Package 'longclust'

December 21, 2023

Type Package

Title Model-Based Clustering and Classification for Longitudinal Data

Version 1.5

Date 2023-12-21

Author Paul D. McNicholas [aut, cre] (<https://orcid.org/0000-0002-2482-523X>),
K. Raju Jampani [aut] (May to Dec 2012),
Sanjeena Subedi [aut] (<https://orcid.org/0000-0002-2462-7786>)

Maintainer Paul D. McNicholas <mcnicholas@math.mcmaster.ca>

Suggests mvtnorm

Depends R (>= 4.3.0)

Description Clustering or classification of longitudinal data based on a mixture of multivariate t or Gaussian distributions with a Cholesky-decomposed covariance structure. Details in McNicholas and Murphy (2010) <doi:10.1002/cjs.10047> and McNicholas and Subedi (2012) <doi:10.1016/j.jspi.2011.11.026>.

License GPL (>= 2)

LazyLoad yes

Repository CRAN

NeedsCompilation yes

Date/Publication 2023-12-21 17:40:02 UTC

R topics documented:

longclust-package *Model-Based Clustering and Classification for Longitudinal Data*

Description

This is a package for clustering or classification of longitudinal data based on a mixture of multivariate t or Gaussian distributions with a Cholesky-decomposed covariance structure.

Details

Package:	longclust
Type:	Package
Version:	1.5
Date:	2023-12-21
License:	GPL-2 or GPL-3
LazyLoad:	yes

This package contains the function `longclustEM`.

Author(s)

P. D. McNicholas, K.R. Jampani and S. Subedi

Maintainer: Paul McNicholas <mcnicholas@math.mcmaster.ca>

See Also

Details, examples, and references are given under `longclustEM`.

longclustEM *Model-Based Clustering and Classification for Longitudinal Data*

Description

Carries out model-based clustering or classification using multivariate t or Gaussian mixture models with Cholesky decomposed covariance structure. EM algorithms are used for parameter estimation and the BIC is used for model selection.

Usage

```
longclustEM(x, Gmin, Gmax, class=NULL, linearMeans = FALSE,
modelSubset = NULL, initWithKMeans = FALSE, criteria = "BIC",
equalDF = FALSE, gaussian=FALSE,  userseed=1004)
```

Arguments

x	A matrix or data frame such that rows correspond to observations and columns correspond to variables.
Gmin	A number giving the minimum number of components to be used.
Gmax	A number giving the maximum number of components to be used.
class	If NULL then model-based clustering is performed. If a vector with length equal to the number of observations, then model-based classification is performed. In this latter case, the ith entry of class is either zero, indicating that the component membership of observation i is unknown, or it corresponds to the component membership of observation i.
linearMeans	If TRUE, then means are modelled using linear models.
modelSubset	A vector of strings giving the models to be used. If set to NULL, all models are used.
initWithKMeans	If TRUE, the components are initialized using k-means algorithm.
criteria	A string that denotes the criteria used for evaluating the models. Its value should be "BIC" or "ICL".
equalDF	If TRUE, the degrees of freedom of all the components will be the same.
gaussian	If TRUE, a mixture of Gaussian distributions is used in place of a mixture of t-distributions.
userseed	The random number seed to be used.

Value

Gbest	The number of components for the best model.
zbest	A matrix that gives the probabilities for any data element to belong to any component in the best model.
nubest	A vector of Gbest integers, that give the degrees of freedom for each component in the best model.
mubest	A matrix containing the means of the components for the best model (one per row).
Tbest	A list of Gbest matrices, giving the T matrices of the components for the best model.
Dbest	A list of Gbest matrices, giving the D matrices of the components for the best model.

Author(s)

Paul D. McNicholas, K. Raju Jampani and Sanjeena Subedi

References

Paul D. McNicholas and T. Brendan Murphy (2010). Model-based clustering of longitudinal data. *The Canadian Journal of Statistics* **38**(1), 153-168.

Paul D. McNicholas and Sanjeena Subedi (2012). Clustering gene expression time course data using mixtures of multivariate t-distributions. *Journal of Statistical Planning and Inference* **142**(5), 1114-1127.

Examples

```
library(mvtnorm)
m1 <- c(23,34,39,45,51,56)
S1 <- matrix(c(1.00, -0.90, 0.18, -0.13, 0.10, -0.05, -0.90,
1.31, -0.26, 0.18, -0.15, 0.07, 0.18, -0.26, 4.05, -2.84,
2.27, -1.13, -0.13, 0.18, -2.84, 2.29, -1.83, 0.91, 0.10,
-0.15, 2.27, -1.83, 3.46, -1.73, -0.05, 0.07, -1.13, 0.91,
-1.73, 1.57), 6, 6)
m2 <- c(16,18,15,17,21,17)
S2 <- matrix(c(1.00, 0.00, -0.50, -0.20, -0.20, 0.19, 0.00,
2.00, 0.00, -1.20, -0.80, -0.36,-0.50, 0.00, 1.25, 0.10,
-0.10, -0.39, -0.20, -1.20, 0.10, 2.76, 0.52, -1.22,-0.20,
-0.80, -0.10, 0.52, 1.40, 0.17, 0.19, -0.36, -0.39, -1.22,
0.17, 3.17), 6, 6)
m3 <- c(8, 11, 16, 22, 25, 28)
S3 <- matrix(c(1.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00,
1.00, -0.20, -0.64, 0.26, 0.00, 0.00, -0.20, 1.04, -0.17,
-0.10, 0.00, 0.00, -0.64, -0.17, 1.50, -0.65, 0.00, 0.00,
0.26, -0.10, -0.65, 1.32, 0.00, 0.00, 0.00, 0.00, 0.00,
0.00, 1.00), 6, 6)
m4 <- c(12, 9, 8, 5, 4 ,2)
S4 <- diag(c(1,1,1,1,1,1))
data <- matrix(0, 40, 6)
data[1:10,] <- rmvnorm(10, m1, S1)
data[11:20,] <- rmvnorm(10, m2, S2)
data[21:30,] <- rmvnorm(10, m3, S3)
data[31:40,] <- rmvnorm(10, m4, S4)
clus <- longclustEM(data, 3, 5, linearMeans=TRUE)
summary(clus)
plot(clus,data)
```

`plot.longclust` *Plots the components of the model.*

Description

Displays a series of two plots, one containing all the components in different colors, and one containing subplots one per each component.

Usage

```
## S3 method for class 'longclust'
plot(x, data, ...)
```

Arguments

`x`	An object of type longclust returned by longclustEM.
`data`	The data matrix used in computing clus.
`...`	Default arguments.

Author(s)

Paul D. McNicholas, K. Raju Jampani and Sanjeena Subedi

Examples

```
library(mvtnorm)
m1 <- c(23,34,39,45,51,56)
S1 <- matrix(c(1.00, -0.90, 0.18, -0.13, 0.10, -0.05, -0.90,
1.31, -0.26, 0.18, -0.15, 0.07, 0.18, -0.26, 4.05, -2.84,
2.27, -1.13, -0.13, 0.18, -2.84, 2.29, -1.83, 0.91, 0.10,
-0.15, 2.27, -1.83, 3.46, -1.73, -0.05, 0.07, -1.13, 0.91,
-1.73, 1.57), 6, 6)
m2 <- c(16,18,15,17,21,17)
S2 <- matrix(c(1.00, 0.00, -0.50, -0.20, -0.20, 0.19, 0.00,
2.00, 0.00, -1.20, -0.80, -0.36,-0.50, 0.00, 1.25, 0.10,
-0.10, -0.39, -0.20, -1.20, 0.10, 2.76, 0.52, -1.22,-0.20,
-0.80, -0.10, 0.52, 1.40, 0.17, 0.19, -0.36, -0.39, -1.22,
0.17, 3.17), 6, 6)
m3 <- c(8, 11, 16, 22, 25, 28)
S3 <- matrix(c(1.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 1.00,
-0.20, -0.64, 0.26, 0.00, 0.00, -0.20, 1.04, -0.17, -0.10,
0.00, 0.00, -0.64, -0.17, 1.50, -0.65, 0.00, 0.00, 0.26, -0.10,
-0.65, 1.32, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 1.00), 6, 6)
m4 <- c(12, 9, 8, 5, 4 ,2)
S4 <- diag(c(1,1,1,1,1,1))
data <- matrix(0, 40, 6)
data[1:10,] <- rmvnorm(10, m1, S1)
data[11:20,] <- rmvnorm(10, m2, S2)
data[21:30,] <- rmvnorm(10, m3, S3)
data[31:40,] <- rmvnorm(10, m4, S4)
clus <- longclustEM(data, 3, 5, linearMeans=TRUE)
plot(clus,data)
```

---

print.longclust *Brief overview of the longclust object*

Description

Prints the number of components, probabily matrix, degrees of freedom and the component means of the computed best model.

Usage

```
  ## S3 method for class 'longclust'
print(x, ...)
```

Arguments

x	An object of type longclust, computed by longclustEM.
...	Default Arguments

Author(s)

Paul D. McNicholas, K. Raju Jampani and Sanjeena Subedi

Examples

```
library(mvtnorm)
m1 <- c(23,34,39,45,51,56)
S1 <- matrix(c(1.00, -0.90, 0.18, -0.13, 0.10, -0.05, -0.90,
1.31, -0.26, 0.18, -0.15, 0.07, 0.18, -0.26, 4.05, -2.84,
2.27, -1.13, -0.13, 0.18, -2.84, 2.29, -1.83, 0.91, 0.10,
-0.15, 2.27, -1.83, 3.46, -1.73, -0.05, 0.07, -1.13, 0.91,
-1.73, 1.57), 6, 6)
m2 <- c(16,18,15,17,21,17)
S2 <- matrix(c(1.00, 0.00, -0.50, -0.20, -0.20, 0.19, 0.00, 2.00,
0.00, -1.20, -0.80, -0.36,-0.50, 0.00, 1.25, 0.10, -0.10, -0.39,
-0.20, -1.20, 0.10, 2.76, 0.52, -1.22,-0.20, -0.80, -0.10, 0.52,
1.40, 0.17, 0.19, -0.36, -0.39, -1.22, 0.17, 3.17), 6, 6)
m3 <- c(8, 11, 16, 22, 25, 28)
S3 <- matrix(c(1.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 1.00,
-0.20, -0.64, 0.26, 0.00, 0.00, -0.20, 1.04, -0.17, -0.10, 0.00,
0.00, -0.64, -0.17, 1.50, -0.65, 0.00, 0.00, 0.26, -0.10, -0.65,
1.32, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 1.00), 6, 6)
m4 <- c(12, 9, 8, 5, 4 ,2)
S4 <- diag(c(1,1,1,1,1,1))
data <- matrix(0, 40, 6)
data[1:10,] <- rmvnorm(10, m1, S1)
data[11:20,] <- rmvnorm(10, m2, S2)
data[21:30,] <- rmvnorm(10, m3, S3)
data[31:40,] <- rmvnorm(10, m4, S4)
clus <- longclustEM(data, 3, 5, linearMeans=TRUE)
print(clus)

## The function is currently defined as
function (tch, ...)
{
    cat("Number of Clusters:", tch$Gbest, "\n")
    cat("z:\n")
    print(tch$zbest)
    cat("\n")
    for (g in 1:tch$Gbest) {
        cat("Cluster: ", g, "\n")
        cat("v: ", tch$nubest[g], "\n")
        cat("mean:", tch$mubest[g, ], "\n\n")
    }
  }
```

---

summary.longclust *Summary of the longclust object*

Description

Prints all the items in the object.

Usage

```
## S3 method for class 'longclust'
summary(object, ...)
```

Arguments

object	An object of type longclust, returned by longclustEM.
...	Default arguments.

Author(s)

Paul D. McNicholas, K. R. Jampani and Sanjeena Subedi

Examples

```
library(mvtnorm)
m1 <- c(23,34,39,45,51,56)
S1 <- matrix(c(1.00, -0.90, 0.18, -0.13, 0.10, -0.05, -0.90,
1.31, -0.26, 0.18, -0.15, 0.07, 0.18, -0.26, 4.05, -2.84,
2.27, -1.13, -0.13, 0.18, -2.84, 2.29, -1.83, 0.91, 0.10,
-0.15, 2.27, -1.83, 3.46, -1.73, -0.05, 0.07, -1.13, 0.91,
-1.73, 1.57), 6, 6)
m2 <- c(16,18,15,17,21,17)
S2 <- matrix(c(1.00, 0.00, -0.50, -0.20, -0.20, 0.19, 0.00,
2.00, 0.00, -1.20, -0.80, -0.36,-0.50, 0.00, 1.25, 0.10,
-0.10, -0.39, -0.20, -1.20, 0.10, 2.76, 0.52, -1.22,-0.20,
-0.80, -0.10, 0.52, 1.40, 0.17, 0.19, -0.36, -0.39, -1.22,
0.17, 3.17), 6, 6)
m3 <- c(8, 11, 16, 22, 25, 28)
S3 <- matrix(c(1.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00,
1.00, -0.20, -0.64, 0.26, 0.00, 0.00, -0.20, 1.04, -0.17,
-0.10, 0.00, 0.00, -0.64, -0.17, 1.50, -0.65, 0.00, 0.00,
0.26, -0.10, -0.65, 1.32, 0.00, 0.00, 0.00, 0.00, 0.00,
0.00, 1.00), 6, 6)
m4 <- c(12, 9, 8, 5, 4 ,2)
S4 <- diag(c(1,1,1,1,1,1))
data <- matrix(0, 40, 6)
data[1:10,] <- rmvnorm(10, m1, S1)
data[11:20,] <- rmvnorm(10, m2, S2)
data[21:30,] <- rmvnorm(10, m3, S3)
data[31:40,] <- rmvnorm(10, m4, S4)
clus <- longclustEM(data, 3, 5, linearMeans=TRUE)
summary(clus)
```

Index