Package 'cauchypca'

January 24, 2024

Type Package

Title Robust Principal Component Analysis Using the Cauchy Distribution

Version 1.3

Date 2024-01-24

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, Rfast2, 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. (2024). ``Cauchy robust principal component analysis with applications to high-dimensional data sets". Statistics and Computing, 34: 26. <doi:10.1007/s11222-023-10328-x>.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2024-01-24 19:50:02 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:	1.3
Date:	2024-01-24
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. (2024). Cauchy robust principal component analysis with applications to high-dimensional data sets. Statistics and Computing, 34: 26. https://doi.org/10.1007/s11222-023-10328-x

MLE of the Cauchy distribution

MLE of the Cauchy distribution

Description

MLE of the Cauchy distribution.

Usage

```
cauchy.mle(x, tol = 1e-07)
```

Arguments

`x`	A numerical vector with data.
`tol`	The tolerance level up to which the maximisation stops set to 1e-09 by default.

Details

Instead of maximising the log-likelihood via a numerical optimiser we have used a Newton-Raphson algorithm which is faster. The Cauchy is the t distribution with 1 degree of freedom.

Value

A list including:

`iters`	The number of iterations required for the Newton-Raphson to converge.
`loglik`	The value of the maximised log-likelihood.
`param`	The vector of the parameters.

Author(s)

Michail Tsagris

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

References

Johnson, Norman L. Kemp, Adrianne W. Kotz, Samuel (2005). Univariate Discrete Distributions (third edition). Hoboken, NJ: Wiley-Interscience.

https://en.wikipedia.org/wiki/Wigner_semicircle_distribution

See Also

`cauchy.pca`

Examples

```
x <- rcauchy(1000)
a <- cauchy.mle(x)
```

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 variable-wise median. Alternatively the user can specify their own vector.
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. (2024). Cauchy robust principal component analysis with applications to high-dimensional data sets. Statistics and Computing, 34: 26. https://doi.org/10.1007/s11222-023-10328-x

See Also

cauchy.mle

Examples

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

Index