Package 'RcmdrPlugin.FuzzyClust'

September 4, 2016

Type Package

Title R Commander Plug-in for Fuzzy Clustering Methods (Fuzzy C-Means and Gustafson Kessel)

Version 1.1

Date 2016-09-03

Author Achmad Fauzi Bagus F <uzi.fauz@gmail.com>, Setia Pramana

Maintainer Achmad Fauzi Bagus F <uzi.fauz@gmail.com>

Description The R Commander Plug-in for Fuzzy Clustering Methods. This Plug-in provide Graphical User Interface of 2 methods of Fuzzy Clustering (Fuzzy C-Means /FCM and Gustafson Kessel-Babuska). For validation of clustering, this plug-in use Xie Beni Index, MPC index, and CE index. For statistical test (test of significant differences of grouping/clustering), this plug-in use MANOVA analysis with Pillai trace statistics. For stabilize the result, this package provide soft voting cluster ensemble function. Visualization of result are provided via plug-in that must be load in Rcmdr file.

Depends R (>= 3.2.5)

Imports Rcmdr, doParallel, tcltk2, foreach, clue, ggplot2, MASS, reshape2, tkrplot, iterators, parallel

Suggests knitr, rmarkdown

License GPL-2

LazyData TRUE

RoxygenNote 5.0.1

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2016-09-04 09:36:21

R topics documented:

`biploting` *Biploting Cluster Result*

Description

Biploting Cluster Result

Usage

```
biploting(cluster)
```

Arguments

`cluster`	a cluster object

Details

Make Visualization Biplot from cluster analysis result

Value

biplot a biplot

Examples

```
library(RcmdrPlugin.FuzzyClust)
fuzzy.CM(X=iris[,1:4],K = 3,m = 2,RandomNumber = 1234)->cl
biploting(cl)
```

checkManova *MANOVA analysis of cluster*

Description

MANOVA analysis based on Pillai Statistic

Usage

```
checkManova(clust)
```

Arguments

clust	cluster object

Value

statistic of MANOVA

Examples

```
library(RcmdrPlugin.FuzzyClust)
fuzzy.CM(X=iris[,1:4],K = 3,m = 2,RandomNumber = 1234)->cl
checkManova(cl)
```

data.gen1 *Data Generate 1*

Description

A dataset containing generated data for simulation

Usage

```
data.gen1
```

Format

A data frame with 120 rows and 4 variables:

V1 Variable 1

V2 Variable 2

V3 Variable 3

LABEL Labeling factor

Source

generated randomly

data.gen2 *Data Generate 2*

Description

A dataset containing generated data for simulation

Usage

```
data.gen2
```

Format

A data frame with 40 rows and 4 variables:

v1 Variable 1

V2 Variable 2

V3 Variable 3

LABEL Labeling factor

Source

generated multiariate random

data.gen3 *Data Generate 3*

Description

A dataset containing generated data for simulation

Usage

```
data.gen3
```

Format

A data frame with 30 rows and 3 variables:

v1 Variable 1

V2 Variable 2

LABEL Labeling factor

Source

generated randomly

data.gen4 *Data Generate 4*

Description

A dataset containing generated data for simulation

Usage

```
data.gen4
```

Format

A data frame with 120 rows and 3 variables:

V1 Variable 1

V2 Variable 2

LABEL Labeling factor

Source

generated randomly

EastJava *Data of Education Variables on East Java Indonesia 2014*

Description

A dataset containing the scaled data of Education Variables in Eeast Java, Indonesia 2014

Usage

```
EastJava
```

Format

A data frame with 38 rows and 12 variables:

V1 Proportion of human that illiterate among 100 people, in proportion per 100 people

V2 Expected School Years, in years

V3 Average Years of Schooling, in years

V4 Net Enrollment Rate for Primary School, in proportion per 100 people

V5 Net Enrollment Rate for Secondary School, in proportion per 100 people

V6 Ratio Student per Teacher on Primary School, in proportion

V7 Ratio Student per School on Primary School, in proportion

V8 Ratio Student per Teacher on Secondary School, in proportion

V9 Ratio Student per School on Secondary School, in proportion

V10 Realization of Goverment Budget on Education. in percent

V11 Drop out rate on Primary School, in proportion per 100 people

V12 Drop out rate on Secondary School, in proportion per 100 people

Source

http://bps.go.id/

`fuzzy.CM` *Fuzzy C-Means*

Description

This function used to perform Fuzzy C-Means of X dataset.

Usage

```
fuzzy.CM(X, K = 2, m = 2, max.iteration = 100, threshold = 10^-5,
  RandomNumber = 0)
```

Arguments

`X`	data frame n x p
`K`	specific number of cluster (must be >1)
`m`	fuzzifier / degree of fuzziness
`max.iteration`	maximum iteration to convergence
`threshold`	threshold of convergence
`RandomNumber`	specific seed

Details

This function perform Fuzzy C-Means algorithm by Bezdek (1981). Fuzzy C-Means is one of fuzzy clustering methods to clustering dataset become K cluster. Number of cluster (K) must be greater than 1. To control the overlaping or fuzziness of clustering, parameter m must be specified. Maximum iteration and threshold is specific number for convergencing the cluster. Random Number is number that will be used for seeding to firstly generate fuzzy membership matrix.

Clustering will produce fuzzy membership matrix (U) and fuzzy cluster centroid (V). The greatest value of membership on data point will determine cluster label. Centroid or cluster center can be use to interpret the cluster. Both membership and centroid produced by calculating mathematical distance. Fuzzy C-Means calculate distance with Euclideans norm. So it can be said that cluster will have sperichal shape of geometry.

Value

func.obj objective function that calculated.

U matrix n x K consist fuzzy membership matrix

V matrix K x p consist fuzzy centroid

D matrix n x K consist distance of data to centroid that calculated

Clust.desc cluster description (dataset with additional column of cluster label)

References

Balasko, B., Abonyi, J., & Feil, B. (2002). Fuzzy Clustering and Data Analysis Toolbox: For Use with Matlab. Veszprem, Hungary.

Gustafson, D. E., & Kessel, W. C. (1978). Fuzzy Clustering With A Fuzzy Covariance Matrix. 761-766.

Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The Fuzzy C-Means Clustering Algorithm. Computers and Geosciences Vol 10, 191-203

Examples

```
library(RcmdrPlugin.FuzzyClust)
data(iris)
fuzzy.CM(X=iris[,1:4],K = 3,m = 2,RandomNumber = 1234)->cl
```

---

fuzzy.GK *Gustafson Kessel Improved Covariance Estimation*

Description

This function used to perform Gustafson Kessel Clustering of X dataset.

Usage

```
fuzzy.GK(X, K = 2, m = 1.5, max.iteration = 100, threshold = 10^-5,
  RandomNumber = 0, rho = rep(1, K), gamma = 0)
```

Arguments

X	data frame n x p
K	specific number of cluster (must be >1)
m	fuzzifier / degree of fuzziness
max.iteration	maximum iteration to convergence
threshold	threshold of convergence
RandomNumber	specific seed
rho	cluster volume
gamma	tuning parameter of covariance

Details

This function perform Fuzzy C-Means algorithm by Gustafson Kessel (1968) that improved by Babuska et al (2002). Gustafson Kessel (GK) is one of fuzzy clustering methods to clustering dataset become K cluster. Number of cluster (K) must be greater than 1. To control the overlaping or fuzziness of clustering, parameter m must be specified. Maximum iteration and threshold is specific number for convergencing the cluster. Random Number is number that will be used for seeding to firstly generate fuzzy membership matrix.

Clustering will produce fuzzy membership matrix (U) and fuzzy cluster centroid (V). The greatest value of membership on data point will determine cluster label. Centroid or cluster center can be use to interpret the cluster. Both membership and centroid produced by calculating mathematical distance. Fuzzy C-Means calculate distance with Covariance Cluster norm distance. So it can be said that cluster will have both sperichal and elipsodial shape of geometry.

Babuska improve the covariance estimation via tuning covariance cluster with covariance of data. Tuning parameter determine proportion of covariance data and covariance cluster that will be used to estimate new covariance cluster. Beside improving via tuning, Basbuka improve the algorithm with decomposition of covariance so it will become non singular matrix.

Value

func.obj objective function that calculated.

U matrix n x K consist fuzzy membership matrix

V matrix K x p consist fuzzy centroid

D matrix n x K consist distance of data to centroid that calculated

Clust.desc cluster description (dataset with additional column of cluster label)

References

Babuska, R., Veen, P. v., & Kaymak, U. (2002). Improved Covarians Estimation for Gustafson Kessel Clustering. IEEE, 1081-1084.

Balasko, B., Abonyi, J., & Feil, B. (2002). Fuzzy Clustering and Data Analysis Toolbox: For Use with Matlab. Veszprem, Hungary.

Gustafson, D. E., & Kessel, W. C. (1978). Fuzzy Clustering With A Fuzzy Covariance Matrix. 761-766.

Examples

```
library(RcmdrPlugin.FuzzyClust)
data(iris)
fuzzy.GK(X=iris[,1:4],K = 3,m = 2,RandomNumber = 1234,gamma=0, max.iteration=20)->cl
```

hello *Hello function*

Description

Hello

Usage

```
hello()
```

managedata *Preparing data for clustering.*

Description

This function used to construct data for clustering from dataset with chosen variables.

Usage

```
managedata(var.choice)
```

Arguments

var.choice	Chosen Variables of Dataset

Details

Don't use it from user.

Value

data.cluster Dataset with chosen variables

pluginInput *Input Plugin of Fuzzy Clustering on Rcmdr*

Description

Graphical User Interface on Rcmdr Plugin. This Plugin provide Interface to select variables of dataset that will be used for Fuzzy Clustering, methods selection, and parameter specification

Never use it before open Rcmdr. Its preferable to use plugin menu on Rcmdr

Usage

```
pluginInput()
```

radar.plotting *Radar Ploting Cluster Result*

Description

Radar Ploting Cluster Result

Usage

```
radar.plotting(cluster)
```

Arguments

cluster	a cluster object

Details

Make Visualization Radar Ploting from

Value

radarplot a radarplot

Examples

```
library(RcmdrPlugin.FuzzyClust)
fuzzy.CM(X=iris[,1:4],K = 3,m = 2,RandomNumber = 1234)->cl
checkManova(cl)
```

result.GUI *Result GUI*

Description

Result GUI

Usage

```
result.GUI(parent, cluster, valid, manov, method)
```

Arguments

parent	parent window
cluster	cluster object
valid	validation index object
manov	manova object
method	method of clustering

Details

Not run by users

`soft.vote.ensemble` *Soft Voting Cluster Ensemble*

Description

This function used to perform Soft Voting Cluster Ensemble.

Usage

```
soft.vote.ensemble(data, seed, method = "FCM", K = 2, m = 2, gamma = 0,
  rho = rep(1, K), threshold = 10^-5, max.iteration = 100, core)
```

Arguments

`data`	data frame nxp
`seed`	number of ensemble
`method`	fuzzy clustering method that will be used ("FCM" or "GK")
`K`	specific number of cluster (must be >1)
`m`	fuzzifier / degree of fuzziness
`gamma`	parameter of Gustafson Kessel Clustering
`rho`	parameter of volume clustering in Gustafson Kessel Clustering
`threshold`	threshold of convergence
`max.iteration`	maximum iteration to convergence
`core`	number of core that used for parallelization

Details

Soft vote cluster ensemble used to stabilize the result of cluster analysis. It can be define combine several result of clustering to be one robust result.

The simple method of ensemble is voting method, vote label that resulted and use maximum number of voting as partition. For fuzzy clustering, voting method use membership matrix. This function implemented voting method with sum rule approach. For standarize the label, this function use hungary algorithm for optimal labelization.

Value

func.obj objective function that calculated.

U matrix n x K consist fuzzy membership matrix

V matrix K x p consist fuzzy centroid

D matrix n x K consist distance of data to centroid that calculated

Clust.desc cluster description (dataset with additional column of cluster label)

seeding list of random number that used as seeding

Call call argument

References

Sevillano, X., Alias, F., & Socoro, J. C. (2013). Posisional and Confidence voting-based Consensus Function For Fuzzy Cluster Ensemble. Fuzzy Sets and System, 1-40.

Examples

```
#library(RcmdrPlugin.FuzzyClust)
#soft.vote.ensemble(iris[1:50,1:4],seed=2,method="FCM",core=1,max.iteration=20,threshold=10^-3)->Cl
```

`validation.index` *Validation Index of Fuzzy Clustering*

Description

Validation Index of Fuzzy Clustering

Usage

```
validation.index(cluster)
```

Arguments

`cluster`	Cluster Result from Fuzzy Clustering

Details

This function provide validation index that calculated from fuzzy clustering result. There are 3 index that calculated, Xie Beni, MPC, and CE index. Both three indexes calculated from fuzzy membership and data point.

Xie Beni index calculated compactness and separation of clustering.

The best cluster result can be decided with minimum value of index.

Value

XB.index Xie Beni index

MPC.index Modified Partition Coeeficient

CE.index Classification Entropy

References

Wang, W., & Zhang, Y. (2007). On Fuzzy Cluster Validity Indices. Fuzzy Sets and System, 2095-2117.

Examples

```
library(RcmdrPlugin.FuzzyClust)
fuzzy.CM(X=iris[,1:4],K = 3,m = 2,RandomNumber = 1234)->cl
validation.index(cl)
```

Index