Package ‘nda’

Type Package
Title Generalized Network-Based Dimensionality Reduction and Analysis
Version 0.1.13
Maintainer Zsolt T. Kosztyan <kosztyan.zsolt@gtk.uni-pannon.hu>
License GPL (>= 2)
Encoding UTF-8
LazyData true
URL https://github.com/kzst/nda
Depends R (>= 4.00)
Imports energy, psych, stats, igraph, Matrix, MASS, ppcor, visNetwork
RoxygenNote 7.2.3
NeedsCompilation no
Author Zsolt T. Kosztyan [aut, cre], Marcell T. Kurbucz [aut], Attila I. Katona [aut]
Repository CRAN
Date/Publication 2023-09-27 07:20:06 UTC

R topics documented:

nda-package ......................................................... 2
biplot.nda .......................................................... 3
COVID19_2020 ......................................................... 4
CrimesUSA1990.X .................................................... 4
CrimesUSA1990.Y .................................................... 5
CWTS_2020 ............................................................ 5
data_gen .............................................................. 6
**nda-package**

**Package of Generalized Network-based Dimensionality Reduction and Analyses**

**Description**

The package of Generalized Network-based Dimensionality Reduction and Analysis (GNDA).

**Author(s)**

Zsolt T. Kosztyan*, Marcell T. Kurbucz, Attila I. Katona

e-mail*: kosztyan.zsolt@gtk.uni-pannon.hu

**References**


**See Also**

ndr, plot, biplot, summary, dCor.
Biplot function for Generalized Network-based Dimensionality Reduction and Analysis (GNDA)

## S3 method for class 'nda'

biplot(x, main=NULL, ...)

### Arguments

- **x**: an object of class 'NDA'.
- **main**: main title of biplot.
- **...**: other graphical parameters.

### Author(s)

Zsolt T. Kosztyan*, Marcell T. Kurbucz, Attila I. Katona

e-mail*: kosztyan.zsolt@gtk.uni-pannon.hu

### See Also

plot, summary, ndr, data_gen.

### Examples

# Biplot function without feature selection

# Generate 200 x 50 random block matrix with 3 blocks and lambda=0 parameter

df<-data_gen(200,50,3,0)
p<-ndr(df)
biplot(p)
**COVID19_2020**  
*Covid'19 case datesets of countries (2020), where the data frame has 138 observations of 18 variables.*

**Description**  
Sample datasets for Generalized Network-based Dimensionality Reduction and Analysis (GNDA) Covid'19 of countries (2020), where the data frame has 138 observations of 18 variables.

**Usage**  
```r  
data("COVID19_2020")  
```

**Format**  
A data frame with 138 observations 18 variables.

**Source**  

**Examples**  
```r  
data(COVID19_2020)  
```

---

**CrimesUSA1990.X**  
*Crimes in USA cities in 1990. Independent variables (X)*

**Description**  
Sample datasets for Generalized Network-based Dimensionality Reduction and Analysis (GNDA) Crimes in USA cities in 1990. Independent variables (X)

**Usage**  
```r  
data("CrimesUSA1990.X")  
```

**Format**  
A data frame with 1994 observations 123 variables.

**Source**  
Examples

data(CrimesUSA1990.X)

---

**CrimesUSA1990.Y**  
*Crimes in USA cities in 1990. Dependent variable (Y)*

**Description**

Sample datasets for Generalized Network-based Dimensionality Reduction and Analysis (GNDA)  
Crimes in USA cities in 1990. Dependent variable (Y)

**Usage**

data("CrimesUSA1990.Y")

**Format**

A data frame with 1994 observations 1 variables.

**Source**


**Examples**

data(CrimesUSA1990.Y)

---

**CWTS_2020**  
*CWTS Leiden’s University Ranking 2020 for all scientific fields, within the period of 2016-2019. 1176 observations (i.e., universities), and 42 variables (i.e., indicators).*

**Description**

Sample datasets for Generalized Network-based Dimensionality Reduction and Analysis (GNDA)  
CWTS Leiden’s 2020 dataset, where the data frame has 1176 observations of 42 variables.

**Usage**

data("CWTS_2020")

**Format**

A data frame with 1176 observations of 42 variables.
data_gen

Source

CWTS Leiden Ranking 2020: https://www.leidenranking.com/ranking/2020/list

Examples

data(CWTS_2020)

Generate random block matrix for GNDA

Description

Generate random block matrix for Generalized Network-based Dimensionality Reduction and Analysis (GNDA)

Usage

data_gen(n,m,nfactors=2,lambda=1)

Arguments

n number of rows
m number of columns
nfactors number of blocks (factors, where the default value is 2)
lambda exponential smoothing, where the default value is 1

Details

n, m, nfactors must be integers, and they are not less than 1; lambda should be a positive real number.

Value

M a dataframe of a block matrix

Author(s)

Prof. Zsolt T. Kosztyan, Department of Quantitative Methods, Institute of Management, Faculty of Business and Economics, University of Pannonia, Hungary
e-mail: kzst@gtk.uni-pannon.hu
Examples

# Specification 30 by 10 random block matrices with 2 blocks/factors
df<-data_gen(30,10)
library(psych)
scree(df)
biplot(ndr(df))
# Specification 40 by 20 random block matrices with 3 blocks/factors
df<-data_gen(40,20,3)
library(psych)
scree(df)
biplot(ndr(df))
plot(ndr(df))
# Specification 50 by 20 random block matrices with 4 blocks/factors
# lambda=0.1
df<-data_gen(50,15,4,0.1)
scree(df)
biplot(ndr(df))
plot(ndr(df))

dCor

Calculating distance correlation of two vectors or columns of a matrix

Description

Calculating distance correlation of two vectors or columns of a matrix for Generalized Network-based Dimensionality Reduction and Analysis (GNDA).

The calculation is very slow for large matrices!

Usage

dCor(x, y=NULL)

Arguments

x

a numeric vector, a numeric matrix (in this case y=NULL), or a numeric data frame (in this case y=NULL)

y

a numeric vector (optional)

Details

If x is a numeric vector, y must be specified. If x is a numeric matrix or numeric data frame, y must be ignored from the parameters.

Value

Either a distance correlation value of vectors x and y, or a distance correlation matrix of x.


**Author(s)**

Prof. Zsolt T. Kosztyan, Department of Quantitative Methods, Institute of Management, Faculty of Business and Economics, University of Pannonia, Hungary
e-mail: kosztyan.zsolt@gtk.uni-pannon.hu

**References**


**Examples**

```r
# Specification of distance correlation value of vectors x and y.
x<-rnorm(36)
y<-rnorm(36)
dCor(x,y)
# Specification of distance correlation matrix.
x<-matrix(rnorm(36),nrow=6)
dCor(x)
```

---

**dCov**

*Calculating distance covariance of two vectors or columns of a matrix*

**Description**

Calculating distance covariance of two vectors or columns of a matrix for Generalized Network-based Dimensionality Reduction and Analysis (GNDA).

The calculation is very slow for large matrices!

**Usage**

dCov(x,y=NULL)

**Arguments**

- `x`  
  a numeric vector, a numeric matrix (in this case `y=NULL`), or a numeric data frame (in this case `y=NULL`)
- `y`  
  a numeric vector (optional)

**Details**

If `x` is a numeric vector, `y` must be specified. If `x` is a numeric matrix or numeric data frame, `y` must be ignored from the parameters.

**Value**

Either a distance covariance value of vectors `x` and `y`, or a distance covariance matrix of `x`. 
Author(s)

Prof. Zsolt T. Kosztyan, Department of Quantitative Methods, Institute of Management, Faculty of Business and Economics, University of Pannonia, Hungary

e-mail: kosztyan.zsolt@gtk.uni-pannon.hu

References


Examples

```r
# Specification of distance covariance value of vectors x and y.
x<-rnorm(36)
y<-rnorm(36)
dCov(x,y)
# Specification of distance covariance matrix.
x<-matrix(rnorm(36),nrow=6)
dCov(x)
```

---

**fs.dimred**

Feature selection for PCA, FA, and (G)NDA

Description

This function drops variables that have low communality values and/or are common indicators (i.e., correlates more than one latent variables).

Usage

```r
fs.dimred(fn,DF,min_comm=0.25,com_comm=0.25)
```

Arguments

- **fn**: It is a list variable of the output of a principal (PCA), a fa (FA), or an ndr (NDA) function.
- **DF**: Numeric data frame, or a numeric matrix of the data table
- **min_comm**: Scalar between 0 to 1. Minimal communality value, which a variable has to be achieved. The default value is 0.25.
- **com_comm**: Scalar between 0 to 1. The minimal difference value between loadings. The default value is 0.25.
Details
This function only works with principal, and fa, and ndr functions.
This function drops each variable that has a low communality value (under min_comm value). In other words, that variable does not fit enough of any latent variable.
This function also drops so-called common indicators, which correlate highly with more than one latent variable. And the difference in the correlation is either lower than the com_comm value or the greatest absolute factor loading value is not twice greater than the second greatest factor loading.

Value

- **dropped_low**: Numeric data frame or numeric matrix. Set of indicators (i.e. variables), which are dropped by their low communalities. This value is NULL if a correlation matrix is used as an input or there is no dropped indicator.
- **dropped_com**: Numeric data frame or numeric matrix. Set of dropped common indicators (i.e. common variables). This value is NULL if a correlation matrix is used as an input or there is no dropped indicator.
- **remain_DF**: Numeric data frame or numeric matrix. Set of retained indicators
- **...**: Other outputs came from principal, fa, or in ndr

Author(s)
Zsolt T. Kosztyan*, Marcell T. Kurbucz, Attila I. Katona
e-mail*: kosztyan.zsolt@gtk.uni-pannon.hu

References

See Also
principal, fa, ndr.

Examples

data<-I40_2020

library(psych)

# Principal Component Analysis (PCA)

pca<-principal(data,nfactors=2,covar=TRUE)
pca

# Feature selection with default values
PCA <- fs.dimred(pca, data)

# List of dropped, low communality value indicators
print(colnames(PCA$dropped_low))

# List of dropped, common communality value indicators
print(colnames(PCA$dropped_com))

# List of retained indicators
print(colnames(PCA$retained_DF))

# Principal Component Analysis (PCA) of correlation matrix
pca <- principal(cor(data, method = "spearman"), nfactors = 2, covar = TRUE)
pca

# Feature selection
min_comm <- 0.25 # Minimal communality value
com_comm <- 0.20 # Minimal common communality value
PCA <- fs.dimred(pca, cor(data, method = "spearman"), min_comm, com_comm)
PCA

---

**fs.KMO**  
*Feature selection for KMO*

**Description**

Drop variables if their MSA_i value is lower than a threshold, in order to increase the overall KMO (MSA) value.

**Usage**

fs.KMO(data, min_MSA=0.5, cor.mtx=FALSE)

**Arguments**

- **data**: A numeric data frame
- **min_MSA**: A numeric value. Minimal MSA value for variable i
- **cor.mtx**: Boolean value. The input is either a correlation matrix (cor.mtx=TRUE), or not (cor.mtx=FALSE)

**Details**

Low Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy does not suggest using principal component or factor analysis. Therefore, this function drop variables with low KMO/MSA values.
Value

data Cleaned data or the cleaned correlation matrix.

Author(s)

Zsolt T. Kosztyan*, Marcell T. Kurbucz, Attila I. Katona

e-mail*: kosztyan.zsolt@gtk.uni-pannon.hu

References


See Also

summary.

Examples

library(psych)
data(I40_2020)
data<-I40_2020
KMO(fs.KMO(data,min_MSA=0.7,cor.mtx=FALSE))

GOVDB2020

Governmental and economic data of countries (2020), where the data frame has 138 observations of 2161 variables.

Description

Sample datasets for Generalized Network-based Dimensionality Reduction and Analysis (GNDA) Governmental and economic data of countries (2020), where the data frame has 138 observations of 2161 variables.

Usage

data("GOVDB2020")

Format

A data frame with 138 observations of 2161 variables.

Source

Examples

data(GOVDB2020)

| I40_2020 | NUTS2 regional development data (2020) of I4.0 readiness, where the data frame has 414 observations of 101 variables. |

Description

Sample datasets for Generalized Network-based Dimensionality Reduction and Analysis (GNDA) NUTS2 regional development data (2020), where the data frame has 414 observations of 101 variables.

Usage

data("COVID19_2020")

Format

A data frame with 414 observations of 101 variables.

Source


Examples

data(I40_2020)

| ndr | Generalized Network-based Dimensionality Reduction and Analysis (GNDA) |

Description

The main function of Generalized Network-based Dimensionality Reduction and Analysis (GNDA).

Usage

ndr(r, covar=FALSE, cor_method=1, cor_type=1, min_R=0, min_comm=2, Gamma=1, null_modell_type=4, mod_mode=6, min_evalue=0, min_communality=0, com_communalities=0, use_rotation=FALSE)
Arguments

r A numeric data frame
covar If this value is FALSE (default), it finds the correlation matrix from the raw data.
If this value is TRUE, it uses the matrix r as a correlation/similarity matrix.
cor_method Correlation method (optional). ‘1’ Pearson’s correlation (default), ‘2’ Spearman’s correlation, ‘3’ Kendall’s correlation, ‘4’ Distance correlation
cor_type Correlation type (optional). ‘1’ Bivariate correlation (default), ‘2’ partial correlation, ‘3’ semi-partial correlation
min_R Minimal square correlation between indicators (default: 0).
min_comm Minimal number of indicators per community (default: 2).
Gamma Gamma parameter in multiresolution null model (default: 1).
null_modell_type '1' Differential Newmann-Grivan’s null model, '2' The null model is the mean of square correlations between indicators, '3' The null model is the specified minimal square correlation, '4' Newmann-Grivan’s modell (default)
mod_mode Community-based modularity calculation mode: '1' Louvain modularity, '2' Fast-greedy modularity, '3' Leading Eigen modularity, '4' Infomap modularity, '5' Walktrap modularity, '6' Leiden modularity (default)
min_evalue Minimal eigenvector centrality value (default: 0)
min_communality Minimal communality value of indicators (default: 0)
com_commonalities Minimal common communalities (default: 0)
use_rotation FALSE no rotation (default), TRUE varimax rotation

Details
NDA both works on low and high simple size datasets. If min_evalue=min_communality=com_commonalities=0 than there is no feature selection.

Value

communality Communalities estimates for each item. These are merely the sum of squared factor loadings for that item. It can be interpreted in correlation matrices.
loadings A standard loading matrix of class “loadings”.
uniqueness Uniqueness values of indicators.
factors Number of found factors.
scores Estimates of the factor scores are reported (if covar=FALSE).
n.obs Number of observations specified or found.
fn Factor name: NDA
Call Callback function
**Author(s)**

Zsolt T. Kosztyan*, Marcell T. Kurbucz, Attila I. Katona  
e-mail*: kosztyan.zsolt@gtk.uni-pannon.hu

**References**


**See Also**

`plot`, `biplot`, `summary`.

**Examples**

```
# Dimension reduction

data(swiss)
df<-swiss
p<-ndr(df)
summary(p)
plot(p)
biplot(p)

# Data reduction

# Distance is Euclidean's distance
# covar=TRUE means only the distance matrix is considered.

q<-ndr(1-normalize(as.matrix(dist(df))),covar=TRUE)
summary(q)
plot(q)
```

---

**normalize**

*Min-max normalization*

**Description**

Min-max normalization for data matrices and data frames

**Usage**

```r
normalize(x,type="all")
```

**Arguments**

- `x` : a data frame or data matrix
- `type` : type of normalization. "row" normalization row by row, "col" normalization column by column, "all" normalization for the entire data frame/matrix (default)
Value

Returns a normalized data.frame/matrix.

Author(s)

Zsolt T. Kosztyan, University of Pannonia
e-mail: kosztyan.zsolt@gtk.uni-pannon.hu

Examples

```r
mtx<-matrix(rnorm(20),5,4)
n_mtx<-normalize(mtx) # Fully normalized matrix
r_mtx<-normalize(mtx,type="row") # Normalize row by row
c_mtx<-normalize(mtx,type="col") # Normalize col by col
print(n_mtx) # Print fully normalized matrix
```

Description

Calculating partial distance correlation of columns of a matrix for Generalized Network-based Dimensionality Reduction and Analysis (GNDA).
The calculation is very slow for large matrices!

Usage

```r
dCcor(x)
```

Arguments

- `x`: a numeric matrix, or a numeric data frame

Value

Partial distance correlation matrix of `x`.

Author(s)

Prof. Zsolt T. Kosztyan, Department of Quantitative Methods, Institute of Management, Faculty of Business and Economics, University of Pannonia, Hungary
e-mail: kosztyan.zsolt@gtk.uni-pannon.hu

References

Examples

# Specification of partial distance correlation matrix.
x<-matrix(rnorm(36),nrow=6)
pdCor(x)

plot.nda

Plot function for Generalized Network-based Dimensionality Reduction and Analysis (GNDA)

Description

Plot variable network graph

Usage

## S3 method for class 'nda'
plot(x, cuts=0.3, interactive=TRUE, edgescale=1.0, labeldist=-1.5, ...)

Arguments

x an object of class 'NDA'.
cuts minimal square correlation value for an edge in the correlation network graph
(default 0.3).
interactive Plot interactive visNetwork graph or non-interactive igraph plot
(default TRUE).
edgescale Proportion scale value of edge width.
labeldist Vertex label distance in non-interactive igraph plot
(default value =-1.5).
... other graphical parameters.

Author(s)

Zsolt T. Kosztyan*, Marcell T. Kurbucz, Attila I. Katona
e-mail*: kosztyan.zsolt@gtk.uni-pannon.hu

See Also

biplot, summary.ndr.

Examples

# Plot function with feature selection
data("CrimesUSA1990.X")
df<-CrimesUSA1990.X
p<-ndr(df)
biplot(p,main="Biplot of CrimesUSA1990 without feature selection")
# Plot function with feature selection
# minimal eigen values (min_evalue) is 0.0065
# minimal communality value (min_communality) is 0.1
# minimal common communality value (com_communalities) is 0.1
p<-ndr(df,min_evalue = 0.0065,min_communality = 0.1,com_communalities = 0.1)

# Plot with default (cuts=0.3)
plot(p)

# Plot with higher cuts
plot(p,cuts=0.6)

---

**spdCor**

*Calculating semi-partial distance correlation of columns of a matrix*

**Description**

Calculating semi-partial distance correlation of two columns of a matrix for Generalized Network-based Dimensionality Reduction and Analysis (GNDA).

The calculation is very slow for large matrices!

**Usage**

```
spdCor(x)
```

**Arguments**

- `x` a numeric matrix, or a numeric data frame

**Value**

Semi-partial distance correlation matrix of x.

**Author(s)**

Prof. Zsolt T. Kosztyan, Department of Quantitative Methods, Institute of Management, Faculty of Business and Economics, University of Pannonia, Hungary

e-mail: kosztyan.zsolt@gtk.uni-pannon.hu

**References**

Examples

# Specification of semi-partial distance correlation matrix.
x<-matrix(rnorm(36),nrow=6)
spdCor(x)

summary.nda

Summary function of Generalized Network-based Dimensionality Reduction and Analysis (GNDA)

Description

Print summary of Generalized Network-based Dimensionality Reduction and Analysis (GNDA)

Usage

## S3 method for class 'nda'
summary(object, digits = getOption("digits"), ...)

Arguments

object an object of class 'nda'.
digits the number of significant digits to use when add.stats = TRUE.
... additional arguments affecting the summary produced.

Author(s)

Zsolt T. Kosztyan*, Marcell T. Kurbucz, Attila I. Katona
e-mail*: kzst@gtk.uni-pannon.hu

See Also

biplot, plot, ndr.

Examples

# Example of summary function of NDA without feature selection

data("CrimesUSA1990.X")
df<-CrimesUSA1990.X
p<-ndr(df)
summary(p)

# Example of summary function of NDA with feature selection
# minimal eigen values (min_evalue) is 0.0065
# minimal communality value (min_communality) is 0.1
# minimal common communality value (com_communalities) is 0.1
p<-ndr(df,min_evalue = 0.0065,min_communality = 0.1,com_communalities = 0.1)
summary(p)
Index

* array
  data_gen, 6
dCor, 7
dCov, 8
pdCor, 16
spdCor, 18
* control chart
  plot.nda, 17
* correlation matrix
  dCor, 7
dCov, 8
pdCor, 16
spdCor, 18
* datasets
  COVID19_2020, 4
  CrimesUSA1990.X, 4
  CrimesUSA1990.Y, 5
  CWTS_2020, 5
  GOVDB2020, 12
  I40_2020, 13
* dimensionality
  fs.dimred, 9
  fs.KMO, 11
  ndr, 13
* distance correlation
  dCor, 7
dCov, 8
pdCor, 16
spdCor, 18
* matrix
  normalize, 15
* multivariate
  data_gen, 6
dCor, 7
dCov, 8
fs.dimred, 9
fs.KMO, 11
ndr, 13
pdCor, 16
plot.nda, 17
spdCor, 18
summary.nda, 19
* nonparametric
  ndr, 13
* package
  nda-package, 2
* plot
  biplot.nda, 3
* random block matrix
  data_gen, 6
* reduction
  fs.dimred, 9
  fs.KMO, 11
  ndr, 13
biplot, 2, 15, 17, 19
biplot.nda, 3
COVID19_2020, 4
CrimesUSA1990.X, 4
CrimesUSA1990.Y, 5
CWTS_2020, 5
data_gen, 3, 6
dCor, 2, 7
dCov, 8
fa, 10
fs.dimred, 9
fs.KMO, 11
GOVDB2020, 12
I40_2020, 13
nda (nda-package), 2
nda-package, 2
ndr, 2, 3, 10, 13, 17, 19
normalize, 15
pdCor, 16
plot, 2, 3, 15, 19
plot.nda, 17
principal, 10

spdCor, 18
summary, 2, 3, 12, 15, 17
summary.nda, 19