Package 'compHclust'

May 4, 2017

Version 1.0-3

Date 2017-05-04

Title Complementary Hierarchical Clustering

Imports graphics, stats

Description Performs the complementary hierarchical clustering procedure and returns X' (the expected residual matrix) and a vector of the relative gene importances.

License GPL (>= 2)

NeedsCompilation yes

Author Gen Nowak [aut, cre],
Robert Tibshirani [aut]

Maintainer Gen Nowak <gen.nowak@gmail.com>

Repository CRAN

Date/Publication 2017-05-04 07:14:50 UTC

R topics documented:

---

compHclust   *Complementary Hierarchical Clustering*

Description

Performs the complementary hierarchical clustering procedure and returns X' (the expected residual matrix) and a vector of the relative gene importances.

Usage

```
compHclust(x, xhc)
```

1

Arguments

x   A numeric matrix X, where interest lies in clustering its columns.

xhc   An object of class `hclust`, specifically, a hierarchical clustering of the columns of X.

Details

This function performs the complementary hierarchical clustering procedure, as described in Nowak and Tibshirani (2008). Although applicable to any numeric matrix X, we typically think of X as microarray data with the rows as genes and the columns as samples, with the number of genes much greater than the number of samples. The goal of the procedure is to uncover structures present in the data that arise from 'weak' genes.

Given X and a hierarchical clustering of the columns of X, the function returns X', which represents a modified version of X with the structural features arising from the strong genes removed. Using information present in the hierarchical clustering, we perform a series of linear regressions and set X' to be the expected value of the resulting residuals. Details are given in Nowak and Tibshirani (2008). The user can then apply a hierarchical clustering algorithm to cluster the columns of X' to discover any important structures arising from the weaker genes.

The function also returns a vector of length equal to the number of rows of X, where the *i*th element is equal to the relative gene importance of the *i*th gene. The relative gene importance lies between 0 and 1, with a value close to 1 indicating that a gene (row) was strongly influential in the hierarchical clustering of the columns of X.

Value

A list with components:

x.prime   The expected residual matrix X'.

gene.imp   A vector of the relative gene importances.

Author(s)

Gen Nowak <gen.nowak@gmail.com> and Robert Tibshirani

References

Nowak, G. and Tibshirani, R. (2008) Complementary hierarchical clustering. *Biostatistics*, **9**(3), 467–483.

See Also

hclust

Examples

```
## Creating example microarray data with rows as genes and columns as
## samples.  Rows 1-20 represent the 'strong' genes which differentiate
## samples 1-4 from samples 5-8.  Rows 31-50 represent the 'weak' genes
## which differentiate the odd numbered samples from the even numbered
```

```
## samples.
set.seed(4872)
x <- matrix(0,nrow=50,ncol=8)
x[1:20,1:4] <- 8
x[1:20,5:8] <- -8
x[31:50,c(1,3,5,7)] <- 4
x[31:50,-c(1,3,5,7)] <- -4
x <- x + matrix(rnorm(50*8),ncol=8)

## Hierarchically cluster the columns of x.
x.hc <- hclust(as.dist(1-cor(x)))

## Run complementary hierarchical clustering.
x.chc <- compHclust(x,x.hc)
xp <- x.chc$x.prime
x.gi <- x.chc$gene.imp

## Hierarchically cluster the columns of x'.
xp.hc <- hclust(as.dist(1-cor(xp)))
xp.gi <- compHclust(xp,xp.hc)$gene.imp

## We use the function 'compHclust.heatmap' to display the
## initial and complementary clusterings.
## The initial clustering.
compHclust.heatmap(x,x.hc,x.gi,d.title="Initial Clustering")
## The complementary clustering.
compHclust.heatmap(xp,xp.hc,xp.gi,d.title="Complementary Clustering")
```

---

compHclust.heatmap   *Heat Map for Complementary Hierarchical Clustering*

Description

Displays a heat map of X, a dendrogram of the clustering of the columns of X and a bar plot of the relative gene importances.

Usage

```
compHclust.heatmap(x, xhc, gi, d.title = "Cluster Dendrogram",
                   hm.lab = TRUE, hm.lab.cex = 1, d.ht = 0.25,
                   gi.width = 0.5, d.mar = c(0, 4, 4, 2),
                   hm.mar = c(5, 4, 2, 2))
```

Arguments

x	A numeric matrix X, where interest lies in clustering its columns.
xhc	An object of class hclust, specifically, a hierarchical clustering of the columns of X.
gi	A vector of the relative gene importances, as returned by compHclust.

d.title	The title for the dendrogram.
hm.lab	Logical. If true, the columns of the heat map are labeled with column numbers.
hm.lab.cex	The magnification to be used for the column labels relative to the current setting of cex. See [axis](#) and [par](#).
d.ht	The relative height of the plotting region for the dendrogram. Note that the relative height of the plotting region for the heat map is set to 1. See [layout](#).
gi.width	The relative width of the plotting region for the relative gene importance plot. Note that the relative width of the plotting region for the heat map is set to 1. See [layout](#).
d.mar	The margins of the plotting region for the dendrogram. See [par](#).
hm.mar	The margins of the plotting region for the heat map. See [par](#).

Details

Given a numeric matrix X, a hierarchical clustering of the columns of X and a vector of the relative gene importances as returned by [compHclust](#), this function displays a heat map of X with a dendrogram above and a bar plot of the relative gene importances to the right. The columns of X are reordered to correspond with the leaves of the dendrogram.

This function can be fragile - depending on the dimensions of X, some of the arguments such as the margins, heights and widths of the plotting regions may need to be tweaked in order for the figure to look nice. However, it provides a quick and easy way of displaying the output of [compHclust](#) and seeing which genes (rows) may be most influential in the clustering of the samples (columns).

For examples of its usage, see the help file for [compHclust](#).

Author(s)

Gen Nowak <gen.nowak@gmail.com> and Robert Tibshirani

References

Nowak, G. and Tibshirani, R. (2008) Complementary hierarchical clustering. *Biostatistics*, **9**(3), 467–483.

See Also

[compHclust](#)

Index