Package: outliersHD (via r-universe)

July 7, 2026

Type Package

Title Detection of Outliers in High Dimensional Data

Version 1.0

Date 2026-06-29

Author Michail Tsagris [aut, cre]

Maintainer Michail Tsagris <mtsagris@uoc.gr>

Depends R (>= 4.0)

Imports Rfast, Rfast2, Rnanoflann, stats

Description Algorithms to detect high-dimensional outliers. The minimum diagonal product of Ro, Zou, Wang and Yin (2015) <doi:10.1093/biomet/asv021>, the algorithm of Wilkinson (2018) <doi:10.1109/TVCG.2017.2744685>, and the distances of distances of Lee and Jeon (2025) <doi:10.48550/arXiv.2511.02199>.

License GPL (>= 2)

NeedsCompilation no

Repository https://cran.r-universe.dev

Date/Publication 2026-07-04 08:10:02 UTC

RemoteUrl https://github.com/cran/outliersHD

RemoteRef HEAD

RemoteSha ae50458f6ad19581d73ef853b9834ff49f6f62a1

Contents

`outliersHD-package` *Detection of Outliers in High Dimensional Data*

Description

Algorithms to detect high-dimensional outliers. The minimum diagonal product (MDP) of Ro, Zou, Wang and Yin (2015), the algorithm of Wilkinson that relies on nearest neighbours, and the distances of distances (DOD of Lee and Jeon (2025).

Details

Package:	outliersHD
Type:	Package
Version:	1.0
Date:	2026-06-29

Maintainers

Michail Tsagris <mtsagris@uoc.gr>.

Author(s)

Michail Tsagris <mtsagris@uoc.gr>

References

Ro K., Zou C., Wang Z. and Yin G. (2015). Outlier detection for high-dimensional data. Biometrika, 102(3): 589–599.

Wilkinson L. (2018). Visualizing big data outliers through distributed aggregation. IEEE Transactions on Visualization and Computer Graphics 24(1): 256–266.

Tsagris M., Papadakis M., Alenazi A. and Alzeley O. (2024). Computationally Efficient Outlier Detection for High-Dimensional Data Using the MDP Algorithm. Computation, 12(9): 185.

Seong-ho Lee and Yongho Jeon (2025). DOD: Detection of outliers in high dimensional data with distance of distances. https://arxiv.org/abs/2511.02199

ahd *Detection of high dimensional outliers using nearest neighbours*

Description

Detection of high dimensional outliers using nearest neighbours.

Usage

```
ahd(x, a = 0.01, k = 10, p = 0.5, tn = 50)
```

Arguments

x	A matrix with numerical data with more columns (p) than rows (n), i.e. n<p.
a	Threshold for determining the cutoff for outliers. Observations are considered outliers if they fall in the (1-a) tail of the distribution of the nearest neighbor distances between exemplars.
k	The number of nearest neighbours to consider.
p	The proportion of possible outliers.
tn	Sample size to calculate an empirical threshold.

Details

For more information see Wilkinson (2018) and the R package "stray" that has implemented the algortihm. Our implementation is a faster (and slightly different) version of theirs.

Value

A list including:

scores	The score values of each observation.
outliers	The indices of the possible outlier(s).

Author(s)

Michail Tsagris.

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

References

Wilkinson L. (2018). Visualizing big data outliers through distributed aggregation. IEEE Transactions on Visualization and Computer Graphics 24(1): 256–266.

See Also

dod, rmdp

Examples

```
x <- matrix(rnorm(20 * 50), ncol = 50)
x <- rbind(x, matrix(rnorm(2 * 50, 5, 1), ncol = 50) )
a <- ahd(x)
```

---

dod *Detection of high dimensional outliers using DOD*

Description

Detection of high dimensional outliers using DOD.

Usage

```
dod(x, co = 0.1, a = 0.1)
```

Arguments

x	A matrix with numerical data with more columns (p) than rows (n), i.e. n<p.
co	This is to compute the c parameter ($c = co\sqrt{pn}$). In the paper co=0.1, and this is the default value in the function as well.
a	The parameter (a>0 and a<0.5) that represents the maximum proportion of outliers. It serves as a tuning parameter controling the maximum false positive rate.

Details

High dimensional outliers (n«p) are detected using distances of distances.

Value

A vector with the index of the detected outlier(s).

Author(s)

Michail Tsagris.

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

References

Seong-ho Lee and Yongho Jeon (2025). DOD: Detection of outliers in high dimensional data with distance of distances. https://arxiv.org/abs/2511.02199

See Also

rmdp, ahd

Examples

```
x <- matrix(rnorm(20 * 50), ncol = 50)
x <- rbind(x, matrix(rnorm(2 * 50, 5, 1), ncol = 50) )
a <- dod(x)
```

---

rmdp *Detection of high dimensional outliers using the RMDP*

Description

Detection of high dimensional outliers using the RMDP.

Usage

```
rmdp(x, alpha = 0.05, itertime = 100, parallel = FALSE)
```

Arguments

x	A matrix with numerical data with more columns (p) than rows (n), i.e. n<p.
alpha	The significance level, i.e. used to decide whether an observation is said to be considered a possible outlier. The default value is 0.05.
itertime	The number of iterations the algorithm will be ran. The higher the sample size, the larger this number must be. With 50 observations in R^1000 maybe this has to be 1000 in order to produce stable results.
parallel	A logical value for parallel version.

Details

High dimensional outliers (n«p) are detected using a properly constructed MCD. The variances of the variables are used and the determinant is simply their product.

Value

A list including:

runtime	The duration of the process.
dis	The final estimated Mahalanobis type normalised distances.
wei	A bollean variable vector specifying whether an observation is "clean" (TRUE) or a possible outlier (FALSE).

Author(s)

Michail Tsagris.

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

References

Ro K., Zou C., Wang Z. and Yin G. (2015). Outlier detection for high-dimensional data. Biometrika, 102(3): 589–599.

Tsagris M., Papadakis M., Alenazi A. and Alzeley O. (2024). Computationally Efficient Outlier Detection for High-Dimensional Data Using the MDP Algorithm. Computation, 12(9): 185.

See Also

dod, ahd

Examples

```
x <- matrix(rnorm(20 * 50), ncol = 50)
x <- rbind(x, matrix(rnorm(2 * 50, 5, 1), ncol = 50) )
a <- rmdp(x, itertime = 5)
```

Index