Package 'cauchypca'

November 9, 2022

Type Package

Title Robust Principal Component Analysis Using the Cauchy Distribution

Version 1.0

URL

Date 2022-11-09

Author Michail Tsagris [aut, cre],
Aisha Fayomi [ctb],
Yannis Pantazis [ctb],
Andrew T.A. Wood [ctb]

Maintainer Michail Tsagris <mtsagris@uoc.gr>

Depends R (>= 4.0)

Imports doParallel, foreach, parallel, Rfast, stats

Description A new robust principal component analysis algorithm is implemented that relies upon the Cauchy Distribution. The algorithm is suitable for high dimensional data even if the sample size is less than the number of variables. The methodology is described in this paper: Fayomi A., Pantazis Y., Tsagris M. and Wood A.T.A. (2022). Cauchy robust principal component analysis with applications to high-dimensional data sets. <arXiv:2211.03181>.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2022-11-09 15:50:21 UTC

R topics documented:

cauchypca-package *Robust Principal Component Analysis Using the Cauchy Distribution*

Description

A new robust principal component analysis algorithm is implemented that relies upon the Cauchy Distribution. The algorithm is suitable for high dimensional data even if the sample size is less than the number of variables.

Details

Package:	cauchypca
Type:	Package
Version:	1.0
Date:	2022-11-09
License:	GPL-2

Maintainers

Michail Tsagris <mtsagris@uoc.gr>.

Author(s)

Michail Tsagris <mtsagris@uoc.gr>, Aisha Fayomi <afayomi@kau.edu.sa>, Yannis Pantazis <pantazis@iacm.forth.gr> and Andrew T.A. Wood <Andrew.Wood@anu.edu.au>.

References

Fayomi A., Pantazis Y., Tsagris M. and Wood A.T.A. (2022). Cauchy robust principal component analysis with applications to high-dimensional data sets. <arXiv:2211.03181>.

Robust PCA using the Cauchy distribution
Robust PCA using the Cauchy distribution

Description

Robust PCA using the Cauchy distribution.

Usage

```
cauchy.pca(x, k = 1, center = "sm", scale = "mad", trials = 20, parallel = FALSE)
```

Arguments

x	A numerical matrix with the data.
k	The number of eigenvectors to extract.
center	The way to center the data. This can be either "sm" corresponding to the spatial median, "med" corresponding to the classical column-wise median or a vector supplied by the user.
scale	This is the method to scale the data. The default value is "mad" corresponding to the mean absolute deviation, computed column-wise. Alternatively the user can provide their own vector.
trials	The number of trials to attempt. How many times the algorithm will be performed with different starting values (different starting vectors).
parallel	If you want parallel computations set this equal to TRUE.

Details

This is the main function used to extract the Cauchy robust eigenvectors.

Value

A list including:

runtime	The duration (in seconds) of the algorithm.
loglik	The minimum maximum Cauchy log-likelihood.
mu	The estimated location parameter of the Cauchy ditribution.
su	The estimated scale parameter of the Cauchy ditribution.
loadings	A matrix with the robust eigenvectors.

Author(s)

Michail Tsagris, Aisha Fayomi, Yannis Pantazis and Andrew T.A. Wood.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Fayomi A., Pantazis Y., Tsagris M. and Wood A.T.A. (2022). Cauchy robust principal component analysis with applications to high-dimensional data sets. <arXiv:2211.03181>.

Examples

```
x <- as.matrix( iris[, 1:4] )
cauchy.pca(x, k = 1)
```

Index