

Package ‘MatrixCorrelation’

May 5, 2017

Encoding latin1

Type Package

Title Matrix Correlation Coefficients

Version 0.9.1

Date 2017-05-05

Author Kristian Hovde Liland [aut, cre],
Tormod Næs [ctb],
Ulf Geir Indahl [ctb]

Maintainer Kristian Hovde Liland <kristian.liland@nofima.no>

Description Computation and visualization of matrix correlation coefficients.
The main method is the Similarity of Matrices Index, while various related measures like r1, r2, r3, r4, Yanai's GCD, RV, RV2 and adjusted RV are included for comparison.

Depends R (>= 3.0.1)

Imports Rcpp, plotrix, pracma, progress, RSpectra

License GPL-2

LazyData TRUE

RoxygenNote 6.0.1

LinkingTo progress, Rcpp, RcppArmadillo

NeedsCompilation yes

Repository CRAN

Date/Publication 2017-05-05 21:07:37 UTC

R topics documented:

allCorrelations	2
candy	3
cor.test_eq	4
MatrixCorrelation	5
PCAcv	5

plot.SMI	6
PSI	8
r1	9
RV	10
significant	11
SMI	13

Index	15
--------------	-----------

allCorrelations	<i>All correlations</i>
-----------------	-------------------------

Description

Compare all correlation measures in the package (or a subset)

Usage

```
allCorrelations(X1, X2, ncomp1, ncomp2, methods = c("SMI", "RV", "RV2",
  "RVadj", "PSI", "r1", "r2", "r3", "r4", "GCD"), digits = 3, plot = TRUE,
  xlab = "", ylab = "", ...)
```

Arguments

X1	first matrix to be compared (data.frames are also accepted).
X2	second matrix to be compared (data.frames are also accepted).
ncomp1	maximum number of subspace components from the first matrix.
ncomp2	maximum number of subspace components from the second matrix.
methods	character vector containing a subset of the supported methods: "SMI", "RV", "RV2", "RVadj", "PSI", "r1", "r2", "r3", "r4", "GCD".
digits	number of digits for numerical output.
plot	logical indicating if plotting should be performed (default = TRUE).
xlab	optional x axis label.
ylab	optional y axis label.
...	additional arguments for SMI or plot.

Details

For each of the three coefficients a single scalar is computed to describe the similarity between the two input matrices.

Value

A single value measuring the similarity of two matrices.

Author(s)

Kristian Hovde Liland

References

- RV: Robert, P.; Escoufier, Y. (1976). "A Unifying Tool for Linear Multivariate Statistical Methods: The RV-Coefficient". *Applied Statistics* 25 (3): 257-265.
- RV2: Smilde, AK; Kiers, HA; Bijlsma, S; Rubingh, CM; van Erk, MJ (2009). "Matrix correlations for high-dimensional data: the modified RV-coefficient". *Bioinformatics* 25(3): 401-5.
- Adjusted RV: Maye, CD; Lorent, J; Horgan, GW. (2011). "Exploratory analysis of multiple omics datasets using the adjusted RV coefficient". *Stat Appl Genet Mol Biol.* 10(14).

See Also

[SMI](#), [RV](#) (RV2/RVadj), [r1](#) (r2/r3/r4/GCD).

Examples

```
X1 <- scale( matrix( rnorm(100*300), 100,300), scale = FALSE)
usv <- svd(X1)
X2 <- usv$u[,-3] %*% diag(usv$d[-3]) %*% t(usv$v[,-3])

allCorrelations(X1,X2, ncomp1 = 5,ncomp2 = 5)
```

candy

Candy data

Description

Measurements from sensory analysis (professional tasting) on a number of candy products obtained by sensory labs. The two labs and the associated data sets are parts of a larger study described in Tomic et al. (2010),

Usage

```
data(candy)
```

Format

Two matrices of dimension 18 x 6.

References

Tomic, O., Luciano, G., Nilsen, A., Hyldig, G., Lorensen, K., Næs, T. (2010). Analysing sensory panel performance in a proficiency test using the PanelCheck software. *European Food Research and Technology.* 230. 3, 497-511

`cor.test_eq`*Test for no correlation between paired sampes*

Description

Permutation test for squared Pearson correlation between to vectors of samples.

Usage

```
cor.test_eq(x, y, B = 10000)
```

Arguments

<code>x</code>	first vector to be compared (or two column matrix/data.frame).
<code>y</code>	second vector to be compared (ommit if included in x).
<code>B</code>	integer number of permutations, default = 10000.

Details

This is a convenience function combining `SMI` and `significant` for the special case of vector vs vector comparisons. The nullhypothesis is that the correlation between the vectors is +/-1, while significance signifies a deviance toward 0.

Value

A value indicating if the two input vectors are significantly different.

Author(s)

Kristian Hovde Liland

References

Similarity of Matrices Index - Ulf Geir Indahl, Tormod NÅ's, Kristian Hovde Liland

See Also

[plot.SMI](#) (`print.SMI/summary.SMI`), [RV](#) (`RV2/RVadj`), [r1](#) (`r2/r3/r4/GCD`), [allCorrelations](#) (matrix correlation comparison), [PCAcv](#) (cross-validated PCA).

Examples

```
a <- (1:5) + rnorm(5)
b <- (1:5) + rnorm(5)
cor.test_eq(a,b)
```

MatrixCorrelation *Similarity of Matrices Coefficients*

Description

Computation and visualization of matrix correlation coefficients. The main method is the Similarity of Matrices Index, while various related measures like r1, r2, r3, r4, Yanai's GCD, RV, RV2 and adjusted RV are included for comparison.

Author(s)

Maintainer: Kristian Hovde Liland <kristian.liland@nofima.no>

Other contributors:

- Tormod Næs [contributor]
- Ulf Geir Indahl [contributor]

See Also

[SMI](#), [plot.SMI](#) (print.SMI/summary.SMI), [RV](#) (RV2/RVadj), [r1](#) (r2/r3/r4/GCD), [allCorrelations](#) (matrix correlation comparison).

PCAcv *Principal Component Analysis cross-validation error*

Description

PRESS values for PCA as implemented by Eigenvector and described by Bro et al. (2008).

Usage

```
PCAcv(X, ncomp)
```

Arguments

X	matrix object to perform PCA on.
ncomp	integer number of components.

Details

For each number of components predicted residual sum of squares are calculated based on leave-one-out cross-validation. The implementation ensures no over-fitting or information bleeding.

Value

A vector of PRESS-values.

Author(s)

Kristian Hovde Liland

References

R. Bro, K. Kjeldahl, A.K. Smilde, H.A.L. Kiers, Cross-validation of component models: A critical look at current methods. *Anal Bioanal Chem* (2008) 390: 1241-1251.

See Also

[plot.SMI](#) (print.SMI/summary.SMI), [RV](#) (RV2/RVadj), [r1](#) (r2/r3/r4/GCD), [allCorrelations](#) (matrix correlation comparison).

Examples

```
X1 <- scale( matrix( rnorm(100*300), 100,300), scale = FALSE)
PCAcv(X1,10)
```

plot.SMI

Result functions for the Similarity of Matrices Index (SMI)

Description

Plotting, printing and summary functions for SMI, plus significance testing.

Usage

```
## S3 method for class 'SMI'
plot(x, y = NULL, x1lab = attr(x, "mat.names")[[1]],
     x2lab = attr(x, "mat.names")[[2]], main = "SMI", signif = 0.05,
     xlim = c(-(pq[1] + 1)/2, (pq[2] + 1)/2), ylim = c(0.5, (sum(pq) + 3)/2),
     B = 10000, cex = 1, cex.sym = 1, frame = NULL, frame.col = "red",
     frame.lwd = 2, replicates = NULL, ...)
```

```
## S3 method for class 'SMI'
print(x, ...)
```

```
## S3 method for class 'SMI'
summary(object, ...)
```

```
is.signif(x, signif = 0.05, B = 10000, ...)
```

Arguments

<code>x</code>	object of class SMI.
<code>y</code>	not used.
<code>x1lab</code>	optional label for first matrix.
<code>x2lab</code>	optional label for second matrix.
<code>main</code>	optional heading (default = SMI).
<code>signif</code>	significance level for testing (default=0.05).
<code>xlim</code>	optional plotting limits.
<code>ylim</code>	optional plotting limits.
<code>B</code>	number of permutations (for significant, default=10000).
<code>cex</code>	optional text scaling (default = 1)
<code>cex.sym</code>	optional scaling for significance symbols (default = 1)
<code>frame</code>	two element integer vector indicating framed components.
<code>frame.col</code>	color for framed components.
<code>frame.lwd</code>	line width for framed components.
<code>replicates</code>	vector of replicates for significance testing.
<code>...</code>	additional arguments for plot.
<code>object</code>	object of class SMI.

Details

For plotting a diamond plot is used. High SMI values are light and low SMI values are dark. If orthogonal projections have been used for calculating SMIs, significance symbols are included in the plot unless `signif=NULL`.

Value

`plot` silently returns `NULL`. `print` and `summary` return the printed matrix.

Author(s)

Kristian Hovde Liland

References

Similarity of Matrices Index - Ulf G. Indahl, Tormod NÅs, Kristian Hovde Liland

See Also

[SMI](#), [PCAcv](#) (cross-validated PCA).

Examples

```

X1 <- scale( matrix( rnorm(100*300), 100,300), scale = FALSE)
usv <- svd(X1)
X2 <- usv$u[,-3] %*% diag(usv$d[-3]) %*% t(usv$v[,-3])

smi <- SMI(X1,X2,5,5)
plot(smi, B = 1000) # default B = 10000
print(smi)
summary(smi)
is.signif(smi, B = 1000) # default B = 10000

```

 PSI

Procrustes Similarity Index

Description

An index based on the RV coefficient with Procrustes rotation.

Usage

```
PSI(X1, X2)
```

Arguments

X1 first matrix to be compared (data.frames are also accepted).
 X2 second matrix to be compared (data.frames are also accepted).

Value

The Procrustes Similarity Index

References

Sibson, R; 1978. "Studies in the Robustness of Multidimensional Scaling: Procrustes Statistics".
 Journal of the Royal Statistical Society. Series B (Methodological), Vol. 40, No. 2, pp. 234-238.

Examples

```

X1 <- scale( matrix( rnorm(100*300), 100,300), scale = FALSE)
usv <- svd(X1)
X2 <- usv$u[,-3] %*% diag(usv$d[-3]) %*% t(usv$v[,-3])
PSI(X1,X2)

```


Description

Matrix similarity as described by Ramsey et al. (1984).

Usage

$r1(X1, X2)$

$r2(X1, X2)$

$r3(X1, X2)$

$r4(X1, X2)$

$GCD(X1, X2, ncomp1 = Rank(X1) - 1, ncomp2 = Rank(X2) - 1)$

Arguments

X1	first matrix to be compared (data.frames are also accepted).
X2	second matrix to be compared (data.frames are also accepted).
ncomp1	(GCD) number of subspace components from the first matrix (default: full subspace).
ncomp2	(GCD) number of subspace components from the second matrix (default: full subspace).

Details

Details can be found in Ramsey's paper:

- r1: inner product correlation
- r2: orientation-independent inner product correlation
- r3: spectra-independent inner product correlations (including orientation)
- r4: Spectra-Independent inner product Correlations
- GCD: Yanai's GCD Measure. To reproduce the original GCD, use all components.

Value

A single value measuring the similarity of two matrices.

Author(s)

Kristian Hovde Liland

References

Ramsay, JO; Berg, JT; Styan, GPH; 1984. "Matrix Correlation". Psychometrika 49(3): 403-423.

See Also

[SMI, RV](#) (RV2/RVadj).

Examples

```
X1 <- matrix(rnorm(100*300),100,300)
usv <- svd(X1)
X2 <- usv$u[,-3] %*% diag(usv$d[-3]) %*% t(usv$v[,-3])

r1(X1,X2)
r2(X1,X2)
r3(X1,X2)
r4(X1,X2)
GCD(X1,X2)
GCD(X1,X2, 5,5)
```

RV

RV coefficients

Description

Three different RV coefficients: RV, RV2 and adusted RV.

Usage

```
RV(X1, X2)

RV2(X1, X2)

RVadjMaye(X1, X2)

RVadjGhaziri(X1, X2)

RVadj(X1, X2, version = c("Maye", "Ghaziri"))
```

Arguments

X1	first matrix to be compared (data.frames are also accepted).
X2	second matrix to be compared (data.frames are also accepted).
version	Which version of RV adjusted to apply: "Maye" (default) or "Ghaziri" RV adjusted is run using the RVadj function.

Details

For each of the four coefficients a single scalar is computed to describe the similarity between the two input matrices.

Value

A single value measuring the similarity of two matrices.

Author(s)

Kristian Hovde Liland, Benjamin Leutner (RV2)

References

- RV: Robert, P.; Escoufier, Y. (1976). "A Unifying Tool for Linear Multivariate Statistical Methods: The RV-Coefficient". *Applied Statistics* 25 (3): 257-265.
- RV2: Smilde, AK; Kiers, HA; Bijlsma, S; Rubingh, CM; van Erk, MJ (2009). "Matrix correlations for high-dimensional data: the modified RV-coefficient". *Bioinformatics* 25(3): 401-5.
- Adjusted RV: Maye, CD; Lorent, J; Horgan, GW. (2011). "Exploratory analysis of multiple omics datasets using the adjusted RV coefficient". *Stat Appl Genet Mol Biol.* 10(14).
- Adjusted RV: El Ghaziri, A; Qannari, E.M. (2015) "Measures of association between two datasets; Application to sensory data", *Food Quality and Preference* 40 (A): 116-124.

See Also

[SMI](#), [r1](#) (r2/r3/r4/GCD).

Examples

```
X1 <- matrix(rnorm(100*300),100,300)
usv <- svd(X1)
X2 <- usv$u[, -3] %*% diag(usv$d[-3]) %*% t(usv$v[, -3])

RV(X1,X2)
RV2(X1,X2)
RVadj(X1,X2)
```

significant

Significance estimation for Similarity of Matrices Index (SMI)

Description

Permutation based hypothesis testing for SMI. The nullhypothesis is that a linear function of one matrix subspace is included in the subspace of another matrix.

Usage

```
significant(smi, B = 10000, replicates = NULL)
```

Arguments

<code>smi</code>	<code>smi</code> object returned by call to SMI.
<code>B</code>	integer number of permutations, default = 10000.
<code>replicates</code>	integer vector of replicates.

Details

For each combination of components significance is estimated by sampling from a null distribution of no similarity, i.e. when the rows of one matrix is permuted `B` times and corresponding SMI values are computed. If the vector `replicates` is included, replicates will be kept together through permutations.

Value

A matrix containing P-values for all combinations of components.

Author(s)

Kristian Hovde Liland

References

Similarity of Matrices Index - Ulf G. Indahl, Tormod N  s, Kristian Hovde Liland

See Also

[plot.SMI](#) (`print.SMI/summary.SMI`), [RV](#) (`RV2/RVadj`), [r1](#) (`r2/r3/r4/GCD`), [allCorrelations](#) (matrix correlation comparison).

Examples

```
X1 <- scale( matrix( rnorm(100*300), 100,300), scale = FALSE)
usv <- svd(X1)
X2 <- usv$u[,-3] %*% diag(usv$d[-3]) %*% t(usv$v[,-3])

(smi <- SMI(X1,X2,5,5))
significant(smi, B = 1000) # default B = 10000
```

 SMI *Similarity of Matrices Index (SMI)*

Description

A similarity index for comparing coupled data matrices.

Usage

```
SMI(X1, X2, ncomp1 = Rank(X1) - 1, ncomp2 = Rank(X2) - 1,
    projection = "Orthogonal", Scores1 = NULL, Scores2 = NULL)
```

Arguments

X1	first matrix to be compared (data.frames are also accepted).
X2	second matrix to be compared (data.frames are also accepted).
ncomp1	maximum number of subspace components from the first matrix.
ncomp2	maximum number of subspace components from the second matrix.
projection	type of projection to apply, defaults to "Orthogonal", alternatively "Procrustes".
Scores1	user supplied score-matrix to replace singular value decomposition of first matrix.
Scores2	user supplied score-matrix to replace singular value decomposition of second matrix.

Details

A two-step process starts with extraction of stable subspaces using Principal Component Analysis or some other method yielding two orthonormal bases. These bases are compared using Orthogonal Projection (OP / ordinary least squares) or Procrustes Rotation (PR). The result is a similarity measure that can be adjusted to various data sets and contexts and which includes explorative plotting and permutation based testing of matrix subspace equality.

Value

A matrix containing all combinations of components. Its class is "SMI" associated with print, plot, summary methods.

Author(s)

Kristian Hovde Liland

References

A similarity index for comparing coupled matrices - Ulf Geir Indahl, Tormod NÅ's, Kristian Hovde Liland

See Also

`plot.SMI` (`print.SMI/summary.SMI`), `RV` (`RV2/RVadj`), `r1` (`r2/r3/r4/GCD`), `allCorrelations` (matrix correlation comparison), `PCAcv` (cross-validated PCA).

Examples

```
# Simulation
X1 <- scale( matrix( rnorm(100*300), 100,300), scale = FALSE)
usv <- svd(X1)
X2 <- usv$u[,-3] %*% diag(usv$d[-3]) %*% t(usv$v[,-3])

(smi <- SMI(X1,X2,5,5))
plot(smi, B = 1000 ) # default B = 10000

# Sensory analysis
data(candy)
plot( SMI(candy$Panel1, candy$Panel2, 3,3, projection = "Procrustes"),
      frame = c(2,2), B = 1000 ) # default B = 10000
```

Index

`allCorrelations`, [2](#), [4-6](#), [12](#), [14](#)

`candy`, [3](#)

`cor.test_eq`, [4](#)

`GCD (r1)`, [9](#)

`is.signif (plot.SMI)`, [6](#)

`MatrixCorrelation`, [5](#)

`MatrixCorrelation-package`
(`MatrixCorrelation`), [5](#)

`PCAcv`, [4](#), [5](#), [7](#), [14](#)

`plot.SMI`, [4-6](#), [6](#), [12](#), [14](#)

`print.SMI (plot.SMI)`, [6](#)

`PSI`, [8](#)

`r1`, [3-6](#), [9](#), [11](#), [12](#), [14](#)

`r2 (r1)`, [9](#)

`r3 (r1)`, [9](#)

`r4 (r1)`, [9](#)

`RV`, [3-6](#), [10](#), [10](#), [12](#), [14](#)

`RV2 (RV)`, [10](#)

`RVadj (RV)`, [10](#)

`RVadjGhaziri (RV)`, [10](#)

`RVadjMaye (RV)`, [10](#)

`significant`, [11](#)

`SMI`, [3](#), [5](#), [7](#), [10](#), [11](#), [13](#)

`summary.SMI (plot.SMI)`, [6](#)