

Package ‘MatrixMixtures’

June 11, 2021

Title Model-Based Clustering via Matrix-Variate Mixture Models

Version 1.0.0

Description Implements finite mixtures of matrix-variate contaminated normal distributions via expectation conditional-maximization algorithm for model-based clustering, as described in Tomarchio et al.(2020) <[arXiv:2005.03861](https://arxiv.org/abs/2005.03861)>. One key advantage of this model is the ability to automatically detect potential outlying matrices by computing their a posteriori probability of being typical or atypical points. Finite mixtures of matrix-variate t and matrix-variate normal distributions are also implemented by using expectation-maximization algorithms.

License GPL (>= 2)

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

Imports doSNOW, foreach, snow, withr

Depends R (>= 2.10)

NeedsCompilation no

Author Salvatore D. Tomarchio [aut],
Michael P.B. Gallagher [aut, cre],
Antonio Punzo [aut],
Paul D. McNicholas [aut]

Maintainer Michael P.B. Gallagher <michael_gallagher@baylor.edu>

Repository CRAN

Date/Publication 2021-06-11 08:00:02 UTC

R topics documented:

MatrixMixt	2
SimX	3
Index	4

Description

Fits, by using expectation-maximization algorithms, mixtures of matrix-variate distributions (normal, t, contaminated normal) to the given data. Can be run in parallel. The Bayesian information criterion (BIC) is used to select the number of groups.

Usage

```
MatrixMixt(
  X,
  G = 1:3,
  mod,
  tol = 1e-05,
  maxiter = 10000,
  ncores = 1,
  verbose = TRUE
)
```

Arguments

X	A list of dimension N, where N is the sample size. Each element of the list corresponds to an observed p x r matrix.
G	A vector containing the numbers of groups to be tried.
mod	The matrix-variate distribution to be used for the mixture model. Possible values are: "MVN" for the normal distribution, "MVT" for the t distribution "MVCN" for the contaminated normal.
tol	Threshold for Aitken's acceleration procedure. Default value is 1.0e-05.
maxiter	Maximum number of iterations of the algorithms. Default value is 10000.
ncores	A positive integer indicating the number of cores used for running in parallel. Default value is 1.
verbose	Logical indicating whether the running output should be displayed.

Value

A list with the following elements:

flag	Convergence flag (TRUE - success, FALSE - failure).
pig	Vector of the estimated mixing proportions (length G).
nu	Vector of the estimated degree of freedoms (length G). Only for "MVT".
alpha	Vector of the estimated inliers proportions (length G). Only for "MVCN".
eta	Vector of the estimated inflation parameters (length G). Only for "MVCN".

M	Array of the mean matrices ($p \times r \times G$).
Sigma	Array of the estimated row covariance matrices ($p \times p \times G$).
Psi	Array of the estimated column covariance matrices ($r \times r \times G$).
class	Vector of estimated data classification.
z	Matrix of estimated posterior probabilities ($N \times G$).
v	Matrix of estimated inlier probabilities ($N \times G$). Only for "MVCN".
lik	Estimated log-likelihood.
BIC	Estimated BIC.

Examples

```
data(SimX)
res <- MatrixMxt(X = SimX, G = 2, mod = "MVCN")
```

SimX	<i>Simulated Data</i>
------	-----------------------

Description

A simulated dataset with 2 groups and 80 observations. Each group consists of 40 observations, 5 of which are outliers.

Usage

```
data(SimX)
```

Format

An object of class `list` of length 80.

Index

* **datasets**

SimX, 3

MatrixMixt, 2

SimX, 3