bioDist
November 11, 2009

R topics documented:

closest.top ................................................................. 1
cor.dist ................................................................. 2
euc ................................................................. 3
KLdist.matrix ............................................................ 4
KLD.matrix ............................................................ 5
man ................................................................. 6
mutualInfo ............................................................ 7
spearman.dist .......................................................... 8
tau.dist .............................................................. 9

Index 10

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.
cor.dist

Author(s)
Beiying Ding

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

Examples

data(sample.ExpressionSet)
sE <- sample.ExpressionSet[1:100,]
d1 <- Kdist.matrix(sE, sample = FALSE)
closest.top(featureNames(sE)[1], d1, 5)

Description
Calculate pairwise Pearson correlational distances, i.e. 1-COR or 1-|COR|, for all columns of a matrix and save as a ‘dist’ object

Usage

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:

abs if TRUE, then 1-|COR| else 1-COR, default is TRUE.

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 displayed, default is FALSE.

sample for ExpressionSet methods: if TRUE, then distances are computed between samples, otherwise, they are computed between genes.

Details
The cor function is used to compute the pairwise distances.

Value
Pairwise Pearson correlational distance object

Author(s)
Beiying Ding
See Also

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

Examples

x <- matrix(rnorm(200), nrow = 5)
cor.dist(x)

euc(x, nrow = 5)
euc(x)

dist

Description

Calculate pairwise Euclidean distances for all columns of a matrix and save as a 'dist' object

Usage

euc(x, ...)

Arguments

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

Value

Pairwise Euclidean distance object

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)
KLdist.matrix  Discrete version of Kullback-Leibler Distance (KLD)

Description

Calculate the KLD by binning continuous data. KL distance is calculated using the formula

\[
KLD(f_1(x), f_2(x)) = \frac{1}{N} \sum_{i=1}^{N} \Delta_i \ast f_1(x_i) \ast \log \frac{f_1(x_i)}{f_2(x_i)}
\]

Usage

KLdist.matrix(x, ...)

Arguments

x  n by p matrix or ExpressionSet; if x is an ExpressionSet, then the values returned by the exprs function are used.
...
arguments passed to KLdist.matrix:
  nbin  number of bins to calculate discrete probabilities; default is 10.
symmetrize  if TRUE, then symmetrize; default is FALSE.
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 displayed; default is FALSE.
sample  for ExpressionSet methods: if TRUE, then the distances are computed between samples, otherwise, between genes.

Details

The data are binned, and then the KL distance between the two discrete distributions is computed and used.

The list method is meant for use when samples sizes are unequal.

Value

Pairwsie KLD object

Author(s)

Beiying Ding

See Also

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

Examples

x <- matrix(rnorm(100), nrow = 5)
KLdist.matrix(x, symmetrize = TRUE, nbin = 3)
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 ExpressionSet; if x is an 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).
supp upper and lower limits of the integral; default is c(-3, 3).
subdivisions subdivisions for the integration; default is 1000.
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 displayed; default is FALSE.
sample for ExpressionSet methods: if TRUE, then distances are computed between samples, otherwise, they are computed between genes.

Details
The pairwise distances between the rows of x are computed.
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))
man        

Description

Calculate pairwise Manhattan distances for all rows of a matrix and save as a dist object.

Usage

man(x, ...)

Arguments

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

... arguments passed to man:

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 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.

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:
nbin number of bins to calculate discrete probabilities; default is 10.
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 displayed; default is FALSE.
sample for ExpressionSet methods, if TRUE, then distances are computed between samples, otherwise, between genes.

Details
For mutualInfo each row of x is divided into nbin 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
spearman.dist

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:
            abs     if TRUE, then 1-SPEAR| else 1-SPEAR; default is TRUE.
            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 displayed; default is FALSE.
            sample  for 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

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**

```r
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`:
  - `abs` if TRUE, then 1-|TAU| else 1-TAU; default is TRUE.
  - `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 displayed; default is FALSE.
  - `sample` for 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

*Topic manip
  KLD.matrix, 4
  KLDist.matrix, 3
  man, 5
  mutualInfo, 6
  spearman.dist, 7
  tau.dist, 8
  closest.top, 1
  cor.dist, 1, 2, 4–9
  cor.dist, ExpressionSet-method (cor.dist), 2
  cor.dist, matrix-method (cor.dist), 2
  dist, 5, 7, 8
  euc, 1, 2, 3, 4, 6, 8, 9
  euc, ExpressionSet-method (euc), 3
  euc, matrix-method (euc), 3
  KLD.matrix, 1–3, 4, 6–9
  KLD.matrix, ExpressionSet-method (KLD.matrix), 4
  KLD.matrix, list-method (KLD.matrix), 4
  KLD.matrix, matrix-method (KLD.matrix), 4
  KLDist.matrix, 1, 2, 3, 5–9
  KLDist.matrix, ExpressionSet-method (KLDist.matrix), 3
  KLDist.matrix, list-method (KLDist.matrix), 3
  KLDist.matrix, matrix-method (KLDist.matrix), 3
  man, 1–4, 5, 8, 9
  man, ExpressionSet-method (man), 5
  man, matrix-method (man), 5
  MIdist (mutualInfo), 6
  MIdist, ExpressionSet-method (mutualInfo), 6
  MIdist, matrix-method (mutualInfo), 6
  mutualInfo, 1–5, 6, 6, 8, 9