

Package ‘fastGeneMI’

May 2, 2018

Type Package

Title A range of mutual information estimators for gene regulatory network inference from microarray expression data.

Version 1.0

Date 2017-02-08

Author Jonathan Ish-Horowicz

Maintainer Jonathan Ish-Horowicz <jonathan.ish-horowicz17@imperial.ac.uk>

Description Provides implementations of several mutual information estimators for use with the gene network inference algorithms CLR, ARACNE and MRNET.

License MIT

LazyData true

Imports Rcpp (>= 0.12.9), infotheo, PRROC, minet

Suggests Rdpack

LinkingTo Rcpp, RcppArmadillo

RoxygenNote 6.0.1

RdMacros Rdpack

R topics documented:

fastGeneMI-package	1
get.mim.bspline	2
get.mim.CS	3
get.mim.ML	3
get.mim.MM	4
get.mim.shrink	5
get.pr	5

fastGeneMI-package *A range of mutual information estimators for gene regulatory network inference from microarray expression data.*

Description

Provides implementations of several mutual information estimators for use with the gene network inference algorithms CLR, ARACNE and MRNET.

Details

The DESCRIPTION file: This package was not yet installed at build time.

Index: This package was not yet installed at build time.

Overview

Author(s)

Jonathan Ish-Horowicz

Maintainer: Jonathan Ish-Horowicz <jonathan.ish-horowicz17@imperial.ac.uk>

References

References

See Also

Links

get.mim.bspline *B-spline mutual information estimate*

Description

Returns a symmetric matrix of pairwise B-spline mutual information estimates from a samples x genes matrix of expression values. The ij^{th} element is the mutual information estimate between the expression of genes i and j in nats.

Usage

```
get.mim.bspline(expr.data, order, n.bins = as.integer(nrow(expr.data)^(1/3)),
  n.cores = 1L)
```

Arguments

<code>expr.data</code>	A samples x genes matrix of expression values.
<code>order</code>	The order of the B-spline used to smooth the histograms. This must be greater than 1 + the number of bins.
<code>n.bins</code>	The number of bins into which the expression data will be discretised. The default is $N^{1/3}$ for N samples.
<code>n.cores</code>	The number of cores to use for the computation. The default is 1.

Value

A symmetric matrix of pairwise mutual information between the expression of pairs of genes. These are the mutual information estimates between pairs of columns of `expr.data`.

<code>get.mim.CS</code>	<i>Chao-Shen Likelihood mutual information estimate</i>
-------------------------	---

Description

Returns a symmetric matrix of pairwise Chao-Shen mutual information estimates from a samples x genes matrix of expression values. The ij^{th} element is the mutual information estimate between the expression of genes i and j in nats.

Usage

```
get.mim.CS(expr.data, discretisation = c("equalwidth", "equalfreq", "bb"),
  n.bins = as.integer(nrow(expr.data)^(1/3)), n.cores = 1L)
```

Arguments

<code>expr.data</code>	A samples x genes matrix of expression values.
<code>discretisation</code>	The types of bins into which the expression data will be discretised. Can choose equal width bins ("equalwidth"), equal frequency bins ("equalfreq") or Bayesian Blocks bins ("bb")
<code>n.bins</code>	The number of bins into which the expression data will be discretised. The default is $N^{1/3}$ for N samples. Ignored if using Bayesian Blocks discretisation.
<code>n.cores</code>	The number of cores to use for the computation. The default is 1.

Value

A symmetric matrix of pairwise mutual information between the expression of pairs of genes. These are the mutual information estimates between pairs of columns of `expr.data`.

get.mim.ML	<i>Maximum Likelihood mutual information estimate</i>
------------	---

Description

Returns a symmetric matrix of pairwise Maximum Likelihood mutual information estimates from a samples x genes matrix of expression values. The ij^{th} element is the mutual information estimate between the expression of genes i and j .

Usage

```
get.mim.ML(expr.data, discretisation = c("equalwidth", "equalfreq", "bb"),
  n.bins = as.integer(nrow(expr.data)^(1/3)), n.cores = 1L)
```

Arguments

expr.data	A samples x genes matrix of expression values.
discretisation	The types of bins into which the expression data will be discretised. Can choose equal width bins ("equalwidth"), equal frequency bins ("equalfreq") or Bayesian Blocks bins ("bb")
n.bins	The number of bins into which the expression data will be discretised. The default is $N^{1/3}$ for N samples. Ignored if using Bayesian Blocks discretisation.
n.cores	The number of cores to use for the computation. The default is 1.

Value

A symmetric matrix of pairwise mutual information between the expression of pairs of genes. These are the mutual information estimates between pairs of columns of `expr.data`.

\@references Daub CO, Steuer R, Selbig J and Kloska S (2004). "Estimating mutual information using B-spline functions—an improved similarity measure for analysing gene expression data." *BMC Bioinformatics*, **5**(1), pp. 118.

get.mim.MM	<i>Miller-Madow mutual information estimate</i>
------------	---

Description

Returns a symmetric matrix of pairwise Miller-Madow mutual information estimates from a samples x genes matrix of expression values. The ij^{th} element is the mutual information estimate between the expression of genes i and j .

Usage

```
get.mim.MM(expr.data, discretisation = c("equalwidth", "equalfreq", "bb"),
  n.bins = as.integer(nrow(expr.data)^(1/3)), n.cores = 1L)
```

Arguments

<code>expr.data</code>	A samples x genes matrix of expression values.
<code>discretisation</code>	The types of bins into which the expression data will be discretised. Can choose equal width bins ("equalwidth"), equal frequency bins ("equalfreq") or Bayesian Blocks bins ("bb")
<code>n.bins</code>	The number of bins into which the expression data will be discretised. The default is $N^{1/3}$ for N samples. Ignored if using Bayesian Blocks discretisation.
<code>n.cores</code>	The number of cores to use for the computation. The default is 1.

Value

A symmetric matrix of pairwise mutual information between the expression of pairs of genes. These are the mutual information estimates between pairs of columns of `expr.data`.

<code>get.mim.shrink</code>	<i>Shrinkage mutual information estimate</i>
-----------------------------	--

Description

Returns a symmetric matrix of pairwise Shrinkage mutual information estimates from a samples x genes matrix of expression values. The ij^{th} element is the mutual information estimate between the expression of genes i and j in nats.

Usage

```
get.mim.shrink(expr.data, discretisation = c("equalwidth", "equalfreq", "bb"),
  n.bins = as.integer(nrow(expr.data)^(1/3)), n.cores = 1L)
```

Arguments

<code>expr.data</code>	A samples x genes matrix of expression values.
<code>discretisation</code>	The types of bins into which the expression data will be discretised. Can choose equal width bins ("equalwidth"), equal frequency bins ("equalfreq") or Bayesian Blocks bins ("bb")
<code>n.bins</code>	The number of bins into which the expression data will be discretised. The default is $N^{1/3}$ for N samples. Ignored if using Bayesian Blocks discretisation.
<code>n.cores</code>	The number of cores to use for the computation. The default is 1.

Value

A symmetric matrix of pairwise mutual information between the expression of pairs of genes. These are the mutual information estimates between pairs of columns of `expr.data`.

get.pr	<i>Compute the precision and recall from a matrix of mutual information estimates</i>
--------	---

Description

From a symmetric matrix of mutual information values between the expression of pairs of genes, infer a network then compute the precision, recall and the area under precision-recall curve by comparing the inferred network to the gold standard

Usage

```
get.pr(mim, gs.net, inf.algo = c("clr", "mrnet", "aracne"), n.reg = NULL)
```

Arguments

mim	A symmetric matrix of mutual information values. The ij^{th} element is the mutual information between the expression of genes i and j
gs.net	The symmetric adjacency matrix of the gold standard regulatory network. Should contain a 0 for no edge, 1 for an edge and NA if unknown.
inf.algo	The inference algorithm used to infer the network from the mutual information estimates. Must be one of "clr", "mrnet" or "aracne"
n.reg	The number of genes that are designated as potential regulators. If it is an integer then the first n.reg genes are marked as regulators. If a vector then those gene indexes contained in the vector are regulators. Only interactions involving regulators are used to evaluate the precision and recall. Default is NULL, in which case all the genes are potential regulators.

Value

pr	A two-column matrix of the precision and recall values of the precision-recall curve. The first column contains the recall and the second column contains the precision.
auprc	The area under the precision-recall curve.