Package ‘bioDist’

April 5, 2024

Title Different distance measures
Version 1.74.0
Author B. Ding, R. Gentleman and Vincent Carey
Description A collection of software tools for calculating distance measures.
Maintainer Bioconductor Package Maintainer
   <maintainer@bioconductor.org>
License Artistic-2.0
Depends R (>= 2.0), methods, Biobase,KernSmooth
Suggests locfit
biocViews Clustering, Classification
git_url https://git.bioconductor.org/packages/bioDist
git_branch RELEASE_3_18
git_last_commit 89a00b0
git_last_commit_date 2023-10-24
Repository Bioconductor 3.18
Date/Publication 2024-04-05

R topics documented:

  closest.top ......................................................... 2
  cor.dist ............................................................. 3
  euc ................................................................. 4
  KLD.matrix ......................................................... 5
  KLDist.matrix ...................................................... 6
  man ................................................................. 7
  mutualInfo .......................................................... 8
  spearman.dist ..................................................... 9
  tau.dist ............................................................ 10

Index 12
closest.top

Find the closest genes.

Description

Find the closest genes to the supplied target gene based on the supplied distances.

Usage

closest.top(x, dist.mat, top)

Arguments

x the name of the gene (feature) to use.
dist.mat either a dist object or a matrix of distances.
top the number of closest genes desired.

Details

The feature named x must be in the supplied distances. If so, then the top closest other features are returned.

Value

A vector of names of the top closest features.

Author(s)

Beiying Ding

See Also

cor.dist, spearman.dist, tau.dist.euc, man.KLdist.matrix, KLD.matrix, mutualInfo

Examples

data(sample.ExpressionSet)
sE <- sample.ExpressionSet[1:100,]
d1 <- KLdist.matrix(sE, sample = FALSE)
closest.top(featureNames(sE)[1], d1, 5)
Description
Calculate pairwise Pearson correlational distances, i.e. 1-COR or 1-|COR|, and saves as a 'dist' object.

Usage
```r
cor.dist(x, ...)
```

Arguments
- `x`: n by p matrix or ExpressionSet; if x is an ExpressionSet, then the function uses its 'exprs' slot.
- `...`: arguments passed to `cor.dist`:
  - `absif`: TRUE, then 1-|COR| else 1-COR, default is TRUE.
  - `diagif`: TRUE, then the diagonal of the distance matrix will be displayed, default is FALSE.
  - `upperif`: TRUE, then the upper triangle of the distance matrix will be displayed, default is FALSE.
  - `sample`: for objects of classes that extend eSet: if TRUE, then distances are computed between samples(columns), otherwise, they are computed between features(rows).

Details
The `cor` function is used to compute the pairwise distances between rows of an input matrix, except if the input is an object of a class that extends eSet and `sample` is TRUE.

Value
Pairwise Pearson correlational distance object

Author(s)
Beiying Ding

See Also
- `spearman.dist`, `tau.dist`, `euc`, `man`, `KLdist.matrix`, `KLD.matrix`, `mutualInfo`

Examples
```r
x <- matrix(rnorm(200), nrow = 5)
cor.dist(x)
```
Description
Calculate pairwise Euclidean distances and saves the result as a 'dist' object

Usage
euc(x, ...)

Arguments
x
n by p matrix or an object of a class that extends cSet; if x is a matrix, pairwise
distances are calculated between the rows of a matrix. If x is an object of a class
that extends cSet, the method makes use of the 'exprs' method and pairwise
distances are calculated between samples(columns) if sample is TRUE

... arguments passed to euc:
• diag if TRUE, then the diagonal of the distance matrix will be displayed;
default is FALSE.
• upper if TRUE, then the upper triangle of the distance matrix will be dis-
  played; default is FALSE.
• sampleFor objects of classes that extends cSet, pairwise distances are cal-
  culated between samples(columns) if sample is TRUE; default value is
  TRUE

Details
The method calculates pairwise euclidean distances, assuming that all samples have the same num-
ber of observations

Value
An object of class dist with the pairwise Euclidean distance between rows except in case of objects
of class that extend cSet when sample is TRUE

Author(s)
Beiying Ding

See Also
spearman.dist, tau.dist, man.KLdist.matrix, KLD.matrix, mutualInfo

Examples
x <- matrix(rnorm(200), nrow = 5)
euc(x)
KLD.matrix

Continuous version of Kullback-Leibler Distance (KLD)

Description

Calculate KLD by estimating by smoothing \( \log(f(x)/g(x)) \cdot f(x) \) and then integrating.

Usage

KLD.matrix(x, ...)

Arguments

x

n by p matrix or list or an object of a class that extends eSet; if x is an an object of a class that extends eSet (eg ExpressionSet), then the function works against its 'exprs' slot.

... arguments passed to KLD.matrix:

• method use locfit or density to estimate integrand; default is c("locfit", "density") (i.e. both methods).
• suppupper and lower limits of the integral; default is NULL in which case the limits of the integral are calculated from the range of the data.
• subdivisions subdivisions for the integration; default is 1000.
• diagif TRUE, then the diagonal of the distance matrix will be displayed; default is FALSE.
• upperif TRUE, then the upper triangle of the distance matrix will be displayed; default is FALSE.
• samplefor ExpressionSet methods: if TRUE, then distances are computed between samples, otherwise, they are computed between genes.

Details

The distance is computed between rows of the input matrix (except if the input is an object of a class that extends eSet and sample is TRUE.

The presumption is that all samples have the same number of observations. The list method is meant for use when samples sizes are unequal.

Value

An object of class dist with the pairwise, between rows, Kullback-Leibler distances.

Author(s)

Beiying Ding, Vincent Carey

See Also

cor.dist, spearman.dist, tau.dist, dist, KLdist.matrix, mutualInfo
Examples

x <- matrix(rnorm(100), nrow = 5)
KLD.matrix(x, method = "locfit", supp = range(x))

Description

Calculate the KLD by binning continuous data.

KL distance is calculated using the formula

$$KLD(f_1(x), f_2(x)) = \sum_{i=1}^{N} f_1(x_i) \log \frac{f_1(x_i)}{f_2(x_i)}$$

Usage

KLD.matrix(x, ...)

Arguments

x
n by p matrix or a list or an object of a class that extends eSet. If x is an object of
a class derived from eSet (ExpressionSet, SnpSet etc), then the values returned
by the exprs function are used.

... arguments passed to KLD.matrix:

gridsize the number of grid points used to select the optimal bin width of the
histogram used to estimate density. If no value is supplied, the grid size is
calculated internally; default is NULL.
symmetrize if TRUE, then symmetrize; the default is FALSE.
diag if TRUE, then the diagonal of the distance matrix will be displayed; the
default is FALSE.
upper if TRUE, then the upper triangle of the distance matrix will be displayed;
default is FALSE.
sample for eSet methods: if TRUE, then the distances are computed between
samples, otherwise, between features; the default is TRUE.

Details

The data are binned, and then the KL distance between the two discrete distributions is computed
and used. The distance is computed between rows of the input matrix (except if the input is an
object of a class that extends eSet and sample is TRUE.

The presumption is that all samples have the same number of observations. The list method is meant
for use when samples sizes are unequal.
Description
Calculate pairwise Manhattan distances and saves as a dist object.

Usage
man(x, ...)

Arguments
x n by p matrix or an object of class that extends eSet. If x is an object of class that extends eSet, (eg ExpressionSet) then the function uses its 'exprs' slot.

... arguments passed to man:
• diagif TRUE, then the diagonal of the distance matrix will be displayed; default is FALSE.
• upperif TRUE, then the upper triangle of the distance matrix will be displayed; default is FALSE.

Details
This is just an interface to dist with the right parameters set.

Value
An instance of the dist class with the pairwise Manhattan distances between the rows of x in case of a matrix or between the features (rows) in case of a class that extends eSet.

Examples
x <- matrix(rnorm(100), nrow = 5)
KLdist.matrix(x, symmetrize = TRUE)
Author(s)
Beiying Ding

See Also
cor.dist, spearman.dist, tau.dist.euc, KLdist.matrix, KLD.matrix, mutualInfo

Examples
x <- matrix(rnorm(200), nrow = 5)
man(x)

mutualInfo
Mutual Information

Description
Calculate mutual information via binning

Usage
mutualInfo(x, ...)
MIdist(x, ...)

Arguments
x an n by p matrix or ExpressionSet; if x is an ExpressionSet, then the function uses its 'exprs' slot.
...
arguments passed to mutualInfo and MIdist:

• nbins number of bins to calculate discrete probabilities; default is 10.
• diagif TRUE, then the diagonal of the distance matrix will be displayed; default is FALSE.
• upperif TRUE, then the upper triangle of the distance matrix will be displayed; default is FALSE.
• samplefor ExpressionSet methods, if TRUE, then distances are computed between samples, otherwise, between genes.

Details
For mutualInfo each row of x is divided into nbins groups and then the mutual information is computed, treating the data as if they were discrete.

For MIdist we use the transformation proposed by Joe (1989), \( \delta^* = (1 - \exp(-2\delta))^{1/2} \) where \( \delta \) is the mutual information. The MIdist is then \( 1 = \delta^* \). Joe argues that this measure is then similar to Kendall’s tau, tau.dist.
Value

An object of class dist which contains the pairwise distances.

Author(s)

Robert Gentleman

References


See Also

dist, KLdist.matrix, cor.dist, KLD.matrix

Examples

x <- matrix(rnorm(100), nrow = 5)
mutualInfo(x, nbin = 3)

spearman.dist

Spearman correlational distance

Description

Calculate pairwise Spearman correlational distances, i.e. 1-SPEAR or 1-|SPEAR|, for all rows of a
matrix and return a dist object.

Usage

spearman.dist(x, ...)

Arguments

x n by p matrix or ExpressionSet; if x is an ExpressionSet, then the function uses its 'exprs' slot.

... arguments passed to spearman.dist:

  • absif TRUE, then 1-|SPEAR| else 1-SPEAR; default is TRUE.
  • diagif TRUE, then the diagonal of the distance matrix will be displayed; default is FALSE.
  • upperif TRUE, then the upper triangle of the distance matrix will be dis-
    played; default is FALSE.
  • samplefor the ExpressionSet method: if TRUE (the default), then distances
    are computed between samples.
Details

We call cor with the appropriate arguments to compute the row-wise correlations.

Value

One minus the Spearman correlation, between rows of x, are returned, as an instance of the dist class.

Author(s)

Beiying Ding

See Also

cor.dist, tau.dist, euc, man, KLdist.matrix, KLD.matrix, mutualInfo, dist

Examples

```r
x <- matrix(rnorm(200), nrow = 5)
spearman.dist(x)
```

tau.dist

Kendall’s tau correlational distance

Description

Calculate pairwise Kendall’s tau correlational distances, i.e. 1-TAU or 1-|TAU|, for all rows of the input matrix and return an instance of the dist class.

Usage

tau.dist(x, ...)

Arguments

- `x`: n by p matrix or ExpressionSet; if x is an ExpressionSet, then the function uses its 'exprs' slot.
- `...`: arguments passed to tau.dist:
  - `absif`: TRUE, then 1-TAU; else 1-|TAU|; default is TRUE.
  - `diagif`: TRUE, then the diagonal of the distance matrix will be displayed; default is FALSE.
  - `upperif`: TRUE, then the upper triangle of the distance matrix will be displayed; default is FALSE.
  - `samplefor`: the ExpressionSet method: if TRUE (the default), then distances are computed between samples.
Details
Row-wise correlations are computed by calling the `cor` function with the appropriate arguments.

Value
One minus the row-wise Kendall’s tau correlations are returned as an instance of the `dist` class. Note that this can be extremely slow for large data sets.

Author(s)
Beiying Ding

See Also
`cor.dist`, `spearman.dist`, `euc`, `man`, `Kldist.matrix`, `KLD.matrix`, `mutualInfo`

Examples
```r
x <- matrix(rnorm(200), nrow = 5)
tau.dist(x)
```
Index

* manip
  KLD.matrix, 5
  KLdist.matrix, 6
  man, 7
  mutualInfo, 8
  spearman.dist, 9
  tau.dist, 10

  closest.top, 2
  cor.dist, 2, 3, 5, 7–11
  cor.dist, eSet-method (cor.dist), 3
  cor.dist, matrix-method (cor.dist), 3

  dist, 5, 9, 10

  euc, 2, 3, 4, 7, 8, 10, 11
  euc, eSet-method (euc), 4
  euc, matrix-method (euc), 4

  KLD.matrix, 2–4, 5, 7–11
  KLD.matrix, eSet-method (KLD.matrix), 5
  KLD.matrix, list-method (KLD.matrix), 5
  KLD.matrix, matrix-method (KLD.matrix), 5
  KLdist.matrix, 6
  KLdist.matrix, 2–5, 8–11
  KLdist.matrix (KLdist.matrix), 6
  KLdist.matrix, eSet-method
  (KLdist.matrix), 6
  KLdist.matrix, list-method
  (KLdist.matrix), 6
  KLdist.matrix, matrix-method
  (KLdist.matrix), 6

  man, 2–4, 7, 7, 10, 11
  man, eSet-method (man), 7
  man, matrix-method (man), 7
  MIdist (mutualInfo), 8
  MIdist, ExpressionSet-method
  (mutualInfo), 8
  MIdist, matrix-method (mutualInfo), 8
  mutualInfo, 2–5, 7, 8, 8, 10, 11