Package 'ORIClust'

February 19, 2015

Type Package

Title Order-restricted Information Criterion-based Clustering Algorithm

Version 1.0-1

Date 2009-09-10

Author Tianqing Liu, Nan Lin, Ningzhong Shi and Baoxue Zhang

Maintainer Tianqing Liu <tianqingliu@gmail.com>

Description ORIClust is a user-friendly R-based software package for gene clustering. Clusters are given by genes matched to prespecified profiles across various ordered treatment groups. It is particularly useful for analyzing data obtained from short time-course or dose-response microarray experiments.

License GPL-2

LazyLoad yes

Repository CRAN

Date/Publication 2012-10-29 08:57:22

NeedsCompilation no

R topics documented:

`ORIClust-package` *Order-restricted Information Criterion-based Clustering Algorithm*

Description

ORIClust is a user-friendly R-based software package for gene clustering. Clusters are given by genes matched to prespecified profiles across various ordered treatment groups. It is particularly useful for analyzing data obtained from short time-course or dose-response microarray experiments.

Details

Package:	ORIClust
Type:	Package
Version:	1.0
Date:	2009-05-24
License:	GPL-2
LazyLoad:	yes

The main functions are `ORICC1` and `ORICC2`, see the documentation files with examples.

Author(s)

Tianqing Liu, Nan Lin, Ningzhong Shi and Baoxue Zhang

Maintainer: Tianqing Liu <tianqingliu@gmail.com>

References

Liu, T., Lin, N., Shi, N. and Zhang, B. (2009), Information criterion-based clustering with order-restricted candidate profiles in short time-course microarray experiments, BMC Bioinformatics, 10: 146.

`Breast` *Breast cancer cell line data*

Description

This data set comes from a breast cancer cell line microarray study. The experiment was done as follows. First, the MCF-7 breast cancer cell line was treated with 17 beta-estradiol or ethanol (vehicle control). Then, samples were harvested at 1, 4, 12, 24, 36 and 48 hours after treatment. At each time point, M = 8 replicate arrays were prepared with each array consisting of G = 1901 genes.

Usage

```
Breast
```

Format

A matrix containing 1901 rows and 50 columns.

References

Lobenhofer, E., Bennett, L., Cable, P., Li, L., Bushel, P., and Afshari, C. (2002), *Regulation of DNA replication fork genes by 17 beta-estradiol.* Molec. Endocrin., **16**, 1215-1229.

`complete.profile` *complete.profile*

Description

Returns the log-maximum likelihood and the estimator of the mean when there is no inequality constraint.

Usage

```
complete.profile(data,x,n.rep)
```

Arguments

`data`	A vector containing the expressions of one gene.
`x`	A vector consisting of the average expression at time points (1, 2,. . . ,T), where T is the total number of time points.
`n.rep`	A vector consisting of the number of replicate arrays at time points (1, 2,. . . ,T), where T is the total number of time points.

Value

`logelr`	Log-maximum likelihood
`mu`	A vector containing the estimator of the mean

Author(s)

Tianqing Liu, Nan Lin, Ningzhong Shi and Baoxue Zhang

Maintainer:Tianqing Liu <tianqingliu@gmail.com>

`cyclical.max.min` *cyclical.max.min*

Description

Returns the log-maximum likelihood and the estimator of the mean under cyclical profile with maximum at `max1` and minimum at `min1` (`max1` < `min1`).

Usage

```
cyclical.max.min(data,x,n.rep,max1,min1)
```

Arguments

`data`	A vector containing the expressions of one gene.
`x`	A vector consisting of the average expression at time points (1, 2,. . . ,T), where T is the total number of time points.
`n.rep`	A vector consisting of the number of replicate arrays at time points (1, 2,. . . ,T), where T is the total number of time points.
`max1`	Cyclical profile with maximum at `max1`.
`min1`	Cyclical profile with minimum at `min1`.

Value

`logelr`	Log-maximum likelihood
`mu`	A vector containing the estimator of the mean

Author(s)

Tianqing Liu, Nan Lin, Ningzhong Shi and Baoxue Zhang

Maintainer: Tianqing Liu <tianqingliu@gmail.com>

References

Liu, T., Lin, N., Shi, N. and Zhang, B. (2009), *Information criterion-based clustering with order-restricted candidate profiles in short time-course microarray experiments.* BMC Bioinformatics, **10**: 146.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order restricted statistical inference.* New York: Wiley.

Shi, N., Gao, W. and Zhang, B. (2001). *One sided estimation and testing problems for location models from grouped samples.* Commun Statist-Simula, **30**: 885-898.

cyclical.min.max *cyclical.min.max*

Description

Returns the log-maximum likelihood and the estimator of the mean under cyclical profile with minimum at `min1` and maximum at `max1` (`min1` < `max1`).

Usage

```
cyclical.min.max(data,x,n.rep,min1,max1)
```

Arguments

`data`	A vector containing the expressions of one gene.
`x`	A vector consisting of the average expression at time points (1, 2,. . . ,T), where T is the total number of time points.
`n.rep`	A vector consisting of the number of replicate arrays at time points (1, 2,. . . ,T), where T is the total number of time points.
`min1`	Cyclical profile with minimum at `min1`
`max1`	cyclical profile with maximum at `max1`

Value

`logelr`	Log-maximum likelihood
`mu`	A vector containing the estimator of the mean

Author(s)

Tianqing Liu, Nan Lin, Ningzhong Shi and Baoxue Zhang

Maintainer: Tianqing Liu <tianqingliu@gmail.com>

References

Liu, T., Lin, N., Shi, N. and Zhang, B. (2009), *Information criterion-based clustering with order-restricted candidate profiles in short time-course microarray experiments.* BMC Bioinformatics, **10**: 146.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order restricted statistical inference.* New York: Wiley.

Shi, N., Gao, W. and Zhang, B. (2001). *One sided estimation and testing problems for location models from grouped samples.* Commun Statist-Simula, **30**: 885-898.

`decreasing` *decreasing*

Description

Returns the log-maximum likelihood and the estimator of the mean under the monotone decreasing profile.

Usage

```
decreasing(data,x,n.rep)
```

Arguments

`data`	A vector containing the expressions of one gene.
`x`	A vector consisting of the average expression at time points (1, 2,. . . ,T), where T is the total number of time points.
`n.rep`	A vector consisting of the number of replicate arrays at time points (1, 2,. . . ,T), where T is the total number of time points.

Value

`logelr`	Log-maximum likelihood
`mu`	A vector containing the estimator of the mean

Author(s)

Tianqing Liu, Nan Lin, Ningzhong Shi and Baoxue Zhang

Maintainer: Tianqing Liu <tianqingliu@gmail.com>

References

Liu, T., Lin, N., Shi, N. and Zhang, B. (2009), *Information criterion-based clustering with order-restricted candidate profiles in short time-course microarray experiments.* BMC Bioinformatics, **10**: 146.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order restricted statistical inference.* New York: Wiley.

Shi, N., Gao, W. and Zhang, B. (2001). *One sided estimation and testing problems for location models from grouped samples.* Commun Statist-Simula, **30**: 885-898.

down.up *down.up*

Description

Returns the log-maximum likelihood and the estimator of the mean under down-up profile with minimum at h.

Usage

```
down.up(data,x,n.rep,h)
```

Arguments

data	A vector containing the expressions of one gene.
x	A vector consisting of the average expression at time points (1, 2,. . . ,T), where T is the total number of time points.
n.rep	A vector consisting of the number of replicate arrays at time points (1, 2,. . . ,T), where T is the total number of time points.
h	Down-up profile with minimum at h.

Value

logelr	Log-maximum likelihood
mu	A vector containing the estimator of the mean

Author(s)

Tianqing Liu, Nan Lin, Ningzhong Shi and Baoxue Zhang

Maintainer: Tianqing Liu <tianqingliu@gmail.com>

References

Liu, T., Lin, N., Shi, N. and Zhang, B. (2009), *Information criterion-based clustering with order-restricted candidate profiles in short time-course microarray experiments.* BMC Bioinformatics, **10**: 146.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order restricted statistical inference.* New York: Wiley.

Shi, N., Gao, W. and Zhang, B. (2001). *One sided estimation and testing problems for location models from grouped samples.* Commun Statist-Simula, **30**: 885-898.

`flat.pattern` *flat.pattern*

Description

Returns the log-maximum likelihood and the estimator of the mean under the equality constraint that all means are equal.

Usage

```
flat.pattern(data,x,n.rep)
```

Arguments

`data`	A vector containing the expressions of one gene.
`x`	A vector consisting of the average expression at time points (1, 2,. . . ,T), where T is the total number of time points.
`n.rep`	A vector consisting of the number of replicate arrays at time points (1, 2,. . . ,T), where T is the total number of time points.

Value

`logelr`	Log-maximum likelihood
`mu`	A vector containing the estimator of the mean

Author(s)

Tianqing Liu, Nan Lin, Ningzhong Shi and Baoxue Zhang

Maintainer: Tianqing Liu <tianqingliu@gmail.com>

References

Liu, T., Lin, N., Shi, N. and Zhang, B. (2009), *Information criterion-based clustering with order-restricted candidate profiles in short time-course microarray experiments.* BMC Bioinformatics, **10**: 146.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order restricted statistical inference.* New York: Wiley.

Shi, N., Gao, W. and Zhang, B. (2001). *One sided estimation and testing problems for location models from grouped samples.* Commun Statist-Simula, **30**: 885-898.

increasing *increasing*

Description

Returns the log-maximum likelihood and the estimator of the mean under the monotone increasing profile.

Usage

```
increasing(data,x,n.rep)
```

Arguments

data	A vector containing the expressions of one gene.
x	A vector consisting of the average expression at time points (1, 2,. . . ,T), where T is the total number of time points.
n.rep	A vector consisting of the number of replicate arrays at time points (1, 2,. . . ,T), where T is the total number of time points.

Value

logelr	Log-maximum likelihood
mu	A vector containing the estimator of the mean

Author(s)

Tianqing Liu, Nan Lin, Ningzhong Shi and Baoxue Zhang

Maintainer: Tianqing Liu <tianqingliu@gmail.com>

References

Liu, T., Lin, N., Shi, N. and Zhang, B. (2009), *Information criterion-based clustering with order-restricted candidate profiles in short time-course microarray experiments.* BMC Bioinformatics, **10**: 146.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order restricted statistical inference.* New York: Wiley.

Shi, N., Gao, W. and Zhang, B. (2001). *One sided estimation and testing problems for location models from grouped samples.* Commun Statist-Simula, **30**: 885-898.

isodecre *isodecre*

Description

Isotonic regression of a with weights w under monotone decreasing profile.

Usage

```
isodecre(a, w)
```

Arguments

a	A vector consisting of the average expression at time points (1, 2,. . . ,T), where T is the total number of time points.
w	The weights.

Value

is	A vector containing the estimator of the mean

Author(s)

Tianqing Liu, Nan Lin, Ningzhong Shi and Baoxue Zhang

Maintainer: Tianqing Liu <tianqingliu@gmail.com>

References

Liu, T., Lin, N., Shi, N. and Zhang, B. (2009), *Information criterion-based clustering with order-restricted candidate profiles in short time-course microarray experiments.* BMC Bioinformatics, **10**: 146.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order restricted statistical inference.* New York: Wiley.

Shi, N., Gao, W. and Zhang, B. (2001). *One sided estimation and testing problems for location models from grouped samples.* Commun Statist-Simula, **30**: 885-898.

`isoincre` *isoincre*

Description

Isotonic regression of `a` with weights `w` under monotone increasing profile.

Usage

```
isoincre(a, w)
```

Arguments

`a`	A vector consisting of the average expression at time points (1, 2,. . . ,T), where T is the total number of time points.
`w`	The weights.

Value

`is`	A vector containing the estimator of the mean

Author(s)

Tianqing Liu, Nan Lin, Ningzhong Shi and Baoxue Zhang

Maintainer: Tianqing Liu <tianqingliu@gmail.com>

References

Liu, T., Lin, N., Shi, N. and Zhang, B. (2009), *Information criterion-based clustering with order-restricted candidate profiles in short time-course microarray experiments.* BMC Bioinformatics, **10**: 146.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order restricted statistical inference.* New York: Wiley.

Shi, N., Gao, W. and Zhang, B. (2001). *One sided estimation and testing problems for location models from grouped samples.* Commun Statist-Simula, **30**: 885-898.

ORICC1 *One-stage ORICC*

Description

One-stage ORICC is a computationally efficient information criterion-based clustering algorithm for selecting and clustering genes according to their time-course or dose-response profiles. This algorithm takes account of the ordering in time-course or dose-response experiments by embedding the order-restricted inference into a model selection framework. This algorithm mainly consist of two steps. In the first step, candidate profiles are defined in terms of inequalities among mean expression levels at different time points or doses levels. In the second step, genes are assigned to the best matched profiles determined by an information criterion for order-restricted inference.

Usage

```
ORICC1(data,data.col,id.col,n.rep,n.top,transform,
       name.profile,cyclical.profile,complete.profile,
       onefile,plot.format)
```

Arguments

data	A matrix containing the gene expressions.
data.col	Column indices of the gene expression data.
id.col	Column index of the gene ID. Defaults to 1.
n.rep	A vector consisting of the number of replicate arrays at time points (1, 2,. . . ,T), where T is the total number of time points.
n.top	The number of genes kept for the final clustering result. Genes are ranked based on expression variation across time or dose levels. Defaults to all genes ORICC1 selects
transform	Transformation of the original data: 0=None, 1=natural log, 2=square root, 3=cubic root. Defaults to 0.
name.profile	A character string specifying the collection of candidate profiles. This option only supports monotone, up-down and down-up profiles specified as by "decreasing"; "increasing". paste("up down max at",i,sep=" "); paste("down up min at",j,sep=" "); If name.profile="all", the 'decreasing', 'increasing' and all 'up-down' and 'down-up' profiles will be included. If name.profile=NULL, 'decreasing', 'increasing' and all 'up-down' and 'down-up' profiles will be absent. Defaults to NULL. One can also specify several up-down or down-up profiles together as follows.

```
profile1=paste("up down max at",c(2,4),sep=" ");
profile2=paste("down up min at",c(3,5),sep=" ");
name.profile=c(profile1,profile2);
```

then up-down profile with maxima at 2 and 4 as well as down-up profile with minima at 3 and 5 will be included.

`cyclical.profile`

A matrix with 2 columns. Each element of the matrix must be a number in the set {2,3,...,T-1}. Each row of the matrix represents a cyclical profile with minima at the first entry of the row and maxima at the 2nd entry. As a result, two elements in the same row must be different. For example, if

`cyclical.profile=matrix(c(2,3,4,3),2,2,byrow=T)`, then the cyclical profile with minima at 2 and maxima at 3 and the cyclical profile with minima at 4 and maxima at 3 will be included as candidate profiles.

If `cyclical.profile=NULL`, all cyclical profiles will be absent. Defaults to NULL.

`complete.profile`

The `complete.profile` means a profile in which there is no defined inequality constraint.

If the `complete.profile` is a candidate profile,

`complete.profile=1`, otherwise,

`complete.profile=NULL`. Defaults to NULL.

`onefile` logical: if true (the default) multiple figures for different clusters are output in one file. If FALSE, each cluster is plotted in a seperate file. Defaults to TRUE.

`plot.format` The format of the output file containing plots of gene clusters.Users can choose between 'eps' and 'jpg'. Defaults to 'eps'.

Details

The gene expression dataset should be in a tab-delimited txt file, in which the first two columns contain the gene names and their accession numbers or descriptions, and the remaining columns, in their orders, are the geneexpression data (contain multiple columns, i.e. `data.col`).The dataset is assumed to have been processed so that each row contains the expressions of only one gene.

Value

The results are displayed in a graphical form. The graphics can be stored in a JPG or EPS format. Both the raw gene expression values and the estimated mean expressions are output to external files 'cluster of raw data.txt' and 'cluster of fitted mean data.txt', respectively.

Author(s)

Tianqing Liu, Nan Lin, Ningzhong Shi and Baoxue Zhang

Maintainer: Tianqing Liu <tianqingliu@gmail.com>

References

Liu, T., Lin, N., Shi, N. and Zhang, B. (2009), *Information criterion-based clustering with order-restricted candidate profiles in short time-course microarray experiments.* BMC Bioinformatics, **10**: 146.

Examples

```
data(Breast)
ORICC1(Breast,data.col=3:50,id.col=1,n.rep=rep(8,6),
       n.top=50,transform=1,name.profile="all",plot.format="eps")
```

ORICC2 *Two-stage ORICC*

Description

It is a computationally efficient two-stage algorithm by adding a pre-screening stage. It first screens out genes that show no significant changes over time, and then applies the one-stage algorithm to a much smaller set of remained genes.

Usage

```
ORICC2(data,data.col,id.col,n.rep,n.top,transform,
       name.profile,cyclical.profile,
       onefile,plot.format)
```

Arguments

`data`	A matrix containing the gene expressions.
`data.col`	Column indices of the gene expression data.
`id.col`	Column index of the gene ID. Defaults to 1.
`n.rep`	A vector consisting of the number of replicate arrays at time points (1, 2,. . . ,T), where T is the total number of time points.
`n.top`	The number of genes kept for the final clustering result. Genes are ranked based on expression variation across time or dose levels. Defaults to all genes ORICC2 selects
`transform`	Transformation of the original data: 0=None, 1=natural log, 2=square root, 3=cubic root. Defaults to 0.

`name.profile`	A character string specifying the collection of candidate profiles. This option only supports monotone, up-down and down-up profiles specified as by
	`"decreasing";`
	`"increasing".`
	`paste("up down max at",i,sep=" ");`

paste("down up min at",j,sep=" ");

If name.profile="all", the 'decreasing', 'increasing' and all 'up-down' and 'down-up' profiles will be included.

If name.profile=NULL, 'decreasing', 'increasing' and all 'up-down' and 'down-up' profiles will be absent. Defaults to NULL.

One can also specify several up-down or down-up profiles together as follows.

```
profile1=paste("up down max at",c(2,4),sep=" ");
profile2=paste("down up min at",c(3,5),sep=" ");
name.profile=c(profile1,profile2);
```

then up-down profile with maxima at 2 and 4 as well as down-up profile with minima at 3 and 5 will be included.

`cyclical.profile`

A matrix with 2 columns. Each element of the matrix must be a number in the set {2,3,...,T-1 }. Each row of the matrix represents a cyclical profile with minima at the first entry of the row and maxima at the 2nd entry. As a result, two elements in the same row must be different. For example, if

`cyclical.profile=matrix(c(2,3,4,3),2,2,byrow=T)`, then the cyclical profile with minima at 2 and maxima at 3 and the cyclical profile with minima at 4 and maxima at 3 will be included as candidate profiles.

If `cyclical.profile=NULL`, all cyclical profiles will be absent. Defaults to NULL.

`onefile` logical: if true (the default) multiple figures for different clusters are output in one file. If FALSE, each cluster is plotted in a seperate file. Defaults to TRUE.

`plot.format` The format of the output file containing plots of gene clusters.Users can choose between 'eps' and 'jpg'. Defaults to 'eps'.

Details

The gene expression dataset should be in a tab-delimited txt file, in which the first two columns contain the gene names and their accession numbers or descriptions, and the remaining columns, in their orders, are the geneexpression data (contain multiple columns, i.e. `data.col`).The dataset is assumed to have been processed so that each row contains the expressions of only one gene.

Value

The results are displayed in a graphical form. The graphics can be stored in a JPG or EPS format. Both the raw gene expression values and the estimated mean expressions are output to external files 'cluster of raw data.txt' and 'cluster of fitted mean data.txt', respectively.

Author(s)

Tianqing Liu, Nan Lin, Ningzhong Shi and Baoxue Zhang

Maintainer: Tianqing Liu <tianqingliu@gmail.com>

References

Liu, T., Lin, N., Shi, N. and Zhang, B. (2009), *Information criterion-based clustering with order-restricted candidate profiles in short time-course microarray experiments.* BMC Bioinformatics, **10**: 146.

Examples

```
data(Breast)
ORICC2(Breast,data.col=3:50,id.col=1,n.rep=rep(8,6),
       n.top=50,transform=1,name.profile="all",plot.format="eps")
```

`up.down` *up.down*

Description

Returns the log-maximum likelihood and the estimator of the mean under under up-down profile with maximum at `h`.

Usage

```
up.down(data,x,n.rep,h)
```

Arguments

`data`	A vector containing the expressions of one gene.
`x`	A vector consisting of the average expression at time points (1, 2,. . . ,T), where T is the total number of time points.
`n.rep`	A vector consisting of the number of replicate arrays at time points (1, 2,. . . ,T), where T is the total number of time points.
`h`	Up-down profile with maximum at `h`.

Value

`logelr`	Log-maximum likelihood
`mu`	A vector containing the estimator of the mean

Author(s)

Tianqing Liu, Nan Lin, Ningzhong Shi and Baoxue Zhang

Maintainer: Tianqing Liu <tianqingliu@gmail.com>

References

Liu, T., Lin, N., Shi, N. and Zhang, B. (2009), *Information criterion-based clustering with order-restricted candidate profiles in short time-course microarray experiments.* BMC Bioinformatics, **10**: 146.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order restricted statistical inference.* New York: Wiley.

Shi, N., Gao, W. and Zhang, B. (2001). *One sided estimation and testing problems for location models from grouped samples.* Commun Statist-Simula, **30**: 885-898.

Index