

Package ‘highTtest’

January 12, 2022

Type Package

Title Simultaneous Critical Values for t-Tests in Very High Dimensions

Version 1.2

Date 2022-01-12

Description Implements the method developed by Cao and Kosorok (2011) for the significance analysis of thousands of features in high-dimensional biological studies. It is an asymptotically valid data-driven procedure to find critical values for rejection regions controlling the k-familywise error rate, false discovery rate, and the tail probability of false discovery proportion.

License GPL-2

Depends methods

Suggests colorfulVennPlot

Author Hongyuan Cao [aut],
Michael Kosorok [aut],
Shannon Holloway [aut, cre]

Maintainer Shannon Holloway <sthollow@ncsu.edu>

NeedsCompilation no

Repository CRAN

Date/Publication 2022-01-12 22:42:41 UTC

R topics documented:

highTtest	2
highTtest-class	4
plot-methods	6
vennD-methods	6

Index	7
--------------	----------

highTtest

*Simultaneous critical values for t-tests in very high dimensions***Description**

Implements the method developed by Cao and Kosorok (2011) for the significance analysis of thousands of features in high-dimensional biological studies. It is an asymptotically valid data-driven procedure to find critical values for rejection regions controlling the k-familywise error rate, false discovery rate, and the tail probability of false discovery proportion.

Usage

```
highTtest(dataSet1, dataSet2, gammas, compare = "BOTH", cSequence = NULL,
tSequence = NULL)
```

Arguments

dataSet1	data.frame or matrix containing the dataset for subset 1 for the two-sample t-test.
dataSet2	data.frame or matrix containing the dataset for subset 2 for the two-sample t-test.
gammas	vector of significance levels at which feature significance is to be determined.
compare	one of ("ST", "BH", "Both", "None"). In addition to the Cao-Kosorok method, obtain feature significance indicators using the Storey-Tibshirani method (ST) (Storey and Tibshirani, 2003), the Benjamini-Hochberg method (BH), (Benjamini and Hochberg, 1995), "both" the ST and the BH methods, or do not consider alternative methods (none).
cSequence	A vector specifying the values of c to be considered in estimating the proportion of alternative hypotheses. If no vector is provided, a default of seq(0.01,6,0.01) is used. See Section 2.3 of Cao and Kosorok (2011) for more information.
tSequence	A vector specifying the search space for the critical t value. If no vector is provided, a default of seq(0.01,6,0.01) is used.

Details

The Storey-Tibshirani (2003), ST, method implemented in **highTtest** is adapted from the implementation written by Alan Dabney and John D. Storey and available from

<http://www.bioconductor.org/packages/release/bioc/html/qvalue.html>.

The comparison capability is included only for convenience and reproducibility of the original manuscript. For a complete analysis based on the ST method, the user is referred to the **qvalue** package available through the bioconductor archive.

The following methods retrieve individual results from a highTtest object, x:

BH(x): Retrieves a matrix of logical values. The rows correspond to features, the columns to levels of significance. Matrix elements are TRUE if feature was determined to be significant by the Benjamini-Hochberg (1995) method.

`CK(x)`: Retrieves a matrix of logical values. The rows correspond to features, the columns to levels of significance. Matrix elements are TRUE if feature was determined to be significant by the Cao-Kosorok (2011) method.

`pi_alt(x)`: Retrieves the estimated proportion of alternative hypotheses obtained by the Cao-Kosorok (2011) method.

`pvalue(x)`: Retrieves the vector of p-values calculated using the two-sample t-statistic.

`ST(x)`: Retrieves a matrix of logical values. The rows correspond to features, the columns to levels of significance. Matrix elements are TRUE if feature was determined to be significant by the Storey-Tibshirani (2003) method.

A simple x-y plot comparing the number of significant features as a function of the level significance level can be generated using

`plot(x, ...)`: Generates a plot of the number of significant features as a function of the level of significance as calculated for each method (CK, BH, and/or ST). Additional plot controls can be passed through the ellipsis.

When comparisons to the ST and BH methods are requested, Venn diagrams can be generated using provided that package `colorfulVennPlot` is installed.

`vennD(x, gamma, ...)`: Generates two- and three-dimensional Venn diagrams comparing the features selected by each method. Implements methods of package **colorfulVennPlot**. In addition to the `highTtest` object, the level of significance, `gamma`, must also be provided. Most control argument of the **colorfulVennPlot** package can be passed through the ellipsis.

Value

Returns an object of class `highTtest`.

Author(s)

Authors: Hongyuan Cao, Michael R. Kosorok, and Shannon T. Holloway <sthollow@ncsu.edu>
Maintainer: Shannon T. Holloway <sthollow@ncsu.edu>

References

Cao, H. and Kosorok, M. R. (2011). Simultaneous critical values for t-tests in very high dimensions. *Bernoulli*, 17, 347–394. PMID: PMC3092179.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57, 289–300.

Storey, J. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences, USA*, 100, 9440–9445.

Examples

```
set.seed(123)
x1 <- matrix(c(runif(500), runif(500, 0.25, 1)), nrow=100)
obj <- highTtest(dataSet1=x1[, 1:5],
                dataSet2=x1[, 6:10],
                gammas=seq(0.1, 1, 0.1),
                tSequence=seq(0.001, 3, 0.001))
```

```

#Print number of significant features identified in each method
colSums(CK(obj))
colSums(ST(obj))
colSums(BH(obj))

#Plot the number of significant features identified in each method
plot(obj, main="Example plot")
ltry <- try(library(colorfulVennPlot),silent=TRUE)

if( !is(ltry,"try-error") ) vennD(obj, 0.8, Title="Example vennD")

#Proportion of alternative hypotheses
pi_alt(obj)

#p-values
pvalue(obj)

```

highTtest-class	Class "highTtest"
-----------------	-------------------

Description

Value object returned by call to `highTtest()`.

Objects from the Class

This object should not be created by users.

Slots

CK: Object of class `matrix` or `NULL`. A matrix of logical values. The rows correspond to features, ordered as provided in input `dataSet1`. The columns correspond to levels of significance. Matrix elements are `TRUE` if feature was determined to be significant by the Cao-Kosorok method. The significance value associated with each column is dictated by the input `gammas`.

pi1: Object of class `numeric` or `NULL`. The estimated proportion of alternative hypotheses calculated using the Cao-Kosorok method.

pvalue: Object of class `numeric`. The vector of p-values calculated using the two-sample t-statistic.

ST: Object of class `matrix` or `NULL`. If requested, a matrix of logical values. The rows correspond to features, ordered as provided in input `dataSet1`. The columns correspond to levels of significance. Matrix elements are `TRUE` if feature was determined to be significant by the Storey-Tibshirani (2003) method. The significance value associated with each column is dictated by the input `gammas`.

BH: Object of class `matrix` or `NULL`. If requested, A matrix of logical values. The rows correspond to features, ordered as provided in input `dataSet1`. The columns correspond to levels of significance. Matrix elements are `TRUE` if feature was determined to be significant by the Benjamini-Hochberg (1995) method. The significance value associated with each column is dictated by the input `gammas`.

gammas: Object of class `numeric`. Vector of significant values provided as input for the calculation.

Methods

BH signature(`x = "highTtest"`): Retrieves a matrix of logical values. The rows correspond to features, the columns to levels of significance. Matrix elements are `TRUE` if feature was determined to be significant by the Benjamini-Hochberg (1995) method.

CK signature(`x = "highTtest"`): Retrieves a matrix of logical values. The rows correspond to features, the columns to levels of significance. Matrix elements are `TRUE` if feature was determined to be significant by the Cao-Kosorok (2011) method.

pi_alt signature(`x = "highTtest"`): Retrieves the estimated proportion of alternative hypotheses obtained by the Cao-Kosorok (2011) method.

plot signature(`x = "highTtest"`): Generates a plot of the number of significant features as a function of the level of significance as calculated for each method (CK, BH, and/or ST)

pvalue signature(`x = "highTtest"`): Retrieves the vector of p-values calculated using the two-sample t-statistic.

ST signature(`x = "highTtest"`): Retrieves a matrix of logical values. The rows correspond to features, the columns to levels of significance. Matrix elements are `TRUE` if feature was determined to be significant by the Storey-Tibshirani (2003) method.

vennD signature(`x = "highTtest"`): Generates two- and three-dimensional Venn diagrams comparing the features selected by each method. Implements methods of package **colorfulVennPlot**. In addition to the `highTtest` object, the level of significance, `gamma`, must also be provided.

Author(s)

Authors: Hongyuan Cao, Michael R. Kosorok, and Shannon T. Holloway <sthollow@ncsu.edu>
Maintainer: Shannon T. Holloway <sthollow@ncsu.edu>

References

Cao, H. and Kosorok, M. R. (2011). Simultaneous critical values for t-tests in very high dimensions. *Bernoulli*, 17, 347–394. PMID: PMC3092179.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57, 289–300.

Storey, J. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences, USA*, 100, 9440–9445.

Examples

```
showClass("highTtest")
```

plot-methods

~~ *Methods for Function plot* ~~

Description

Generates a simple x-y plot giving the number of significant features as a function of the level of significance. If comparisons to Storey-Tibshirani and Benjamini-Hochberg methods were requested by the user, these will automatically be included in the plot.

Methods

signature(x = "ANY") Plot method as implemented by other packages.

signature(x = "highTtest") Object returned by a call to highTtest().

vennD-methods

~~ *Methods for Function vennD* ~~

Description

Generates 2- or 3-dimensional Venn diagrams comparing the features selected by the Cao-Kosorok method to those selected by the Storey-Tibshirani (2003) method and/or the Benjamini-Hoshberg (1995) method. This S4 method is simply a wrapper for plotVenn2d() and plotVenn3d() of package **colorfulVennPlot**.

Methods

signature(x = "highTtest", gamma="numeric", ...) Object returned by a call to highTtest(). gamma is the level of significance. Additional control variables for the methods of plotVenn2d() and plotVenn3d() of package **colorfulVennPlot** can be passed through the ellipsis.

Index

* classes

highTtest-class, 4

* methods

plot-methods, 6

vennD-methods, 6

BH (highTtest), 2

BH, highTtest (highTtest), 2

BH, highTtest-method (highTtest-class), 4

CK (highTtest), 2

CK, highTtest (highTtest), 2

CK, highTtest-method (highTtest-class), 4

highTtest, 2

highTtest-class, 4

pi_alt (highTtest), 2

pi_alt, highTtest (highTtest), 2

pi_alt, highTtest-method
(highTtest-class), 4

plot, ANY-method (plot-methods), 6

plot, highTtest (highTtest), 2

plot, highTtest-method (plot-methods), 6

plot-methods, 6

pvalue (highTtest), 2

pvalue, highTtest (highTtest), 2

pvalue, highTtest-method
(highTtest-class), 4

ST (highTtest), 2

ST, highTtest (highTtest), 2

ST, highTtest-method (highTtest-class), 4

vennD (vennD-methods), 6

vennD, highTtest (highTtest), 2

vennD, highTtest, numeric-method
(vennD-methods), 6

vennD-methods, 6