data sets to evaluate the significance of regulatory effects in the fitted large system of structural equations by the two-stage penalized least squares (2SPLS) method proposed by Chen, Zhang and Zhang (2016).

Usage

```
bootsem(y, x, sk, nboots=100)
```

Arguments

y	a data frame containing the endogenous variables Y_1, Y_2, \dots, Y_p in the model.
x	a data frame containing the exogenous variables X_1, X_2, \dots, X_q in the model.
sk	a list with the k -th element specifying S_k which includes the indices of exogenous variables appearing in the structural equation for k -th endogenous variable.
nboots	number of the bootstrap datasets.

Details

Generate the bootstrap data sets by randomly sampling from the original data with replacement, and apply 2SPLS to each bootstrap data set to infer the regulatory effects.

Value

boot.target	the index of target variable.
boot.source	the index of source variable.
boot.freq	the bootstrap frequency of the regulatory effect.
boot.mean	the mean of the regulatory effect.
boot.sd	the standard deviation of the regulatory effects.

Author(s)

Chen Chen <chen1167@stat.purdue.edu>, Dabao Zhang <zhangdb@stat.purdue.edu>.

References

Chen, C., Zhang, M., and Zhang, D. (2016) *A Two-Stage Penalized Least Squares Method for Constructing Large Systems of Structural Equations*. (Submitted)

See Also

[fitsem](#) for constructing large systems of structural equations.

Examples

```
data(simdata)
attach(simdata)
#simsem <- fitsem(y=y,x=x,sk=sk)
#btres <- bootsem(y,x,sk,nboots=200)
#summary(btres)
```

fitsem

Constructing Large Systems of Structural Equations

Description

Use the two-stage penalized least squares (2SPLS) method proposed by Chen, Zhang and Zhang (2016) to construct a large system of structural equations, with the k -th structural equation as $Y_k = Y_{-k}\gamma_k + X_{S_k}\psi_{S_k} + \epsilon_k$.

Usage

```
fitsem(y,x,sk,method='2SPLS')
```

Arguments

y	a data frame containing the endogenous variables Y_1, Y_2, \dots, Y_p in the model.
x	a data frame containing the exogenous variables X_1, X_2, \dots, X_q in the model.
sk	a list with the k -th element specifying S_k which includes the indices of exogenous variables appearing in the structural equation for k -th endogenous variable.
method	the method to be used; the default is "2SPLS".

Details

2SPLS fits a single regression model for each endogenous variable at each stage. It employs the L₂ penalty at the first stage to obtain consistent estimation of a set of well-defined surrogate variables, and the L₁ penalty at the second stage to consistently select regulatory effects among massive candidates.

Value

y.names	the names of the endogenous variables.
x.names	the names of the exogenous variables.
y.source	a list with the k -th element indicating all endogenous variables which affect the k -th endogenous variable, i.e., all endogenous variables having nonzero coefficients on the right-hand of the k -th structural equation.
y.gamma	a list with k -th element the nonzero coefficients of γ_k .
y.sk	the same as the input sk.

Author(s)

Chen Chen <chen1167@stat.purdue.edu>, Dabao Zhang <zhangdb@stat.purdue.edu>.

References

Chen, C., Zhang, M., and Zhang, D. (2016) *A Two-Stage Penalized Least Squares Method for Constructing Large Systems of Structural Equations*. (Submitted)

See Also

[bootsem](#) for bootstrap analysis of the fitted system of structural equations.

Examples

```
data(simdata)
attach(simdata)
#simsem <- fitsem(y=y,x=x,sk=sk)
#summary(simsem)

# Check the exogenous variables affecting Y_1
#simsem$x.names[simsem$y.sk[[1]]]
#cat(simsem$y.names[1], '<- ', simsem$x.names[sk[[1]]], '\n')

# Check the endogenous variables affecting Y_1
#simsem$y.names[simsem$y.source[[1]]]
#cat(simsem$y.names[1], '<- ', simsem$y.names[simsem$y.source[[1]]], '\n')

# Check gamma_1
#simsem$y.gamma[[1]]
```

`getedges`*Retrieve All the Edges from the Fitted Systems of Structural Equations*

Description

The output from `fitsem`, this function lists all the connects in the fitted system by the two-stage penalized least squares (2SPLS) method (Chen, Zhang and Zhang, 2016).

Usage

```
getedges(fsObj)
```

Arguments

`fsObj` a result of a call to `fitsem`.

Details

The output will list each connection in the fitted system in a row.

Value

<code>target</code>	the indices of the target variables.
<code>source</code>	the indices of the source variables.
<code>weight</code>	the corresponding value in γ_k when the connect is included in the k -th structural equation.

Author(s)

Chen Chen <chen1167@stat.purdue.edu>, Dabao Zhang <zhangdb@stat.purdue.edu>.

References

Chen, C., Zhang, M., and Zhang, D. (2016) *A Two-Stage Penalized Least Squares Method for Constructing Large Systems of Structural Equations*. (Submitted)

See Also

`bootsem` for bootstrap analysis of the fitted system of structural equations.

Examples

```
data(simdata)
attach(simdata)
#simsem <- fitsem(y=y,x=x,sk=sk)

#edges <- getedges(simsem)
```

```
# Check the first connect
#cat(simsem$y.names[edges$target[1]], '<- ', simsem$y.names[edges$source[1]], '\n')
```

simdata

Simulated Data for a System of Structural Equations

Description

There are 100 observations sampled from a system of structural equations, which includes 10 endogenous variables and 10 exogenous variables. Each endogenous variables are directly affected by three unique exogenous variables.

Usage

```
data("simdata")
```

Format

A list with two data frames and one list.

y values of 10 endogenous variables.

x values of 10 exogenous variables.

sk a list of indices indicating the unique exogenous variables for each endogenous variable.

Examples

```
data(simdata)
summary(simdata)
attach(simdata)
#simsem <- fitsem(y=y,x=x,sk=sk)
```

yeast

Yeast Genetical Genomics Data

Description

A set of data from a genetical genomics experiment on yeast, including 112 observations on 722 gene expression levels and 732 genotypic values.

Usage

```
data("yeast")
```

Format

A list with two data frames and one list.

geneexp expression levels of 722 genes.

snps genotypic values of 732 SNPs.

ciseqtl a list of the indices of cis-eQTL for each gene.

Details

The experiment included 112 segregants from a cross between two strains BY4716 and RM11-1a (Brem and Kruglyak, 2005). The data here only keep 722 genes which have marginally significant cis-eQTL (p-value < 0.05). The set of cis-eQTL for each gene was filtered to control a pairwise correlation under 0.90, and then further filtered to keep up to three cis-eQTL which have the strongest association with the corresponding gene expression.

Source

Brem, R. B. and Kruglyak, L. (2005) *The landscape of genetic complexity across 5,700 gene expression traits in yeast*. Proceedings of the National Academy of Sciences of the United States of America, 102(5):1572-1577.

References

Chen, C., Zhang, M., and Zhang, D. (2016) *A Two-Stage Penalized Least Squares Method for Constructing Large Systems of Structural Equations*. (Submitted)

Examples

```
data(yeast)
summary(yeast)
attach(yeast)

#yeastsem <- fitsem(y=geneexp,x=snps,sk=ciseqtl)
```

Index

*Topic **datasets**

simdata, [6](#)

yeast, [6](#)

BigSEM (BigSEM-package), [2](#)

BigSEM-package, [2](#)

bootsem, [2](#), [4](#), [5](#)

fitsem, [3](#), [3](#), [5](#)

getedges, [5](#)

simdata, [6](#)

yeast, [6](#)