Package ‘RNAseqNet’

January 19, 2018

Type Package

Title Log-Linear Poisson Graphical Model with Hot-Deck Multiple Imputation

Version 0.1.2

Date 2018-01-19

Maintainer Nathalie Villa-Vialaneix <nathalie.villa-vialaneix@inra.fr>

Description Infer log-linear Poisson Graphical Model with an auxiliary data set. Hot-deck multiple imputation method is used to improve the reliability of the inference with an auxiliary dataset. Standard log-linear Poisson graphical model can also be used for the inference and the Stability Approach for Regularization Selection (StARS) is implemented to drive the selection of the regularization parameter. The method is fully described in <doi:10.1093/bioinformatics/btx819>.

License GPL (>= 3)

Repository CRAN

Depends R (>= 3.1.0), ggplot2

Imports igraph (>= 1.0), hot.deck, PoiClaClu, glmnet, methods, utils

Suggests knitr, rmarkdown

VignetteBuilder knitr

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

NeedsCompilation no

Author Alyssa Imbert [aut], Nathalie Villa-Vialaneix [aut, cre]

Date/Publication 2018-01-19 19:36:18 UTC
## R topics documented:

chooseSigma ................................. 2
GLMnetToGraph ................................ 3
GLMnetwork ................................... 4
GLMpath ...................................... 6
HDImputed .................................. 6
HDPath ...................................... 7
imputedGLMnetwork ......................... 8
imputeHD ..................................... 9
lung .......................................... 11
RNAseqNetUsersGuide ....................... 11
stabilitySelection ......................... 12
stars ......................................... 13
thyroid ...................................... 14
varIntra .................................... 14

## Index

<table>
<thead>
<tr>
<th>Function</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>chooseSigma</td>
<td>2</td>
</tr>
<tr>
<td>GLMnetToGraph</td>
<td>3</td>
</tr>
<tr>
<td>GLMnetwork</td>
<td>4</td>
</tr>
<tr>
<td>GLMpath</td>
<td>6</td>
</tr>
<tr>
<td>HDImputed</td>
<td>6</td>
</tr>
<tr>
<td>HDPath</td>
<td>7</td>
</tr>
<tr>
<td>imputedGLMnetwork</td>
<td>8</td>
</tr>
<tr>
<td>imputeHD</td>
<td>9</td>
</tr>
<tr>
<td>lung</td>
<td>11</td>
</tr>
<tr>
<td>RNAseqNetUsersGuide</td>
<td>11</td>
</tr>
<tr>
<td>stabilitySelection</td>
<td>12</td>
</tr>
<tr>
<td>stars</td>
<td>13</td>
</tr>
<tr>
<td>thyroid</td>
<td>14</td>
</tr>
<tr>
<td>varIntra</td>
<td>14</td>
</tr>
</tbody>
</table>

### chooseSigma

Select the threshold sigma for hd-MI.

#### Description

chooseSigma computes the average intra-donor pool variance for different values of sigma. It helps choosing a sigma that makes a good trade-off between homogeneity within the pool of donors and variety (large enough number of donors in every pool).

#### Usage

```r
chooseSigma(X, Y, sigma_list, seed = NULL)
```

#### Arguments

- **X**: n x p numeric matrix containing RNA-seq expression with missing rows (numeric matrix or data frame)
- **Y**: auxiliary dataset (n’ x q numeric matrix or data frame)
- **sigma_list**: a sequence of increasing positive values for sigma (numeric vector)
- **seed**: single value, interpreted as an in integer, used to initialize the random number generation state

#### Details

The average intra-donor pool variance is described in (Imbert et al., 2018).

#### Value

A data frame with the values of sigma and the corresponding intra-donor pool variances
Author(s)

Alyssa Imbert, <alyssa.imbert@inra.fr>
Nathalie Villa-Vialaneix, <nathalie.villa-vialaneix@inra.fr>

References


See Also

varIntra

Examples

```r
data(lung)
data(thyroid)
nobs <- nrow(lung)
mis_ind <- sample(1:nobs, round(0.2 * nobs), replace = FALSE)
lung[mis_ind, ] <- NA
lung <- na.omit(lung)
sigma_stats <- chooseSigma(lung, thyroid, 1:5)
## Not run: plot(sigma_stats, type = "b")
```

---

**GLMnetToGraph**  
Convert the result of imputedGLMnetwork or a matrix into a network.

**Description**

GLMnetToGraph combines the m inferred networks, obtained from m imputed datasets, into a single stable network or convert a matrix of coefficients of a GLM model into a network (non zero coefficients are converted to edges)

**Usage**

```r
GLMnetToGraph(object, threshold = 0.9)
```

**Arguments**

- **object**: an object of class HDpath as obtained from the function `imputedGLMnetwork` or a squared matrix with zero and non zero values
- **threshold**: the percentage of times, among the m imputed networks, that an edge has to be predicted to be in the final network. Used only for objects of class HDpath. Default to 0.9
Value

an ‘igraph’ object. See igraph

Author(s)

Alyssa Imbert, <alyssa.imbert@inra.fr>
Nathalie Villa-Vialaneix, <nathalie.villa-vialaneix@inra.fr>

References


See Also

imputedGLMnetwork, igraph

Examples

data(lung)
data(thyroid)
nobs <- nrow(lung)
miss_ind <- sample(1:nobs, round(0.2 * nobs), replace = FALSE)
lung[miss_ind,] <- NA
lung <- na.omit(lung)
lambdas <- 4 * 10^(-seq(0, -2, length = 10))
## Not run:
lung_hdmi <- imputedGLMnetwork(lung, thyroid, sigma = 2, lambdas = lambdas,
m = 10, B = 5)
lung_net <- GLMnetToGraph(lung_hdmi, 0.75)
lung_net
plot(lung_net)
## End(Not run)

GLMnetwork

Infer a network from RNA-seq expression.

Description

GLMnetwork infers a network from RNA-seq expression with the log-linear Poisson graphical model of (Allen and Liu, 2012).

Usage

GLMnetwork(counts, lambdas = NULL, normalize = TRUE)
Arguments

- `counts`: a n x p matrix of RNA-seq expression (numeric matrix or data frame)
- `lambdas`: a sequence of decreasing positive numbers to control the regularization (numeric vector). Default to NULL
- `normalize`: logical value to normalize predictors in the log-linear Poisson graphical model. If TRUE, log normalization and scaling are performed prior the model is fit. Default to TRUE

Details

When input `lambdas` are null the default sequence of `glmnet` for the first model (the one with the first column of `count` as the target) is used.

Value

S3 object of class `GLMnetwork`: a list consisting of

- `lambda` regularization parameters used for LLGM path(vector)
- `path` a list having the same length than `lambda`. It contains the estimated coefficients (in a matrix) along the path

Author(s)

Alyssa Imbert, <alyssa.imbert@inra.fr>
Nathalie Villa-Vialaneix, <nathalie.villa-vialaneix@inra.fr>

References


See Also

`stabilitySelection`

Examples

```r
data(lung)
lambdas <- 4 * 10^(seq(0, -2, length = 10))
ref_lung <- GLMnetwork(lung, lambdas = lambdas)
```
### Description
Methods for the result of `GLMnetwork` (GLMpath object)

#### Usage

```r
## S3 method for class 'GLMpath'
summary(object, ...)

## S3 method for class 'GLMpath'
print(x, ...)
```

#### Arguments

- `object` : GLMpath object
- `...` : not used
- `x` : GLMpath object

#### Author(s)

Alyssa Imbert, <alyssa.imbert@inra.fr>
Nathalie Villa-Vialaneix, <nathalie.villa-vialaneix@inra.fr>

#### See Also

`GLMnetwork`

---

### Description
Methods for the result of `imputeHD` (HDImputed object)

#### Usage

```r
## S3 method for class 'HDImputed'
summary(object, ...)

## S3 method for class 'HDImputed'
print(x, ...)
```
**HDpath**

**Arguments**

- object: HDImputed object
- ...: not used
- x: HDImputed object

**Author(s)**

Alyssa Imbert, alyssa.imbert@inra.fr
Nathalie Villa-Vialaneix, nathalie.villa-vialaneix@inra.fr

**See Also**

imputeHD

---

**Methods for 'HDpath' objects.**

**Description**

Methods for the result of `imputedGLMnetwork` (HDpath object)

**Usage**

```r
## S3 method for class 'HDpath'
summary(object, ...)

## S3 method for class 'HDpath'
print(x, ...)

## S3 method for class 'HDpath'
plot(x, ...)
```

**Arguments**

- object: HDpath object
- ...: not used
- x: HDpath object

**Author(s)**

Alyssa Imbert, alyssa.imbert@inra.fr
Nathalie Villa-Vialaneix, nathalie.villa-vialaneix@inra.fr

**See Also**

imputedGLMnetwork
Examples

```r
data(lung)
data(thyroid)
nobs <- nrow(lung)
miss_ind <- sample(1:nobs, round(0.2 * nobs), replace = FALSE)
lung[miss_ind, ] <- NA
lung <- na.omit(lung)
lambdas <- 4 * 10^(-seq(0, -2, length = 10))
## Not run:
lung_hdmi <- imputedGLMnetwork(lung, thyroid, sigma = 2, lambdas = lambdas, m = 10, B = 5)
plot(lung_hdmi)
## End(Not run)
```

imputedGLMnetwork

Multiple hot-deck imputation and network inference from RNA-seq data.

Description

imputedGLMnetwork performs a multiple hot-deck imputation and infers a network for each imputed dataset with a log-linear Poisson graphical model (LLGM).

Usage

```r
imputedGLMnetwork(X, Y, sigma, m = 50, lambdas = NULL, B = 20)
```

Arguments

- **X**: n x p numeric matrix containing RNA-seq expression with missing rows (numeric matrix or data frame)
- **Y**: auxiliary dataset (n’ x q numeric matrix or data frame)
- **sigma**: affinity threshold for donor pool
- **m**: number of replicates in multiple imputation (integer). Default to 50
- **lambdas**: a sequence of decreasing positive numbers to control the regularization (numeric vector). Default to NULL
- **B**: number of iterations for stability selection. Default to 20

Details

When input lambdas are null the default sequence of `glmnet` for the first model (the one with the first column of count as the target) is used. A common default sequence is generated for all imputed datasets using this method.
imputeHD

Value

S3 object of class HDpath: a list consisting of

- path a list of \( m \) data frames, each containing the adjacency matrix of the inferred network obtained from the corresponding imputed dataset. The regularization parameter is selected by StARS
- efreq a numeric matrix of size \( p \times p \), which indicates the number of times an edge has been predicted among the \( m \) inferred networks

Author(s)

Alyssa Imbert, <alyssa.imbert@inra.fr>
Nathalie Villa-Vialaneix, <nathalie.villa-vialaneix@inra.fr>

References


Examples

data(lung)
data(thyroid)
nobs <- nrow(lung)
miss_ind <- sample(1:nobs, round(0.2 * nobs), replace = FALSE)
lung[miss_ind, ] <- NA
lung <- na.omit(lung)
lambdas <- 4 * 10^((seq(0, -2, length = 10)))
## not run:
lung_hdmi <- imputedGLMnetwork(lung, thyroid, sigma = 2, lambdas = lambdas,
                                m = 10, B = 5)

## End(Not run)

imputeHD

Impute missing row datasets with multiple hot deck.

Description

`imputeHD` performs multiple hot-deck imputation on an input data frame with missing rows. Each missing row is imputed with a unique donor. This method requires an auxiliary dataset to compute similarities between individuals and create the pool of donors.

Usage

`imputeHD(X, Y, sigma, m = 50, seed = NULL)`
Arguments

\( X \)  
\( n \times p \) numeric matrix containing RNA-seq expression with missing rows (numeric matrix or data frame)

\( Y \)  
auxiliary dataset (\( n' \times q \) numeric matrix or data frame)

\( \sigma \)  
threshold for hot-deck imputation (numeric, positive)

\( m \)  
number of replicates in multiple imputation (integer). Default to 50

\( \text{seed} \)  
single value, interpreted as an in integer, used to initialize the random number generation state. Default to NULL (not used in this case)

Details

Missing values are identified by matching rownames in \( X \) and \( Y \). If rownames are not provided the missing rows in \( X \) are supposed to correspond to the last rows of \( Y \).

Value

S3 object of class `HDImputed`: a list consisting of

- `donors` a list. Each element of this list contains the donor pool for every missing observations
- `draws` a data frame which indicates which donor was chosen for each missing samples
- `data` a list of \( m \) imputed datasets

Author(s)

Alyssa Imbert, <alyssa.imbert@inra.fr>

Nathalie Villa-Vialaneix, <nathalie.villa-vialaneix@inra.fr>

References


See Also

`chooseSigma`, `imputedGLMnetwork`

Examples

```r
data(lung)
data(thyroid)
nobs <- nrow(lung)
miss_ind <- sample(1:nobs, round(0.2 * nobs), replace = FALSE)
lung[miss_ind, ] <- NA
lung <- na.omit(lung)
imputed_lung <- imputeHD(lung, thyroid, sigma = 2)
```
Description

This data set is a small subset of the full data set from GTEx. It contains RNA-seq expressions measured from lung tissue. The RNA-seq expressions have been normalized with the TMM method.

Format

a data frame with 221 rows and 100 variables (genes). Rownames are identifiers for individuals.

Author(s)

Alyssa Imbert <alyssa.imbert@inra.fr>

Source

The raw data were downloaded from https://gtexportal.org/. The TMM normalisation of RNA-seq expression was performed with the R package edger.

Description

Find the location of the RNAseqNet User’s Guide and optionally opens it.

Usage

RNAseqNetUsersGuide(html = TRUE, view = html)

Arguments

html logical. Should the document returned by the function be the compiled PDF or the Rmd source. Default to TRUE

view logical. Should the document be opened using the default HTML viewer? Default to html. It has no effect if html = FALSE.

Details

The function vignette("RNAseqNet") will find the short RNAseqNet vignette that describes how to obtain the RNAseqNet User’s Guide. The User’s Guide is not itself a true vignette because it is not automatically generated during the package build process. However, the location of the Rmarkdown source is returned by the function if html = FALSE. If the operating system is not Windows, then the HTML viewer used is that given by Sys.getenv("R_BROWSER"). The HTML viewer can be changed using Sys.getenv(R_BROWSER = ).
stabilitySelection

Value
Character string giving the file location. If html = TRUE and view = TRUE, the HTML document reader is started and the User's Guide is opened in it.

Author(s)
Alyssa Imbert, <alyssa.imbert@inra.fr>
Nathalie Villa-Vialaneix, <nathalie.villa-vialaneix@inra.fr>

Examples
RNaseqNetUsersGuide(view = FALSE)
RNaseqNetUsersGuide(html = FALSE)
## Not run: RNaseqNetUsersGuide()

stabilitySelection Selection of the regularization parameter by StARS (Liu et al., 2010).

Description
stabilitySelection implements the regularization parameter selection of (Liu et al., 2010) called 'Stability Approach to Regularization Selection' (StARS).

Usage
stabilitySelection(counts, lambdas = NULL, B = 20)

Arguments
counts a n x p matrix of RNA-seq expression (numeric matrix or data frame)
lambdas a sequence of decreasing positive numbers to control the regularization (numeric vector). Default to NULL
B number of iterations for stability selection. Default to 20

Details
When input lambdas are null the default sequence of glmnet (see GLMnetwork for details).

Value
S3 object of class stabilitySelection: a list consisting of
• lambdas numeric regularization parameters used for regularization path
• B number of iterations for stability selection
• best index of the regularization parameter selected by StARS in lambdas
• variabilities numeric vector having same length than lambdas and providing the variability value as defined by StARS along the path
Author(s)
Alyssa Imbert, <alyssa.imbert@inra.fr> Nathalie Villa-Vialaneix, <nathalie.villa-vialaneix@inra.fr>

References

See Also
GLMnetwork

Examples
```r
data(lung)
lambda <- 4 * 10^seq(0, -2, length = 5))
stability_lung <- stabilitySelection(lung, lambda = lambda, B = 4)
## Not run: plot(stability_lung)
```

stars

Methods for 'stars' objects.

Description
Methods for the result of stabilitySelection (stars object)

Usage
```r
## S3 method for class 'stars'
summary(object, ...)

## S3 method for class 'stars'
print(x, ...)

## S3 method for class 'stars'
plot(x, ...)
```

Arguments
- `object` stars object
- `...` not used
- `x` stars object
varIntra

Author(s)
Alyssa Imbert, <alyssa.imbert@inra.fr>
Nathalie Villa-Vialaneix, <nathalie.villa-vialaneix@inra.fr>

See Also
stabilitySelection

```
thyroid  RNA-seq expression from thyroid tissue (GTEx).
```

Description
This data set is a small subset of the full data set from GTEx. It contains RNA-seq expressions measured from thyroid tissue. The RNA-seq expressions have been normalized with the TMM method.

Format
a data frame with 221 rows and 50 variables (genes). Rownames are identifiers for individuals.

Author(s)
Alyssa Imbert <alyssa.imbert@inra.fr>

Source
The raw data were downloaded from https://gtexportal.org/. The TMM normalisation of RNA-seq expression was performed with the R package edgeR.

varIntra  Average intra-donor pool variance.

Description
varIntra computes the average intra-donor pool variance.

Usage
varIntra(X, Y, donors)
varIntra

Arguments

X  n x p numeric matrix containing RNA-seq expression with missing rows (numeric matrix or data frame)
Y  auxiliary dataset (n’ x q numeric matrix or data frame)
donors  donor pool (a list, as given $donors obtained from the function imputeHD)

Value

varIntra returns a numeric value which is the average intra-donor pool variance, as described in (Imbert et al., 2018).

Author(s)

Alyssa Imbert, <alyssa.imbert@inra.fr>
Nathalie Villa-Vialaneix, <nathalie.villa-vialaneix@inra.fr>

References


See Also

imputeHD, chooseSigma
Index

chooseSigma, 2, 10, 15

glmnet, 5, 8, 12
GLMnetToGraph, 3
GLMnetwork, 4, 6, 12, 13
GLMpath, 6

HDImputed, 6
HDpath, 7

igraph, 4
imputedGLMnetwork, 3, 4, 7, 8, 10
imputeHD, 6, 7, 9, 15

lung, 11

plot.HDpath (HDpath), 7
plot.stars (stars), 13
print.GLMpath (GLMpath), 6
print.HDImputed (HDImputed), 6
print.HDpath (HDpath), 7
print.stars (stars), 13

RNaseqNetUsersGuide, 11

stabilitySelection, 5, 12, 13, 14
stars, 13
summary.GLMpath (GLMpath), 6
summary.HDImputed (HDImputed), 6
summary.HDImputed (HDImputed), 6
summary.HDpath (HDpath), 7
summary.stars (stars), 13

thyroid, 14

varIntra, 3, 14