

Package ‘klic’

May 25, 2018

Title Kernel Learning Integrative Clustering

Version 0.0.1.0

Description

Unsupervised and semi-supervised clustering based on multiple kernel learning algorithms.

Depends R (>= 3.4.1)

License GPL-3

Encoding UTF-8

LazyData true

Imports Rmosek, Matrix, cluster

RoxygenNote 6.0.1.9000

R topics documented:

coca	1
consensusCluster	2
copheneticCorrelation	3
kkmeansTrain	3
lmkkmeansTrain	4
maximiseSilhouette	4
plotSilhouette	5
plotSimilarityMatrix	5
spectrumShift	6

Index	7
--------------	---

coca	<i>Cluster-Of-Clusters Analysis</i>
------	-------------------------------------

Description

This function allows to do Cluster-Of-Clusters-Analysis on a binary matrix where each column is a clustering of the data, each row corresponds to a data point and the element in position (i,j) is equal to 1 if data point i belongs to cluster j, 0 otherwise.

Usage

```
coca(dataset, K, B = 1000, pItem = 0.8, hclustMethod = "average")
```

Arguments

K	number of clusters.
B	number of iterations of the Consensus Cluster step.
pItem	proportion of items sampled at each iteration of the Consensus Cluster step.
hclustMethod	method to be used by the hclust function for the hierarchical clustering step.
data	N X C data matrix, where C is the total number of clusters considered.

Value

The output is a consensus matrix, that is a symmetric matrix where the element in position (i,j) corresponds to the proportion of times that items i and j have been clustered together.

Author(s)

Alessandra Cabassi <ac2051@cam.ac.uk>

References

Monti, S., Tamayo, P., Mesirov, J. and Golub, T., 2003. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. Machine learning, 52(1-2), pp.91-118.

The Cancer Genome Atlas, 2012. Comprehensive molecular portraits of human breast tumours. Nature, 487(7407), pp.61-70.

consensusCluster	<i>Consensus clustering with k-means</i>
------------------	--

Description

This function allows to perform consensus clustering using the k-means clustering algorithm, for a fixed number of clusters. Contrarily to what was suggested in the original paper by Monti et al. (2003), here we consider the number clusters K to be fixed, for simplicity.

Usage

```
consensusCluster(data, K, B = 100, pItem = 0.8, clMethod = "km",
  distHC = "euclidean")
```

Arguments

data	N X P data matrix
K	number of clusters
B	of iterations
pItem	proportion of items sampled at each iteration
clMethod	clustering method

Value

The output is a consensus matrix, that is a symmetric matrix where the element in position (i,j) corresponds to the proportion of times that items i and j have been clustered together.

Author(s)

Alessandra Cabassi <ac2051@cam.ac.uk>

References

Monti, S., Tamayo, P., Mesirov, J. and Golub, T., 2003. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. Machine learning, 52(1-2), pp.91-118.

copheneticCorrelation	<i>Cophenetic correlation coefficient</i>
-----------------------	---

Description

Compute the cophenetic correlation coefficient of a kernel matrix.

Usage

```
copheneticCorrelation(kernelMatrix)
```

Arguments

kernelMatrix kernel matrix.

Author(s)

Alessandra Cabassi <ac2051@cam.ac.uk>

References

Sokal, R.R. and Rohlf, F.J., 1962. The comparison of dendrograms by objective methods. Taxon, 11(2), pp.33-40.

kkmeansTrain	<i>Kernel k-means</i>
--------------	-----------------------

Description

Perform the training step of kernel k-means

Usage

```
kkmeansTrain(K, parameters)
```

Arguments

K kernel matrix
parameters list of parameters

Author(s)

Mehmet Gonen

References

Gönen, M. and Margolin, A.A., 2014. Localized data fusion for kernel k-means clustering with application to cancer biology. In Advances in Neural Information Processing Systems (pp. 1305-1313).

1mkkmeansTrain

Localised multiple kernel k-means

Description

Perform the training step of the localised multiple kernel k-means

Usage

```
1mkkmeansTrain(Km, parameters, verbose = FALSE)
```

Arguments

Km	kernel matrices
parameters	list of parameters
verbose	Boolean flag. If TRUE, at each iteration the iteration number is printed. Defaults to FALSE.

Author(s)

Mehmet Gonen

References

Gönen, M. and Margolin, A.A., 2014. Localized data fusion for kernel k-means clustering with application to cancer biology. In Advances in Neural Information Processing Systems (pp. 1305-1313).

maximiseSilhouette

Choose K that maximises the silhouette from a set kernel matrices and clusterings

Description

Choose K that maximises the silhouette from a set kernel matrices and clusterings

Usage

```
maximiseSilhouette(KM, clLabels, maxK, savePNG = FALSE,  
  fileName = "silhouette")
```

Arguments

maxK	Maximum number of clusters considered.
savePNG	If TRUE, a plot of the silhouette is saved in the working folder. Defaults to FALSE.
fileName	If savePNG is TRUE, this is the name of the png file.
kernelMatrix	N X N X (maxK-1) array of kernel matrices.
clustering	(maxK-1) X N matrix containing the clusterings obtained for different values of K

Author(s)

Alessandra Cabassi <ac2051@cam.ac.uk>

plotSilhouette	<i>Plot silhouette</i>
----------------	------------------------

Description

Plot average silhouette.

Usage

```
plotSilhouette(sil, fileName)
```

Arguments

sil	vector of the average silhouette for K from 2 to some value maxK
fileName	name of the png file

Author(s)

Alessandra Cabassi <ac2051@cam.ac.uk>

plotSimilarityMatrix	<i>Plot similarity matrix</i>
----------------------	-------------------------------

Description

Plot similarity matrix. It is possible to plot a side vector that corresponds to a response variable Y and to order the rows (and columns) according to some clustering structure specified by the variable clusLabels.

Usage

```
plotSimilarityMatrix(X, y = NULL, clusLabels = NULL, colX = NULL,
  colY = NULL, myLegend = NULL, file_name = "myheatmap.png",
  savePNG = FALSE, saveTIKZ = FALSE, saveEPS = FALSE,
  semiSupervised = FALSE, scale = "none", labRow = NA, labCol = NA,
  dendro = "none")
```

Arguments

X	matrix
y	response
clusLabels	cluster labels
colX	colours for the matrix
colY	colours for the response
myLegend	vector of strings with the names of the variables
file_name	name of the file
savePNG	boolean flag. if true, saves the plot as a png file
saveTIKZ	boolean flag. if true saves the plot as a tikz file
semiSupervised	boolean flag. if true, the response is plotted next to the matrix.
scale	parameter scale of heatmap.2. can be "none" or "columns"
labRow	vector of row labels. default NA.
labCol	vector of column labels. default NA.

spectrumShift	<i>Spectrum shift</i>
---------------	-----------------------

Description

Make a symmetric matrix positive semi-definite

Usage

```
spectrumShift(kernelMatrix, coeff = 1.2, verbose = FALSE)
```

Arguments

kernelMatrix	symmetric matrix
--------------	------------------

Author(s)

Alessandra Cabassi <ac2051@cam.ac.uk>

Index

coca, [1](#)
consensusCluster, [2](#)
copheneticCorrelation, [3](#)

kkmeansTrain, [3](#)

lmkkmeansTrain, [4](#)

maximiseSilhouette, [4](#)

plotSilhouette, [5](#)
plotSimilarityMatrix, [5](#)

spectrumShift, [6](#)