Package ‘MNARclust’

December 2, 2021

Type Package

Title Clustering Data with Non-Ignorable Missingness using Semi-Parametric Mixture Models

Version 1.1.0

Description Clustering of data under a non-ignorable missingness mechanism. Clustering is achieved by a semi-parametric mixture model and missingness is managed by using the pattern-mixture approach. More details of the approach are available in Du Roy de Chaumaray et al. (2020) <arXiv:2009.07662>.

Maintainer Matthieu Marbac <matthieu.marbac-lourdelle@ensai.fr>

License GPL (>= 2)

Imports Rcpp, parallel, sn, rmutil

LinkingTo Rcpp, RcppArmadillo

ByteCompile true


Author Marie Du Roy de Chaumaray [aut], Matthieu Marbac [aut, cre, cph]


LazyData true

LazyLoad yes

Encoding UTF-8

Depends R (>= 3.5)

RoxygenNote 7.1.0

NeedsCompilation yes

Repository CRAN

Date/Publication 2021-12-02 13:20:10 UTC
R topics documented:

- MNARclust-package
- echo
- MNARcluster
- rMNAR

Index

MNARclust-package  MNARclust.

Description

Clustering method to analyze continuous or mixed-type data with missingness. The missingness mechanism can be non ignorable. The approach considers a semi-parametric mixture model.

Details

Package: MNARclust
Type: Package
Version: 1.1.0
Date: 2021-12-01
License: GPL-3
LazyLoad: yes

echo  Echocardiogram data set

Description

All the patients suffered heart attacks at some point in the past. Some are still alive and some are not. The survival and still-alive variables, when taken together, indicate whether a patient survived for at least one year following the heart attack.

Format

A data frame with 132 observations on 13 variables (more details on this data set are presented in http://archive.ics.uci.edu/ml/datasets/Echocardiogram).

Details

This data set arise from the UCI machine learning repository (more details on this data set are presented http://archive.ics.uci.edu/ml/datasets/Echocardiogram)
MNARcluster

References

Examples
data(echo)

| MNARcluster | Clustering function |

Description
Clustering method to analyze continuous or mixed-type data with missingness. The missingness mechanism can be non ignorable. The approach considers a semi-parametric mixture model.

Usage
MNARcluster(
  x,  # matrix used for clustering
  K,  # number of components
  nbinit = 20,  # number of random starting points
  nbCPU = 1,  # number of CPU used for parallel computing (only Unix and Linux systems are allowed)
  tol = 0.01,  # stopping rule
  band = band.default(x),  # bandwidth (numeric vector).
  seedvalue = 123  # value of the seed (used to set the initializations of the MM algorithm)
)

Arguments
| x | matrix used for clustering |
| K | number of components |
| nbinit | number of random starting points |
| nbCPU | number of CPU used for parallel computing (only Unix and Linux systems are allowed) |
| tol | stopping rule |
| band | bandwidth (numeric vector). |
| seedvalue | value of the seed (used to set the initializations of the MM algorithm) |

Value
Returns a list containing the proportions (proportions), matrix of probabilities of missingness (rho), the posterior probabilities of classification (classproba), the partition (zhat) and the logarithme of the smoothed-likelihood (logSmoothlike)
rMNAR

Function used to simulate data from mixture model with specific missingness mechanism

Description

Generation of data set to perform the simulation presented in Section 4.1 of Du Roy de Chaumaray (2020)

Usage

rMNAR(n, K, d = 3, delta = 3, gamma = 1, law = "gauss", linkmissing = "logit-X")

Arguments

- **n** sample size (numeric of length 1)
- **K** number of clusters (numeric of length 1)
- **d** number of variables (numeric of length 1)
- **delta** tuning parameter to define the rate of misclassification (numeric of length 1)
- **gamma** tuning parameter to define the rate of missingness (numeric of length 1)
- **law** specifies the distribution of the variables within components (character that must be equal to gauss, student, laplace or skewgauss)
- **linkmissing** specify the missingness mechanism (character that must be equal to MCAR, logit-Z, logit-X or censoring)
Value

rMNAR returns a list containing the observed data (x), the true cluster membership (z), the complete data (xfull), the cluster membership given by the Baye’s rule (zhat), the empirical rates of misclassification (meanerrorclass) and missingness (meanmiss).

References


Examples

```r
set.seed(123)
# Data generation
ech <- rMNAR(n=100, K=3, d=3, delta=2, gamma=1)
# Head of the observed data
head(ech$x)
# Table of the cluster memberships
table(ech$z)
# Empirical rate of misclassification
ech$meanerrorclass
# Empirical rate of missingness
ech$meanmiss
```
Index

* datasets
  echo, 2
* package
  MNARclust-package, 2

  echo, 2

  MNARclust (MNARclust-package), 2
  MNARclust-package, 2
  MNARcluster, 3

  rMNAR, 4