

Package ‘fasthplus’

October 13, 2022

Title Fast Label-Dissimilarity Discordance Estimation with H+

Version 1.0

Description Estimation procedures for assessing fitness of observation labels (i.e., clusters or partitions) given observation dissimilarities, or vice versa. The estimated parameter of interest is modified from G+ (Williams 1971), so we call it H+.

License CC BY 4.0

Encoding UTF-8

RoxygenNote 7.1.2

BugReports <https://github.com/ntdyjack/fasthplus/>

URL <https://github.com/ntdyjack/fasthplus/>

Imports stats

Suggests knitr, rmarkdown, clusterCrit

VignetteBuilder knitr

NeedsCompilation no

Author Nathan Dyjack [aut, cre] (<<https://orcid.org/0000-0001-6944-6252>>),
Stephanie C. Hicks [aut, ctb] (<<https://orcid.org/0000-0002-7858-0231>>)

Maintainer Nathan Dyjack <ntdyjack@gmail.com>

Repository CRAN

Date/Publication 2022-02-07 09:20:05 UTC

R topics documented:

hpb	2
hpe	2
Index	4

hpb *H+ discordance estimation with bootstrapping*

Description

Estimates the H+ discordance metric given data `Dat` and a label vector `L`. Bootstrapping is performed over `r` iterations with `s` points sampled per iterate.

Usage

```
hpb(D, L, r = 30, t)
```

Arguments

<code>D</code>	numeric matrix or data frame with observations in rows (<code>nxm</code>)
<code>L</code>	numeric vector containing a set of length <code>n</code>
<code>r</code>	numeric number of bootstrap iterations
<code>t</code>	numeric pre-bootstrap sample size, 1-5% of <code>n</code>

Value

list, `h` is the estimated H+ value.
`gamma1` and `gamma2` are plausible ranges for what % of A (or Dw)
are strictly greater than B (or Db)

Examples

```
a <- sapply(1:500, function(i) rnorm(n=50, mean=0))
b <- sapply(1:500, function(i) rnorm(n=50, mean=0))
x <- t(cbind(a,b))
l <- c(rep(0,500), rep(1,500))
h <- hpb(D=x, L=l, r=30, t=50)
```

hpe *H+ discordance metric*

Description

Estimates the H+ discordance metric for either (1) two sets (vectors) A and B, or (2) a dissimilarity matrix D and a label vector L. Approximation is calculated using `p+1` percentiles, with an accuracy bound of `1/p`.

Usage

```
hpe(A, B, D, L, p = 101, alg = "brute_force", alpha = F, gammas = F)
```

Arguments

A	numeric vector containing a set of length n
B	numeric vector containing a set of length n
D	distance matrix of dimension nxn
L	numeric or character vector of length n
p	integer representing the number of percentiles
alg	character string ("brute_force" or "grid_search") representing the choice of algorithm used to estimate H+
alpha	logical indicator to return alpha values that parameterize balance of within/between cluster distances
gammas	logical indicator to return estimate for gamma values that parameterize what %Dw is greater than a second %Db

Value

h is the estimated H+ value.

(optional) aw and ab (alphaW and alphaB) are (respectively) the portion of within- and between-cluster distances (or portional sizes of A and B)

(optional) gw and gb (gammaW and gammaB) are plausible ranges for gw100% of Dw (or A) are strictly greater than gw100% Db (or B)

Examples

```
a <- rnorm(n=500, mean=0)
b <- rnorm(n=500, mean=1)
h <- hpe(A=a, B=b, p=101, alg="brute_force")

a <- sapply(1:500, function(i) rnorm(n=50, mean=0))
b <- sapply(1:500, function(i) rnorm(n=50, mean=0))
x <- cbind(a,b)
d <- dist(t(x))
l <- c(rep(0,500), rep(1,500))
h <- hpe(D=d, L=l, p=101, alg="brute_force")
```

Index

hpb, 2
hpe, 2