

Package ‘RMSDp’

January 20, 2025

Type Package

Title Refined Modified Stahel-Donoho (MSD) Estimators for Outlier Detection (Parallel Version)

Version 0.1.1

Suggests testthat (>= 3.0.0)

Depends R (>= 2.10), stats

Imports parallel, doParallel, foreach

Description A parallel function for multivariate outlier detection named modified Stahel-Donoho estimators is contained in this package. The function RMSDp() is for elliptically distributed datasets and recognizes outliers based on Mahalanobis distance. This function is for higher dimensional datasets that cannot be handled by a single core function RMSD() included in 'RMSD' package. See Wada and Tsubaki (2013) <[doi:10.1109/CLOUDCOM-ASIA.2013.86](https://doi.org/10.1109/CLOUDCOM-ASIA.2013.86)> for the detail of the algorithm.

License GPL (>= 3)

Encoding UTF-8

Language en-US

RoxygenNote 7.3.1

Config/testthat/edition 3

LazyData true

NeedsCompilation no

Author Kazumi Wada [aut, cre] (<<https://orcid.org/0000-0002-9578-1588>>)

Maintainer Kazumi Wada <kazwd2008@gmail.com>

Repository CRAN

Date/Publication 2024-06-12 21:00:21 UTC

Contents

RMSDp	2
wine	3
Index	4

RMSDp

Modified Stahel-Donoho Estimators (parallel version)

Description

This function is for multivariate outlier detection. version 0.0.1 2013/06/15 Related paper: DOI: 10.1109/CLOUDCOM-ASIA.2013.86 version 0.0.2 2021/11/15 Outlier detection step added version 0.0.3 2022/08/12 Bug fixed about Random seed setting

Usage

```
RMSDp(inp, cores = 0, nb = 0, sd = 0, pt = 0.999, dv = 10000)
```

Arguments

inp	input data (a numeric matrix)
cores	number of cores used for this function
nb	number of basis
sd	seed (for reproducibility)
pt	threshold for outlier detection (probability)
dv	maximum number of elements processed together on the same core

Value

a list of the following information

- u final mean vector
- V final covariance matrix
- wt final weights
- mah squared squared Mahalanobis distances
- cf threshold to detect outlier (percentile point)
- ot outlier flag (1:normal observation, 2:outlier)

wine

Wine dataset in UCI Machine Learning Repository

Description

A subset of data from the World Health Organization Global Tuberculosis Report ...

Usage

wine

Format

'wine' A data frame with 178 rows and 13 columns:

Alcohol Alcohol

Malic acid Malic acid

Ash Ash

Alcalinity of ash Alcalinity of ash

Magnesium Magnesium

Total phenols Total phenols

Flavonoids Flavonoids

Nonflavanoid phenols Nonflavanoid phenols

Proanthocyanins Proanthocyanins

Color intensity Color intensity

Hue Hue

OD280/OD315 of diluted wines OD280/OD315 of diluted wines

Proline Proline

Source

<<https://archive.ics.uci.edu/dataset/109/wine>>

Index

* **datasets**
 wine, 3

RMSDp, 2

wine, 3