Package ‘MixGHD’

May 28, 2021

Type Package
Title Model Based Clustering, Classification and Discriminant Analysis
   Using the Mixture of Generalized Hyperbolic Distributions
Version 2.3.5
Date 2021-05-27
Maintainer Cristina Tortora <grikris1@gmail.com>
Author Cristina Tortora [aut, cre, cph], Aisha ElSherbiny [com], Ryan P. Browne [aut, cph], Brian C. Franczak [aut, cph], and Paul D. McNicholas [aut, cph], and Donald D. Amos [ctb].
Description Carries out model-based clustering, classification and discriminant analysis using five different models. The models are all based on the generalized hyperbolic distribution. The first model 'MGHD' (Browne and McNicholas (2015) <doi:10.1002/cjs.11246>) is the classical mixture of generalized hyperbolic distributions. The 'MGHFA' (Tortora et al. (2016) <doi:10.1007/s11634-015-0204-z>) is the mixture of generalized hyperbolic factor analyzers for high dimensional data sets. The 'MSGHD' is the mixture of multiple scaled generalized hyperbolic distributions, the 'cMSGHD' is a MSGHD with convex contour plots and the 'MCGHD', mixture of coalesced generalized hyperbolic distributions is a new more flexible model (Tortora et al. (2019)<doi:10.1007/s00357-019-09319-3>. The paper related to the software can be found at <doi:10.18637/jss.v098.i03>.
Imports Bessel,stats, mvtnorm, ghyp, numDeriv, mixture, e1071,cluster, methods
Depends MASS, R (>= 3.1.3)
NeedsCompilation yes
License GPL (>= 2)
Repository CRAN
Date/Publication 2021-05-28 09:30:14 UTC
SystemRequirements GNU make

R topics documented:

   ARI ................................................................. 2
Description

Compares two classifications using the adjusted Rand index (ARI).

Usage

ARI(x=NULL, y=NULL)

Arguments

x A n dimensional vector of class labels.
y A n dimensional vector of class labels.

Details

The ARI has expected value 0 in case of random partition, it is equal to one in case of perfect agreement.

Value

The adjusted Rand index value
banknote

Author(s)
Cristina Tortora
Maintainer: Cristina Tortora <cristina.tortora@sjtu.edu>

References

Examples
```r
##loading banknote data
data(banknote)

##model estimation
res=MGHD(data=banknote[,2:7], G=2 )

#result
ARI(res@map, banknote[,1])
```

---

**banknote**

*Swiss Banknote data*

**Description**
The data set contain 6 measures of 100 genuine and 100 counterfeit Swiss franc banknotes.

**Usage**
```r
data(banknote)
```

**Format**
A data frame with the following variables:

- **Status** the status of the banknote: genuine or counterfeit
- **Length** Length of bill (mm)
- **Left** Width of left edge (mm)
- **Right** Width of right edge (mm)
- **Bottom** Bottom margin width (mm)
- **Top** Top margin width (mm)
- **Diagonal** Length of diagonal (mm)

**References**
bankruptcy  

**Bankruptcy data**

**Description**

The data set contain the ratio of retained earnings (RE) to total assets, and the ratio of earnings before interests and taxes (EBIT) to total assets of 66 American firms recorded in the form of ratios. Half of the selected firms had filed for bankruptcy.

**Usage**

data(bankruptcy)

**Format**

A data frame with the following variables:

- **Y** the status of the firm: 0 bankruptcy or 1 financially sound.
- **RE** ratio
- **EBIT** ratio

**References**


---

**cMSGHD**  

**Convex mixture of multiple scaled generalized hyperbolic distributions (cMSGHD).**

**Description**

Carries out model-based clustering using the convex mixture of multiple scaled generalized hyperbolic distributions. The cMSGHD only allows convex level sets.

**Usage**

cMSGHD(data=NULL,gpar0=NULL,G=2,max.iter=100,label=NULL,eps=1e-2, method="km",scale=TRUE,nr=10, modelSel="AIC")
cMSGHD

Arguments

data  A n x p matrix or data frame such that rows correspond to observations and columns correspond to variables.
gpar0  (optional) A list containing the initial parameters of the mixture model. See the 'Details' section.
G  The range of values for the number of clusters.
max.iter  (optional) A numerical parameter giving the maximum number of iterations each EM algorithm is allowed to use.
label  (optional) A n dimensional vector, if label[i]=k then observation i belongs to group k, if NULL then the data has no known groups.
eps  (optional) A number specifying the epsilon value for the convergence criteria used in the EM algorithms. For each algorithm, the criterion is based on the difference between the log-likelihood at an iteration and an asymptotic estimate of the log-likelihood at that iteration. This asymptotic estimate is based on the Aitken acceleration.
method  (optional) A string indicating the initialization criterion, if not specified kmeans clustering is used. Alternative methods are: hierarchical "hierarchical",random "random", kmedoids "kmedoids", and model based "modelBased"
scale  (optional) A logical value indicating whether or not the data should be scaled, true by default.
 nr  (optional) A number indicating the number of starting value when random is used, 10 by default.
modelSel  (optional) A string indicating the model selection criterion, if not specified AIC is used. Alternative methods are: BIC,ICL, and AIC3

Details

The arguments gpar0, if specified, is a list structure containing at least one p dimensional vector mu, alpha and phi, a pxp matrix gamma, and a px2 matrix cpl containing the vector omega and the vector lambda.

Value

A S4 object of class MixGHD with slots:

index  Value of the index used for model selection (AIC or ICL or BIC or AIC3) for each G, the index used is specified by the user, if not specified AIC is used.
BIC  Bayesian information criterion.
ICL  Integrated completed likelihood.
AIC  Akaike information criterion.
AIC3  Akaike information criterion 3.
gpar  A list of the model parameters
loglik  The log-likelihood values.
map  A vector of integers indicating the maximum a posteriori classifications for the best model.
z  A matrix giving the raw values upon which map is based.
Author(s)

Cristina Tortora, Aisha ElSherbiny, Ryan P. Browne, Brian C. Franczak, and Paul D. McNicholas.
Maintainer: Cristina Tortora <cristina.tortora@sjsu.edu>

References


See Also

MGHD MSGHD

Examples

```r
##Generate random data
set.seed(3)
mu1 <- mu2 <- c(0,0)
Sigma1 <- matrix(c(1,0.85,0.85,1),2,2)
Sigma2 <- matrix(c(1,-0.85,-0.85,1),2,2)
X1 <- mvrnorm(n=150,mu=mu1,Sigma=Sigma1)
X2 <- mvrnorm(n=150,mu=mu2,Sigma=Sigma2)
X <- rbind(X1,X2)
##model estimation
em=cMSGHD(X,G=2,max.iter=30,method="random",nr=2)

#result
plot(em)
```

coef

*Coefficients for objects of class MixGHD*

Description

Coefficients of the estimated model.
## S4 method for signature 'MixGHD'
coef(object)

### Arguments

- **object**: An S4 object of class `MixGHD`.

### Value

The coefficients of the estimated model

### Examples

```r
##loading bankruptcy data
data(bankruptcy)

##model estimation
res=MCGHD(data=bankruptcy[,2:3],G=2,method="kmedoids",max.iter=30)
#coefficients of the model
coeff(res)
```

---

### contourpl

**Contour plot**

### Description

Contour plot for a given set of parameters.

### Usage

`contourpl(input)`

### Arguments

- **input**: An S4 object of class `MixGHD`.

### Value

The contour plot
Author(s)
Cristina Tortora Maintainer: Cristina Tortora <cristina.tortora@sj-su.edu>

Examples

```r
# loading bankruptcy data
data(bankruptcy)

# model estimation
res = MCGHD(data = bankruptcy[,2:3], G = 2, method = "kmedoids", max.iter = 30)
# result
contourpl(res)
```

DA  
Discriminant analysis using the mixture of generalized hyperbolic distributions.

Description
Carries out model-based discriminant analysis using 5 different models: the mixture of multiple scaled generalized hyperbolic distributions (MGHD), the mixture of generalized hyperbolic factor analyzers (MGHFA), the mixture of multiple scaled generalized hyperbolic distributions (MSGHD), the mixture of convex multiple scaled generalized hyperbolic distributions (cMSGHD) and the mixture of coalesced generalized hyperbolic distributions (MCGHD).

Usage

```r
DA(train, trainL, test, testL, method = "MGHD", starting = "km", max.iter = 100, eps = 1e-2, q = 2, scale = TRUE)
```

Arguments

- `train`: A n1 x p matrix or data frame such that rows correspond to observations and columns correspond to variables of the training data set.
- `trainL`: A n1 dimensional vector of membership for the units of the training set. If `trainL[i] = k` then observation belongs to group `k`.
- `test`: A n2 x p matrix or data frame such that rows correspond to observations and columns correspond to variables of the test data set.
- `testL`: A n2 dimensional vector of membership for the units of the test set. If `testL[i] = k` then observation belongs to group `k`.
- `method`: (optional) A string indicating the method to be used for discriminant analysis, if not specified MGHD is used. Alternative methods are: MGHFA, MSGHD, cMSGHD, MCGHD.
- `starting`: (optional) A string indicating the initialization criterion, if not specified kmeans clustering is used. Alternative methods are: hierarchical "hierarchical", random "random", kmedoids "kmedoids", and model based "modelBased"
max.iter (optional) A numerical parameter giving the maximum number of iterations each EM algorithm is allowed to use.

eps (optional) A number specifying the epsilon value for the convergence criteria used in the EM algorithms. For each algorithm, the criterion is based on the difference between the log-likelihood at an iteration and an asymptotic estimate of the log-likelihood at that iteration. This asymptotic estimate is based on the Aitken acceleration.

q (optional) used only if MGHFA method is selected. A numerical parameter giving the number of factors.

scale (optional) A logical value indicating whether or not the data should be scaled, true by default.

**Value**

A list with components

- **model** An S4 object of class `MixGHD` with the model parameters.
- **testMembership** A vector of integers indicating the membership of the units in the test set
- **ARItest** A value indicating the adjusted rand index for the test set.
- **ARItrain** A value indicating the adjusted rand index for the train set.

**Author(s)**

Cristina Tortora, Aisha ElSherbiny, Ryan P. Browne, Brian C. Franczak, and Paul D. McNicholas.

Maintainer: Cristina Tortora <cristina.tortora@sjsu.edu>

**References**


**See Also**

"MixGHD" MGHD MGHFA MSGHD cMSGHD MCGHD ARI MixGHD-class MixGHD

**Examples**

```r
# loading banknote data
data(banknote)
banknote[,1]=as.numeric(factor(banknote[,1]))

# divide the data in training set and test set
train=banknote[c(1:74,126:200),]
test=banknote[75:125,]
```
## model estimation

```r
model=DA(train[,2:7],train[,1],test[,2:7],test[,1],method="MGHD",max.iter=20)
```

# result

```r
model$ARItest
```

---

### dCGHD

**Density of a coalesced generalized hyperbolic distribution (MSGHD).**

---

### Description

Compute the density of a p dimensional coalesced generalized hyperbolic distribution.

### Usage

```r
dCGHD(data,p,mu=rep(0,p),alpha=rep(0,p),sigma=diag(p),lambda=1,omega=1,
      omegav=rep(1,p),lambdav=rep(1,p),wg=0.5,gam=NULL,phi=NULL)
```

### Arguments

- **data**: n x p data set
- **p**: number of variables.
- **mu**: (optional) the p dimensional mean
- **alpha**: (optional) the p dimensional skewness parameter alpha
- **sigma**: (optional) the p x p dimensional scale matrix
- **lambda**: (optional) the 1 dimensional index parameter lambda
- **omega**: (optional) the 1 dimensional concentration parameter omega
- **omegav**: (optional) the p dimensional concentration parameter omega
- **lambdav**: (optional) the p dimensional index parameter lambda
- **wg**: (optional) weight
- **gam**: (optional) the pxp gamma matrix
- **phi**: (optional) the p dimensional vector phi

### Details

The default values are: 0 for the mean and the skweness parameter alpha, diag(p) for sigma, 1 for omega, and 0.5 for lambda.

### Value

A n dimensional vector with the density from a coalesced generalized hyperbolic distribution
Author(s)
Cristina Tortora, Aisha ElSherbiny, Ryan P. Browne, Brian C. Franczak, and Paul D. McNicholas.
Maintainer: Cristina Tortora <cristina.tortora@sjsu.edu>

References

Examples

```r
x = seq(-3,3,length.out=30)
y = seq(-3,3,length.out=30)
xyS1 = matrix(0,nrow=length(x),ncol=length(y))
for(i in 1:length(x)){
  for(j in 1:length(y)){
    xy <- matrix(cbind(x[i],y[j]),1,2)
    xyS1[i,j] = dCGHD(xy,2)
  }
}
contour(x=x,y=y,z=xyS1, levels=c(.005,.01,.025,.05, .1,.25), main="CGHD",ylim=c(-3,3), xlim=c(-3,3))
```

---

dGHD

Density of a generalized hyperbolic distribution (GHD).

Description
Compute the density of a p dimensional generalized hyperbolic distribution.

Usage
dGHD(data,p, mu=rep(0,p),alpha=rep(0,p),sigma=diag(p),omega=1,lambda=0.5, log=FALSE)

Arguments
data n x p data set
p number of variables.
mu (optional) the p dimensional mean
alpha (optional) the p dimensional skewness parameter alpha
sigma (optional) the p x p dimensional scale matrix
omega (optional) the unidimensional concentration parameter omega
lambda (optional) the unidimensional index parameter lambda
log (optional) if TRUE returns the log of the density
Details

The default values are: 0 for the mean and the skweness parameter alpha, diag(p) for sigma, 1 for omega, and 0.5 for lambda.

Value

A n dimensional vector with the density from a generalized hyperbolic distribution

Author(s)

Cristina Tortora, Aisha ElSherbiny, Ryan P. Browne, Brian C. Franczak, and Paul D. McNicholas.
Maintainer: Cristina Tortora <cristina.tortora@sjsu.edu>

References

Canadian Journal of Statistics, 43.2 176-198

Examples

```r
x = seq(-3,3,length.out=50)
y = seq(-3,3,length.out=50)
xyS1 = matrix(0,nrow=length(x),ncol=length(y))
for(i in 1:length(x)){
  for(j in 1:length(y)){
    xy <- matrix(cbind(x[i],y[j]),1,2)
    xyS1[i,j] = dGHD(xy,2)
  }
}
contour(x=x,y=y,z=xyS1, levels=c(.005,.01,.025,.05,.1,.25), main="MGHD",ylim=c(-3,3), xlim=c(-3,3))
```

dMSGHD

Density of a multiple-scaled generalized hyperbolic distribution (MS-GHD).

Description

Compute the density of a p dimensional multiple-scaled generalized hyperbolic distribution.
Usage

dMSGHD(data,p,mu=rep(0,p),alpha=rep(0,p),sigma=diag(p),omegav=rep(1,p),
       lambdav=rep(0.5,p),gam=NULL,phi=NULL,log=FALSE)

Arguments

data n x p data set
p number of variables.
mu (optional) the p dimensional mean
alpha (optional) the p dimensional skewness parameter alpha
sigma (optional) the p x p dimensional scale matrix
omegav (optional) the p dimensional concentration parameter omega
lambdav (optional) the p dimensional index parameter lambda
gam (optional) the pxp gamma matrix
phi (optional) the p dimensional vector phi
log (optional) if TRUE returns the log of the density

Details

The default values are: 0 for the mean and the skewness parameter alpha, diag(p) for sigma, 1 for omega, and 0.5 for lambda.

Value

A n dimensional vector with the density from a multiple-scaled generilzed hyperbolic distribution

Author(s)

Cristina Tortora, Aisha ElSherbiny, Ryan P. Browne, Brian C. Franczak, and Paul D. McNicholas. Maintainer: Cristina Tortora <cristina.tortora@sjsu.edu>

References


Examples

```r
x = seq(-3,3,length.out=50)
y = seq(-3,3,length.out=50)
xyS1 = matrix(0,nrow=length(x),ncol=length(y))
for(i in 1:length(x)){
  for(j in 1:length(y)){
    xy <- matrix(cbind(x[i],y[j]),1,2)
  }
}
```
xyS1[i,j] = dMSGHD(xy,2)
}
}
contour(x=x,y=y,z=xyS1, levels=seq(.005,.25,by=.005), main="MSGHD")

---

**MCGHD**

*Mixture of coalesced generalized hyperbolic distributions (MCGHD)*.

---

**Description**

Carries out model-based clustering using the mixture of coalesced generalized hyperbolic distributions.

**Usage**

```r
MCGHD(data=NULL,gpar0=NULL,G=2,max.iter=100,eps=1e-2,label=NULL,
method="km",scale=TRUE,nr=10, modelSel="AIC")
```

**Arguments**

- **data**
  A n x p matrix or data frame such that rows correspond to observations and columns correspond to variables.

- **gpar0**
  (optional) A list containing the initial parameters of the mixture model. See the 'Details' section.

- **G**
  The range of values for the number of clusters.

- **max.iter**
  (optional) A numerical parameter giving the maximum number of iterations each EM algorithm is allowed to use.

- **eps**
  (optional) A number specifying the epsilon value for the convergence criteria used in the EM algorithms. For each algorithm, the criterion is based on the difference between the log-likelihood at an iteration and an asymptotic estimate of the log-likelihood at that iteration. This asymptotic estimate is based on the Aitken acceleration.

- **label**
  (optional) A n dimensional vector, if label[i]=k then observation i belongs to group k, If label[i]=0 then observation i has no known group, if NULL then the data has no known groups.

- **method**
  (optional) A string indicating the initialization criterion, if not specified kmeans clustering is used. Alternative methods are: hierarchical "hierarchical", random "random", and model based "modelBased"

- **scale**
  (optional) A logical value indicating whether or not the data should be scaled, true by default.

- **nr**
  (optional) A number indicating the number of starting value when random is used, 10 by default.

- **modelSel**
  (optional) A string indicating the model selection criterion, if not specified AIC is used. Alternative methods are: BIC, ICL, and AIC3
Details

The arguments gpar0, if specified, has to be a list structure containing as much element as the number of components G. Each element must include the following parameters: one p dimensional vector \( \mu \), alpha and phi, a pxp matrix gamma, a px2 vector cpl containing the vectors omega and lambda, and a 2-dimensional vector containing the omega0 and lambda0.

Value

A S4 object of class MixGHD with slots:

- **index**
  Value of the index used for model selection (AIC or ICL or BIC or AIC3) for each G, the index used is specified by the user, if not specified AIC is used.

- **BIC**
  Bayesian information criterion.

- **ICL**
  Integrated completed likelihood.

- **AIC**
  Akaike information criterion.

- **AIC3**
  Akaike information criterion 3.

- **gpar**
  A list of the model parameters in the rotated space.

- **loglik**
  The log-likelihood values.

- **map**
  A vector of integers indicating the maximum a posteriori classifications for the best model.

- **par**
  A list of the model parameters.

- **z**
  A matrix giving the raw values upon which map is based.

Author(s)

Cristina Tortora, Aisha ElSherbiny, Ryan P. Browne, Brian C. Franczak, and Paul D. McNicholas.

Maintainer: Cristina Tortora <cristina.tortora@sjsu.edu>

References


C. Tortora, R. P. Browne, A. ElSherbiny, B. C. Franczak, and P. D. McNicholas (2021). Model-Based Clustering, Classification, and Discriminant Analysis using the Generalized Hyperbolic Distribution: MixGHD

R package, *Journal of Statistical Software* 98(3) 1–24, <doi:10.18637/jss.v098.i03>.

See Also

[MGHD](https://example.com), [MSGHD](https://example.com)

Examples

```r
# loading banknote data
data(banknote)

# model estimation
model = MCGHD(banknote[,2:7], G=2, max.iter=20)
```
Mixture of generalized hyperbolic distributions (MGHD).

Description

Carries out model-based clustering and classification using the mixture of generalized hyperbolic distributions.

Usage

MGHD(data=NULL,gpar0=NULL,G=2,max.iter=100,label=NULL,eps=1e-2,
method="kmeans",scale=TRUE,nr=10, modelSel="AIC")

Arguments

data A n x p matrix or data frame such that rows correspond to observations and columns correspond to variables.
gpar0 (optional) A list containing the initial parameters of the mixture model. See the 'Details' section.
G The range of values for the number of clusters.
max.iter (optional) A numerical parameter giving the maximum number of iterations each EM algorithm is allowed to use.
label (optional) A n dimensional vector, if label[i]=k then observation i belongs to group k, If label[i]=0 then observation i has no known group, if NULL then the data has no known groups.
eps (optional) A number specifying the epsilon value for the convergence criteria used in the EM algorithms. For each algorithm, the criterion is based on the difference between the log-likelihood at an iteration and an asymptotic estimate of the log-likelihood at that iteration. This asymptotic estimate is based on the Aitken acceleration.
method (optional) A string indicating the initialization criterion, if not specified kmeans clustering is used. Alternative methods are: hierarchical "hierarchical", random "random", and model based "modelBased" clustering
scale (optional) A logical value indicating whether or not the data should be scaled, true by default.
nr (optional) A number indicating the number of starting value when random is used, 10 by default.
modelSel (optional) A string indicating the model selection criterion, if not specified AIC is used. Alternative methods are: BIC, ICL, and AIC3
Details

The arguments gpars, if specified, is a list structure containing at least one p dimensional vector 
mu, and alpha, a pxp matrix sigma, and a 2 dimensional vector containing omega and lambda.

Value

A S4 object of class MixGHD with slots:

index  Value of the index used for model selection (AIC or ICL or BIC or AIC3) for 
each G, the index used is specified by the user, if not specified AIC is used.

BIC  Bayesian information criterion.

ICL  Integrated completed likelihood.

AIC  Akaike information criterion.

AIC3  Akaike information criterion 3.

gpars  A list of the model parameters.

loglik  The log-likelihood values.

map  A vector of integers indicating the maximum a posteriori classifications for the 
best model.

z  A matrix giving the raw values upon which map is based.

Author(s)

Cristina Tortora, Aisha ElSherbiny, Ryan P. Browne, Brian C. Franczak, and Paul D. McNicholas. 
Maintainer: Cristina Tortora <cristina.tortora@sjsu.edu>

References

Canadian Journal of Statistics, 43.2 176-198

Examples

```r
# loading crabs data
data(crabs)

# model estimation
model = MGHD(data = crabs[,4:8], G = 2)

# result
plot(model)
table(model@map, crabs[,2])

# Classification
# loading bankruptcy data
data(bankruptcy)
```
#70% belong to the training set
label=bankruptcy[,1]
#for a Classification porpuse the label cannot be 0
label[1:33]=2
a=round(runif(20)*65+1)
label[a]=0

##model estimation
model=MGHD(data=bankruptcy[,2:3], G=2, label=label )

#result
table(model@map,bankruptcy[,1])
plot(model)

---

MGHFA

*Mixture of generalized hyperbolic factor analyzers (MGHFA).*

**Description**

Carries out model-based clustering and classification using the mixture of generalized hyperbolic factor analyzers.

**Usage**

```r
MGHFA(data=NULL, gpar0=NULL, G=2, max.iter=100, label =NULL ,q=2,eps=1e-2 , method="kmeans", scale=TRUE ,nr=10)
```

**Arguments**

- `data` A matrix or data frame such that rows correspond to observations and columns correspond to variables.
- `gpar0` (optional) A list containing the initial parameters of the mixture model. See the 'Details' section.
- `G` The range of values for the number of clusters.
- `max.iter` (optional) A numerical parameter giving the maximum number of iterations each EM algorithm is allowed to use.
- `label` (optional) A n dimensional vector, if label[i]=k then observation i belongs to group k, If label[i]=0 then observation i has no known group, if NULL then the data has no known groups.
- `q` The range of values for the number of factors.
- `eps` (optional) A number specifying the epsilon value for the convergence criteria used in the EM algorithms. For each algorithm, the criterion is based on the difference between the log-likelihood at an iteration and an asymptotic estimate of the log-likelihood at that iteration. This asymptotic estimate is based on the Aitken acceleration.
method (optional) A string indicating the initialization criterion, if not specified kmeans clustering is used. Alternative methods are: hierarchical "hierarchical" and model based "modelBased" clustering

scale (optional) A logical value indicating whether or not the data should be scaled, true by default.

nr (optional) A number indicating the number of starting value when random is used, 10 by default.

Details

The arguments gpars, if specified, is a list structure containing at least one p dimensional vector mu, alpha and phi, a pxp matrix gamma, a 2 dimensional vector cpl containing omega and lambda.

Value

A S4 object of class MixGHD with slots:

Index Bayesian information criterion value for each combination of G and q.

BIC Bayesian information criterion.

gpar A list of the model parameters.

loglik The log-likelihood values.

map A vector of integers indicating the maximum a posteriori classifications for the best model.

z A matrix giving the raw values upon which map is based.

Author(s)

Cristina Tortora, Aisha ElSherbiny, Ryan P. Browne, Brian C. Franczak, and Paul D. McNicholas. Maintainer: Cristina Tortora <cristina.tortora@sjsu.edu

References


Examples

## Classification
#70% belong to the training set
data(sonar)
label=sonar[,61]
set.seed(4)
a=round(runif(62)*207+1)
label[a]=0
### model estimation

```r
model <- MGHFA(data=sonar[,1:60], G=2, max.iter=25, q=2, label=label)
```

```r
# result
table(model@map, sonar[,61])
summary(model)
```

---

**MixGHD class**

**Class "MixGHD"**

**Description**

This class pertains to results of the application of function `MGHD`, `MSGHD`, `cMSGHD`, `MCGHD`, and `MGHFA`.

**Objects from the Class**

Objects can be created as a result to a call to `MGHD`, `MSGHD`, `cMSGHD`, `MCGHD`, and `MGHFA`.

**Slots**

- `index` Value of the index used for model selection (AIC or ICL or BIC or AIC3) for each G, the index used is specified by the user, if not specified AIC is used.
- `BIC` Bayesian information criterion value.
- `ICL` ICL index.
- `AIC` AIC index.
- `AIC3` AIC3 index.
- `gpar` A list of the model parameters (in the rotated space for MCGHD).
- `loglik` The log-likelihood values.
- `map` A vector of integers indicating the maximum a posteriori classifications for the best model.
- `par` Only for MCGHD. A list of the model parameters.
- `z` A matrix giving the raw values upon which map is based.

**Methods**

- `plot` signature(x = "MixGHD") Provides plots of `MixGHD-class` by plotting the following elements:
  - the value of the log likelihood for each iteration.
  - Scatterplot of the data of all the possible couples of coordinates coloured according to the cluster. Only for less than 10 variables.
  - If the number of variables is two: scatterplot and contour plot of the data coloured according to the cluster.
summary summary(x = "MixGHD").
Provides a summary of MixGHD-class objects by printing the following elements:

- The number components used for the model
- BIC;
- AIC;
- AIC3;
- ICL;
- A table with the number of element in each cluster.

Author(s)

Cristina Tortora, Aisha ElSherbiny, Ryan P. Browne, Brian C. Franczak, and Paul D. McNicholas.
Maintainer: Cristina Tortora <cristina.tortora@sjsu.edu>

See Also

MixGHD-class

Examples

```r
# loading bankruptcy data
data(bankruptcy)

# model estimation
# res = MCGHD(data = bankruptcy[, 2:3], G = 2, method = "kmedoids", max.iter = 30)
# result
# plot(res)
# summary(res)
```

MixGHD-class

Class MixGHD.

Description

This class pertains to results of the application of function MGH, MCGH, MSGH, cMSGH.

Details

Plot the log likelihood value for each iteration of the EM algorithm. If p=2 it shows a contour plot. If 2<p<10 shows a splom of the data colored according to the cluster membership.
MixGHD-class

Slots

  Index  Bayesian information criterion value for each combination of G and q.
  BIC    Bayesian information criterion value.
  gpar   A list of the model parameters.
  loglik The log-likelihood values.
  map    A vector of integers indicating the maximum a posteriori classifications for the best model.
  z      A matrix giving the raw values upon which map is based.
  method A string indicating the used method: MGHD, MGHFA, MSGHD, cMSGHD, MCGHD.
  data   A matrix or data frame such that rows correspond to observations and columns correspond to variables.
  par    (only for MCGHD) A list of the model parameters in the rotated space.

Methods

  signature(x = "MixGHD", y = "missing") S4 method for plotting objects of MixGHD-class.

Author(s)

  Cristina Tortora, Aisha ElSherbiny, Ryan P. Browne, Brian C. Franczak, and Paul D. McNicholas.
  Maintainer: Cristina Tortora <cristina.tortora@sjsu.edu>

See Also

  MixGHD-class, MGHD, MCGHD, MSGHD, cMSGHD, MGHFA

Examples

  # loading banknote data
data(bankruptcy)

  # model estimation
model=MSGHD(bankruptcy[,2:3],G=2,max.iter=30)

  # result
summary(model)
plot(model)
**MSGHD**

*Mixture of multiple scaled generalized hyperbolic distributions (MSGHD).*

---

**Description**

Carries out model-based clustering using the mixture of multiple scaled generalized hyperbolic distributions.

**Usage**

```r
MSGHD(data=NULL, gpar0=NULL, G=2, max.iter=100, label=NULL, eps=1e-2, method="km", scale=TRUE, nr=10, modelSel="AIC")
```

**Arguments**

- `data`: A n x p matrix or data frame such that rows correspond to observations and columns correspond to variables.
- `gpar0`: (optional) A list containing the initial parameters of the mixture model. See the 'Details' section.
- `G`: The range of values for the number of clusters.
- `max.iter`: (optional) A numerical parameter giving the maximum number of iterations each EM algorithm is allowed to use.
- `label`: (optional) A n dimensional vector, if label[i]=k then observation i belongs to group k, If label[i]=0 then observation i has no known group, if NULL then the data has no known groups.
- `eps`: (optional) A number specifying the epsilon value for the convergence criteria used in the EM algorithms. For each algorithm, the criterion is based on the difference between the log-likelihood at an iteration and an asymptotic estimate of the log-likelihood at that iteration. This asymptotic estimate is based on the Aitken acceleration.
- `method`: (optional) A string indicating the initialization criterion, if not specified kmeans clustering is used. Alternative methods are: hierarchical "hierarchical", random "random", and model based "modelBased" clustering.
- `scale`: (optional) A logical value indicating whether or not the data should be scaled, true by default.
- `nr`: (optional) A number indicating the number of starting value when random is used, 10 by default.
- `modelSel`: (optional) A string indicating the model selection criterion, if not specified AIC is used. Alternative methods are: BIC, ICL, and AIC3.

**Details**

The arguments gpar0, if specified, is a list structure containing at least one p dimensional vector mu, alpha and phi, a p x p matrix gamma, and a p x 2 matrix cpl containing the vector omega and the vector lambda.
Value

A S4 object of class `MixGHD` with slots:

- `index` Value of the index used for model selection (AIC or ICL or BIC or AIC3) for each G, the index used is specified by the user, if not specified AIC is used.
- `BIC` Bayesian information criterion.
- `ICL` Integrated completed likelihood.
- `AIC` Akaike information criterion.
- `AIC3` Akaike information criterion 3.
- `gpar` A list of the model parameters
- `loglik` The log-likelihood values.
- `map` A vector of integers indicating the maximum a posteriori classifications for the best model.
- `z` A matrix giving the raw values upon which map is based.

Author(s)

Cristina Tortora, Aisha ElSherbiny, Ryan P. Browne, Brian C. Franczak, and Paul D. McNicholas.
Maintainer: Cristina Tortora <cristina.tortora@sjsu.edu>

References


See Also

`MGHD`

Examples

```r
#loading banknote data
data(banknote)

#model estimation
model=MSGHD(banknote[,2:7],G=2,max.iter=30)

#result
table(banknote[,1],model@map)
summary(model)
plot(model)
```
Description
Plots the loglikelihood function and for p<10 shows the splom of the data.

Usage
## S4 method for signature 'MixGHD'
plot(x,y)

Arguments
x A object of MixGHD-class;
y Not used; for compatibility with generic plot.

Details
Plot the loglikelihood vale for each iteration of the EM algorithm. If p=2 it shows a contour plot. If 2<p<10 shows a splom of the data colored according to the cluster membership.

Methods
signature(x = "MixGHD", y = "missing") S4 method for plotting objects of MixGHD-class.

Author(s)
Cristina Tortora. Maintainer: Cristina Tortora <cristina.tortora@sjsu.edu>

See Also
MixGHD-class,MGHD,MCGHD,MSGHD,cMSGHD,MGHFA

Examples
##loading banknote data
data(bankruptcy)

##model estimation
model=MCGHD(bankruptcy[,2:3],G=2,max.iter=30)

#result
plot(model)
predict | Membership prediction for objects of class MixGHD

Description

Cluster membership

Usage

```r
## S4 method for signature 'MixGHD'
predict(object)
```

Arguments

- `object`: An S4 object of class `MixGHD`.

Value

The cluster membership

Author(s)

Cristina Tortora
Maintainer: Cristina Tortora <cristina.tortora@sjsu.edu>

Examples

```r
##loading bankruptcy data
data(bankruptcy)

##model estimation
res=MCGHD(data=bankruptcy[,2:3],G=2,method="kmedoids",max.iter=30)
#coefficients of the model
predict(res)
```

---

rCGHD | Pseudo random number generation from a coalesced generalized hyperbolic distribution (MSGHD).

Description

Generate n pseudo random numbers from a p dimensional coalesced generalized hyperbolic distribution.
rCGHD

Usage

rCGHD(n,p,mu=rep(0,p),alpha=rep(0,p),sigma=diag(p),omega=1,lambda=0.5
,omegav=rep(1,p),lambdav=rep(0.5,p),wg=0.5)

Arguments

n       number of observations.
p       number of variables.
mu      (optional) the p dimensional mean
alpha   (optional) the p dimensional skewness parameter alpha
sigma   (optional) the p x p dimensional scale matrix
lambda  (optional) the 1 dimensional index parameter lambda
omega   (optional) the 1 dimensional concentration parameter omega
omegav  (optional) the p dimensional concentration parameter omega
lambdav (optional) the p dimensional index parameter lambda
wg      (optional) the weight

Details

The default values are: 0 for the mean and the skewness parameter alpha, diag(p) for sigma, 1 for omega, and 0.5 for lambda.

Value

A n times p matrix of numbers pseudo randomly generated from a coalesced generalized hyperbolic distribution.

Author(s)

Cristina Tortora, Aisha ElSherbiny, Ryan P. Browne, Brian C. Franczak, and Paul D. McNicholas. Maintainer: Cristina Tortora <cristina.tortora@sjtu.edu>

References


Examples

data=rCGHD(300,2,alpha=c(2,-2),omegav=c(2,2),omega=3)

plot(data)
Pseudo random number generation from a generalized hyperbolic distribution (GHD).

**Description**

Generate n pseudo random numbers from a p dimensional generalized hyperbolic distribution.

**Usage**

```r
rGHD(n, p, mu=rep(0, p), alpha=rep(0, p), sigma=diag(p), omega=1, lambda=0.5)
```

**Arguments**

- `n`: number of observations.
- `p`: number of variables.
- `mu`: (optional) the p dimensional mean
- `alpha`: (optional) the p dimensional skewness parameter alpha
- `sigma`: (optional) the p x p dimensional scale matrix
- `omega`: (optional) the unidimensional concentration parameter omega
- `lambda`: (optional) the unidimensional index parameter lambda

**Details**

The default values are: 0 for the mean and the skweness parameter alpha, diag(p) for sigma, 1 for omega, and 0.5 for lambda.

**Value**

A n times p matrix of numbers psudo randomly generated from a generilzed hyperbolic distribution

**Author(s)**

Cristina Tortora, Aisha ElSherbiny, Ryan P. Browne, Brian C. Franczak, and Paul D. McNicholas. Maintainer: Cristina Tortora <cristina.tortora@sjsu.edu>

**References**


**Examples**

```r
data=rGHD(300,2, alpha=c(2,-2))
plot(data)
```
rMSGHD

Pseudo random number generation from a multiple-scaled generalized hyperbolic distribution (MSGHD).

Description

Generate n pseudo random numbers from a p dimensional multiple-scaled generalized hyperbolic distribution.

Usage

rMSGHD(n, p, mu=rep(0, p), alpha=rep(0, p), sigma=diag(p), omegav=rep(1, p), lambdav=rep(0.5, p))

Arguments

- **n**: number of observations.
- **p**: number of variables.
- **mu**: (optional) the p dimensional mean.
- **alpha**: (optional) the p dimensional skewness parameter alpha.
- **sigma**: (optional) the p x p dimensional scale matrix.
- **omegav**: (optional) the p dimensional concentration parameter omega.
- **lambdav**: (optional) the p dimensional index parameter lambda.

Details

The default values are: 0 for the mean and the skewness parameter alpha, diag(p) for sigma, 1 for omega, and 0.5 for lambda.

Value

A n times p matrix of numbers pseudo randomly generated from a generalized hyperbolic distribution.

Author(s)

Cristina Tortora, Aisha ElSherbiny, Ryan P. Browne, Brian C. Franczak, and Paul D. McNicholas. Maintainer: Cristina Tortora <cristina.tortora@sjsu.edu>

References


Examples

data=rMSGHD(300, 2, alpha=c(2, -2), omegav=c(2, 2))

plot(data)
sonar

Sonar data

Description
The data report the patterns obtained by bouncing sonar signals at various angles and under various
conditions. There are 208 patterns in all, 111 obtained by bouncing sonar signals off a metal cylinder
and 97 obtained by bouncing signals off rocks. Each pattern is a set of 60 numbers (variables) taking
values between 0 and 1.

Usage
data(sonar)

Format
A data frame with 208 observations and 61 columns. The first 60 columns contain the variables.
The 61st column gives the material: 1 rock, 2 metal.

Source
UCI machine learning repository

References
R.P. Gorman and T. J. Sejnowski (1988) Analysis of hidden units in a layered network trained to
classify sonar targets. Neural Networks 1: 75-89

summary
Plot objects of class MixGHD.

Description
Methods for function summary aimed at summarizing the S4 classes included in the MixGHD-package

Arguments
object A object of MixGHD-class.

Methods
signature(object = "MixGHD") S4 method for summarying objects of MixGHD-class.

Author(s)
Cristina Tortora. Maintainer: Cristina Tortora <cristina.tortora@sjsu.edu>
See Also

MixGHD MixGHD-class, MGHD, MCGHD, MSGHD, cMSGHD, MGHFA

Examples

```r
## loading banknote data
data(bankruptcy)

## model estimation
model = MSGHD(bankruptcy[, 2:3], G = 2, max.iter = 30)

# result
summary(model)
```
# Index

* Classification  
  MGHD, 16  
  MGHFA, 18  

* Clustering  
  cMSGHD, 4  
  DA, 8  
  MCGHD, 14  
  MGHD, 16  
  MGHFA, 18  
  MixGHD-class, 21  
  MSGHD, 23  
  plot, 25  
  summary, 30  

* Coalesced Generalized hyperboilc distribution  
  dCGHD, 10  
  rCGHD, 26  

* Generalized hyperboilc distribution  
  cMSGHD, 4  
  DA, 8  
  dGHD, 11  
  MCGHD, 14  
  MGHD, 16  
  MGHFA, 18  
  MSGHD, 23  
  rGHD, 28  

* Multiple-scaled Generalized hyperboilc distribution  
  dMSGHD, 12  
  rMSGHD, 29  

* Plot  
  MixGHD-class, 21  
  plot, 25  

* data sets  
  banknote, 3  
  bankruptcy, 4  
  sonar, 30  

* methods  
  summary, 30  

ARI, 2, 9  
banknote, 3  
bankruptcy, 4  
cMSGHD, 4, 9, 20–22, 25, 31  
coef, 6  
coef, MixGHD, missing-method ( coef ), 6  
coef, MixGHD-method ( coef ), 6  
contourpl, 7  
DA, 8  
dCGHD, 10  
dGHD, 11  
dMSGHD, 12  
MCGHD, 9, 14, 20–22, 25, 31  
MGHD, 6, 9, 15, 16, 20–22, 24, 25, 31  
MGHFA, 9, 18, 20, 22, 25, 31  
MixGHD, 5, 7, 9, 15, 17, 19, 24, 26, 30, 31  
MixGHD ( MixGHD class ), 20  
MixGHD class, 20  
MixGHD-class, 21  
MSGHD, 6, 9, 15, 20–22, 23, 25, 31  
plot, 25  
plot, MixGHD, missing-method ( plot ), 25  
plot, MixGHD-method ( plot ), 25  
plot. MixGHD ( plot ), 25  
predict, 26  
predict, MixGHD, missing-method ( predict ), 26  
predict, MixGHD-method ( predict ), 26  
predict. MixGHD ( predict ), 26  
rCGHD, 26  
rGHD, 28  
rMSGHD, 29  
sonar, 30  
summary, 30