Package 'fastHICA'

May 12, 2015

Type Package

Title Hierarchical Independent Component Analysis: a Multi-Scale Sparse Non-Orthogonal Data-Driven Basis

Version 1.0.2

Date 2015-05-12

Author Piercesare Secchi, Simone Vantini, and Paolo Zanini

Maintainer Paolo Zanini <paolo.zanini@polimi.it>

Depends fastICA, energy

Description It implements HICA (Hierarchical Independent Component Analysis) algorithm. This approach, obtained through the integration between treelets and Independent Component Analysis, is able to provide a multi-scale non-orthogonal data-driven basis, whose elements have a phenomenological interpretation according to the problem under study.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2015-05-12 11:39:40

R topics documented:

basis_hica 2
energy_hica 3
extract_hica 5
similarity_hica 6

Index **8**

basis_hica *Construction of the HICA basis*

Description

This function builds the HICA tree up to a prespecified height providing the corresponding non-orthogonal bases.

Usage

```
basis_hica(X, maxlev = dim(X)[2] - 1, dim.subset = 512)
```

Arguments

X	Data matrix with nrow(X) observations and ncol(X) variables.
maxlev	The maximum level of the tree. This must be an integer between 1 and ncol(X)-1. The default value is set to ncol(X)-1.
dim.subset	The dimension of the subset used for the evaluation of the similarity index (i.e., distance correlation). If this it is greater than nrow(X) all the observations are used, unless a random subsample of dim.subset observations is used. The default value is set to 512.

Value

X	data matrix.
basis	a list with maxlev elements. The ith element of the list contains the basis matrix provided at level i of the tree. Each column of the basis matrix represent a basis element.
aggregation	a matrix with maxlev rows and 3 columns. At each row the first two columns contain the variable indeces merged at the corresponding level of the tree. In the third column the distance correlation of the two merged variables is recorded.

Note

The distance correlation is evaluated through the function dcor of the package "energy". It becomes computationally unfeasible if the number of observations is too large. For this reason it is possibile to choose the dimension of the subsample to be used in the evaluation of the similarity matrix. By default the dimension is set to 512.

Author(s)

Piercesare Secchi, Simone Vantini, and Paolo Zanini.

References

P. Secchi, S. Vantini, and P. Zanini (2014). Hierarchical Independent Component Analysis: a multi-resolution non-orthogonal data-driven basis. *MOX-report 01/2014*, Politecnico di Milano.

See Also

energy_hica, similarity_hica, extract_hica

Examples

```
## Not run:

############################################################
# Example - Independent sources and overlapping loadings #
############################################################

c1=c(0,0,0,0,1,1)
c2=c(1,1,1,1,0,0)
c3=c(1,1,0,0,0,0)

s1=runif(400,0,20)
s2=runif(400,0,20)
s3=runif(400,0,20)

# Here we generate the simulated dataset

X=s1%*%t(c1)+s2%*%t(c2)+s3%*%t(c3)+rnorm(6*400,0,1)

# Here we perform HICA on the simulated dataset

basis=basis_hica(X,5)

# Here we plot the 3 main components of HICA basis
# (according to the energy criterium) for 4th level

energy=energy_hica(basis,6,5,plot=TRUE)
ex4=extract_hica(energy,3,4)
loa4=ex4$C

par( mfrow = c(3,1))
barplot(loa4[,1], ylim = c(-1, 1),main="HICA transform - Level 4",
ylab="1st component",xlab="Coordinate",names.arg=1:6,col="red",mgp=c(2.5,1,0))
barplot(loa4[,2], ylim = c(-1, 1),ylab="2nd component",
xlab="Coordinate",names.arg=1:6,col="green",mgp=c(2.5,1,0))
barplot(loa4[,3], ylim = c(-1, 1),ylab="3rd component",
xlab="Coordinate",names.arg=1:6,col="blue",mgp=c(2.5,1,0))

## End (Not run)
```

---

energy_hica *Energy criterion*

Description

This function implements the energy criterion defined in Secchi, Vantini, and Zanini (2013).

Usage

```
energy_hica(HICA.obj, maxcomp = 1, nlevel = 1, plot = FALSE)
```

Arguments

HICA.obj	An object provided by the function basis_hica.
maxcomp	The maximum space dimension considered.
nlevel	The number of levels analyzed. Specifically the levels from p-nlevel to p-1 are analyzed, where p is the number of variables.
plot	A logical value. If TRUE the energy is plotted.

Details

This function computes the energy according the criterion presented in Secchi, Vantini and Zanini (2013). It is useful to find the best representation. It receives in input the output of the basis_hica function.

Value

energy	A matrix with maxcomp rows and p-1 columns, where p is the number of variables. In position (i,j) it contains the energy of the best i-dimensional space for the jth level of the tree. Only the last nlevel columns are filled.
components	A matrix with maxcomp rows and p-1 columns, where p is the number of variables. In position (i,j), it contains the index of th ith basis element for jth level of the tree. Only the last nlevel columns are filled.
HICA.obj	The same object, output of the function basis_hica, provided in input.

Author(s)

Piercesare Secchi, Simone Vantini, and Paolo Zanini

References

P. Secchi, S. Vantini, and P. Zanini (2014). Hierarchical Independent Component Analysis: a multi-resolution non-orthogonal data-driven basis. *MOX-report 01/2014*, Politecnico di Milano.

See Also

basis_hica, similarity_hica, extract_hica

`extract_hica` *Extraction of score and loading matrices.*

Description

This function extracts the score matrix and the loading matrix given the dimension of the subspace considered and the level of the tree chosen. Furthermore it provides the cumulant energies for the subspace extracted.

Usage

```
extract_hica(energy.obj, comp, level)
```

Arguments

`energy.obj`	An object provided by the function `energy_hica`.
`comp`	Dimension of the subspace.
`level`	Level of the tree.

Value

`X`	data matrix.
`S`	score data matrix.
`C`	loading matrix. Each column represents a basis element.
`cum.energy`	cumulant energy for the subspace extracted.

Author(s)

Piercesare Secchi, Simone Vantini, and Paolo Zanini.

References

P. Secchi, S. Vantini, and P. Zanini (2014). Hierarchical Independent Component Analysis: a multi-resolution non-orthogonal data-driven basis. *MOX-report 01/2014*, Politecnico di Milano.

See Also

`basis_hica`, `similarity_hica`, `energy_hica`

`similarity_hica` *Estimate of the similarity matrix*

Description

This function provides an estimate of the similarity matrix of the original data, before performing HICA algorithm.

Usage

```
similarity_hica(X, dim.subset = 512)
```

Arguments

`X`	Data matrix with nrow(X) observations and ncol(X) variables.
`dim.subset`	The dimension of the subset used for the evaluation of the similarity index (i.e., distance correlation). If this it is greater than nrow(X) all the observations are used, unless a random subset of `dim.subset` observations is used. The default value is set to 512.

Details

This function is auxiliary for the `basis_hica` function. Indeed its output is the estimate of the similarity matrix at the first step of the algorithm.

Value

`similarity_matrix`	similarity matrix of the original data.
`subset`	subset used for the evaluation of distance correlation between variables.

Note

The distance correlation is evaluated through the function `dcor` of the package "energy". It becomes computationally unfeasible if the number of observations is too large. For this reason it is possibile to choose the dimension of the subsample to be used in the evaluation of the similarity matrix. By default the dimension is set to 512.

Author(s)

Piercesare Secchi, Simone Vantini, and Paolo Zanini.

References

P. Secchi, S. Vantini, and P. Zanini (2014). Hierarchical Independent Component Analysis: a multi-resolution non-orthogonal data-driven basis. *MOX-report 01/2014*, Politecnico di Milano.

See Also

basis_hica, energy_hica, extract_hica

Index

basis_hica, 2, *4–7*

energy_hica, *3*, 3, *5*, *7*
extract_hica, *3*, *4*, 5, *7*

similarity_hica, *3–5*, 6