

Package ‘truh’

July 22, 2025

Title Two-Sample Nonparametric Testing Under Heterogeneity

Version 1.0.0

Description Implements the TRUH test statistic for two sample testing under heterogeneity. TRUH incorporates the underlying heterogeneity and imbalance in the samples, and provides a conservative test for the composite null hypothesis that the two samples arise from the same mixture distribution but may differ with respect to the mixing weights. See Trambak Banerjee, Bhaswar B. Bhattacharya, Gourab Mukherjee Ann. Appl. Stat. 14(4): 1777-1805 (December 2020). <[DOI:10.1214/20-AOAS1362](https://doi.org/10.1214/20-AOAS1362)> for more details.

License GPL (>= 3)

Encoding UTF-8

URL <https://github.com/natesmith07/truh>

Imports Rfast, cluster, doParallel, foreach, iterators, fpc, parallel

RoxygenNote 7.1.1

Suggests rmarkdown, knitr

VignetteBuilder knitr

NeedsCompilation no

Author Nathan Smith [aut, cre],
Trambak Banerjee [aut],
Bhaswar Bhattacharya [aut],
Gourab Mukherjee [aut]

Maintainer Nathan Smith <nathan_smith_99@ku.edu>

Repository CRAN

Date/Publication 2021-09-08 08:00:02 UTC

Contents

nearest	2
truh	3
Index	5

nearest

Nearest neighbor computation for the TRUH statistic

Description

For a given d dimensional vector \mathbf{y} , this function finds the nearest neighbor of \mathbf{y} in a $n \times d$ matrix \mathbf{U} .

Usage

```
nearest(y, U, n, d)
```

Arguments

\mathbf{y}	a d dimensional vector.
\mathbf{U}	a $n \times d$ matrix where n represents the sample size and d is the dimension of each sample.
n	the sample size.
d	dimension of each sample.

Value

1. d1 - nearest neighbor of \mathbf{y} in \mathbf{U}
2. d2 - nearest neighbor of d1 in \mathbf{U}

See Also

[truh](#)

Examples

```
library(truh)
n = 100
d = 3
set.seed(1)
y = rnorm(3)
set.seed(2)
U = matrix(rnorm(n*d),nrow=n,ncol=d)
out = nearest(y,U,n,d)
```

truh	<i>TRUH test statistic</i>
------	----------------------------

Description

TRUH test statistic for nonparametric two sample testing under heterogeneity.

Usage

```
truh(V, U, B, fc = 1, ncores = 2, seed = 1)
```

Arguments

V	$m \times d$ matrix where m represents the sample size and d is the dimension of each sample.
U	a $n \times d$ matrix where n represents the sample size and d is the dimension of each sample with $m \ll n$.
B	number of bootstrap samples.
fc	fold change constant. The default value is 1. See equation (2.8) of the referenced paper for more details.
ncores	the number of computing cores available. The default value is 2.
seed	random seed for replicability. The default value is 1.

Value

1. teststat - TRUH test statistic.
2. k.hat - number of clusters detected in the uninfected sample.
3. pval - The maximum p-value across the detected clusters.
4. pval_all - p-value for each cluster.
5. dist.null_all - the approximate bootstrapped based null distribution.

References

Banerjee, Trambak, Bhaswar B. Bhattacharya, and Gourab Mukherjee. "A nearest-neighbor based nonparametric test for viral remodeling in heterogeneous single-cell proteomic data." *The Annals of Applied Statistics* 14, no. 4 (2020): 1777-1805.

See Also

[nearest](#)

Examples

```
library(truh)
n = 500
m = 10
d = 3
set.seed(1)
V = matrix(rnorm(m*d),nrow=m,ncol=d)
set.seed(2)
U = matrix(rnorm(n*d),nrow=n,ncol=d)
out = truh(V,U,100)
```

Index

nearest, [2](#), [3](#)

truh, [2](#), [3](#)