Package ‘clustrd’

February 19, 2015

Title Methods for joint dimension reduction and clustering

Description A collection of methods for joint dimension reduction and clustering

Version 0.1.2

Maintainer Angelos Markos <amarkos@gmail.com>

License GPL

Depends corpcor, e1071, ggplot2, irlba

Collate 'MCAk.r' 'disjMake.r' 'fuzzyMCAk.r' 'groupals.r' 'iFCB.r'
 'FactorialKM.r' 'ReducedKM.r' 'plotrd.r' 'ed.r' 'nrm.r'
 'orth.r' 'rosplit.r'

Author Angelos Markos [aut, cre],
 Alfonso Iodice D’ Enza [aut],
 Michel Van de Velden [aut]

NeedsCompilation no

Repository CRAN

Date/Publication 2013-12-27 16:04:14

R topics documented:

  FactorialKM ................................................................. 2
  fuzzyMCAk ............................................................... 3
  groupals ................................................................. 4
  iFCB ..................................................................... 5
  macro ................................................................. 6
  MCAk ................................................................. 7
  plotrd ................................................................. 8
  ReducedKM ............................................................ 9
  underwear ............................................................ 10

Index 11
Description

Implements Factorial k-means (Vichi and Kiers, 2001) which combines Principal Component Analysis for dimension reduction with k-means for clustering.

Usage

```
FactorialKM(data, nclus, ndim, nstart = 100, smartStart = FALSE)
```

Arguments

data: quantitative dataset
nclus: number of clusters
ndim: dimensionality of the solution
nstart: number of starts
smartStart: If TRUE then starting values are obtained with k-means

Value

- obscoord: object scores
- attcoord: attribute scores
- centroid: cluster centroids
- cluID: cluster membership
- criterion: optimal value of the objective function

Author(s)

Markos, A. <amarkos@gmail.com>, Iodice D’Enza, A. <iodicede@gmail.com> and Van de Velden, M. <vandevelden@ese.eur.nl>

References


See Also

ReducedKM

Examples

```
data(macro)
macro = data.frame(scale(macro, center = TRUE, scale = TRUE))
outf <- FactorialKM(macro,3,2,nstart=1,smartStart=TRUE)
plotrd(outf,what=c("all","none"),obslable=rownames(macro),density=FALSE)
```
Description

Implements Fuzzy Cluster MCA (Hwang, Dillon and Takane, 2010) which combines Multiple Correspondence Analysis for dimension reduction with fuzzy c-means (Bezdek, 1981) for clustering.

Usage

fuzzyMCAk(data, nclus=3, ndim=2, nstart=1)

Arguments

data: categorical dataset
nclus: number of clusters
ndim: dimensionality of the solution
nstart: number of random starts

Value

obscoord: object scores
atrcoord: attribute scores
centroid: cluster centroids
cluID: hard cluster membership
U: fuzzy cluster membership
FPI: Fuzziness Performance Index
MPE: Modified Partition Entropy
criterion: optimal value of the objective function

Author(s)

Markos, A. <amarkos@gmail.com>, Iodice D’Enza, A. <iodicede@gmail.com> and Van de Velden, M. <vandevelden@ese.eur.nl>

References


See Also

groupals, ifcb, MCAk
Examples

```r
data(underwear)
outfMCAk <- fuzzyMCAk(underwear[c(1:200),c(1:2)],nclus=3,ndim=2,nstart=1)
plotrd(outfMCAk)
```

Description

Implements a constrained homogeneity analysis to take into account the cluster structure of objects as described in Van Buuren and Heiser (1989).

Usage

```r
groupals(data,nclus,ndim,nstart=100,smartStart=F,seed=1234)
```

Arguments

data: categorical dataset
nclus: number of clusters
ndim: dimensionality of the solution
nstart: number of random starts
smartStart: If TRUE then starting values are obtained with fuzzy c-means
seed: seed is used to set the random number seed when smartStart = FALSE

Value

- obscoord: object scores
- attcoord: attribute scores
- centroid: cluster centroids
- cluID: cluster membership
- criterion: optimal value of the objective function

Author(s)

Markos, A. <amarkos@gmail.com>, Iodice D’Enza, A. <iodicede@gmail.com> and Van de Velden, M. <vandevelden@ese.eur.nl>

References

**iFCB**

**See Also**

MCAk, fuzzyMCAk, iFCB

**Examples**

```r
data(underwear)
attlab = c(c(1:15), "by", "tr", "vm", "jd", "ml", "bn", "bg", "ck", "a1", "a2", "a3")
outgroupals <- groupals(underwear, nclus=3, ndim=2, nstart=1, smartStart=TRUE, seed=1234)
plotrd(outgroupals, attlabel=attlab)
```

**Description**

Implements iFCB (Iodice D’Enza and Palumbo, 2013) which combines Nonsymmetric Correspondence Analysis for dimension reduction with k-means for clustering.

**Usage**

```r
iFCB(data, nclus, ndim, nstart=100, smartStart=F, seed=1234)
```

**Arguments**

- `data`: categorical dataset
- `nclus`: number of clusters
- `ndim`: dimensionality of the solution
- `nstart`: number of random starts
- `smartStart`: If TRUE then starting values are obtained with k-means
- `seed`: seed is used to set the random number seed when smartStart = FALSE

**Value**

- `obscoord`: object scores
- `attcoord`: attribute scores
- `centroid`: cluster centroids
- `cluID`: cluster membership
- `criterion`: optimal value of the objective function

**Author(s)**

Markos, A. <amarkos@gmail.com>, Iodice D’Enza, A. <iodicede@gmail.com> and Van de Velden, M. <vandevelden@ese.eur.nl>
References


See Also

MCak, fuzzyMCak, groupals

Examples

data(underwear)
attlab = c(1:15, "by", "tr", "vm", "jd", "ml", "bn", "bg", "ck", "a1", "a2", "a3")
outifCB <- iFCB(underwear, nclus=3, ndim=2, nstart=1, smartStart=TRUE, seed=1234)
plotrd(outifCB, atlabel=attlab)

Economic Indicators of 20 OECD countries for 1999

Description

Data on the macroeconomic performance of national economies of 20 countries, members of the OECD (September 1999). The performance of the economies reflects the interaction of six main economic indicators: (GDP), leading indicator (LI), unemployment rate (UR), interest rate (IR), trade balance (TB), net national savings (NNS).

Usage

data(macro)

Format

A data frame with 20 observations on the following 6 variables.

GDP numeric
LI numeric
UR numeric
IR numeric
TB numeric
NNS numeric

Source

Vichi and Kiers 2001

References

Description

Implements Cluster MCA (Hwang, Dillon and Takane, 2006) which combines Multiple Correspondence Analysis for dimension reduction with k-means for clustering.

Usage

MCAk(data, nclus, ndim, nstart=100, smartStart=F, seed=1234)

Arguments

data            categorical dataset
nclus            number of clusters
ndim             dimensionality of the solution
nstart           number of random starts
smartStart       If TRUE then starting values are obtained with k-means
seed             seed is used to set the random number seed when smartStart = FALSE

Value

obscoord         object scores
attcoord         attribute scores
centroid         cluster centroids
cluID            cluster membership
criterion        optimal value of the objective function

Author(s)

Markos, A. <amarkos@gmail.com>, Iodice D’Enza, A. <iodicede@gmail.com> and Van de Velden, M. <vandevelden@ese.eur.nl>

References


See Also

groupals, iFCB, fuzzyMCAk
Examples

```r
data(underwear)
atlab = c(1:15, "by", "tr", "vm", "jd", "m1", "bn", "bg", "ck", "a1", "a2", "a3")
outMCak <- MCak(underwear, nclus=3, ndim=2, nstart=1, smartStart=TRUE, seed=1234)
plotrd(outMCak, attlabel = attlab)
```

Description

Plot function that creates a ggplot2 based density map of the object scores and a scatter plot of both the attribute scores and the centroids. Creates a .png file in the working directory.

Usage

```r
plotrd(clustrdOut, what=c("all", "all"), obslabel = 0, attlabel = 0, density = T, fname = 0)
```

Arguments

- `clustrdOut`: list object created by one of the methods.
- `what`: Vector of two character strings specifying the contents of the plot. First entry sets the rows and the second entry the columns. Allowed values are "all" (all available points, default) "none" (no points are displayed).
- `obslabel`: the object labels; if not provided, no labeling is applied.
- `attlabel`: the attribute labels; if not provided, default labeling is applied.
- `density`: if TRUE a density map of the objects is created.
- `fname`: name of the png file to be created. Default file name is "figure".

Value

a ggplot object

Author(s)

Markos, A. <amarkos@gmail.com>, Iodice D’Enza, A. <iodicede@gmail.com> and Van de Velden, M. <vandevelden@ese.eur.nl>
Description

Implements Reduced k-means (De Soete and Carroll, 1994) which combines k-means for clustering with PCA for dimension reduction.

Usage

ReducedKM(data, nclus, ndim, nstart = 100, smartStart = FALSE)

Arguments

data  quantitative dataset
nclus number of clusters
ndim  dimensionality of the solution
nstart number of starts
smartStart   If TRUE then starting values are obtained with k-means

Value

obscoord  object scores
attdcoord  variable loadings
centroid  cluster centroids
cluID cluster membership
criterion  optimal value of the objective function

Author(s)

Markos, A. <amarkos@gmail.com>, Iodice D’Enza, A. <iodicede@gmail.com> and Van de Velden, M. <vandevelden@ese.eur.nl>

References


See Also

FactorialKM
Examples

```r
data(macro)
macro = data.frame(scale(macro, center = TRUE, scale = TRUE))
outr <- ReducedKM(macro,3,2,nstart=1,smartStart=TRUE)
plotrd(outr,what=c("all","none"),obslabel=rownames(macro),density=FALSE)
```

<table>
<thead>
<tr>
<th>underwear</th>
<th>South Korean Underwear</th>
</tr>
</thead>
</table>

Description

The dataset comes from a large survey conducted by a South Korean underwear manufacturer in 1997 (Yang, 1997). 664 South Korean consumers were asked to provide responses for three multiple-choice items: preferred brand of underwear (8 brands), attributes when considering a brand of underwear to purchase (15 attributes) and consumer age (6 levels).

Usage

```r
data(underwear)
```

Format

A data frame with 664 observations on the following 3 variables.

- brand categorical
- atts categorical
- age categorical

Source

Yang 1997

References


Examples

```r
data(underwear)
```
Index

*Topic clustering
  FactorialKM, 2
  fuzzyMCAk, 3
  groupals, 4
  iFCB, 5
  MCAk, 7
  plotrd, 8
  ReducedKM, 9

*Topic cluster
  FactorialKM, 2
  fuzzyMCAk, 3
  groupals, 4
  iFCB, 5
  MCAk, 7
  plotrd, 8
  ReducedKM, 9

*Topic datasets
  macro, 6
  underwear, 10

FactorialKM, 2, 9
fuzzyMCAk, 3, 5–7
groupals, 3, 4, 6, 7
iFCB, 3, 5, 6, 7

macro, 6
MCAk, 3, 5, 6, 7
plotrd, 8
ReducedKM, 2, 9
underwear, 10