Package ‘VarSelLCM’

March 19, 2015

Type Package
Version 1.0
Date 2015-03-19
Author Matthieu Marbac and Mohammed Sedki
Title Variable Selection for Model-Based Clustering using the Integrated Complete-Data Likelihood of a Latent Class Model
Description Variable Selection for model-based clustering by using a mixture model of Gaussian distributions assuming conditional independence between variables. The algorithm carries out the model selection by optimizing the MICL criterion which has a closed form for such a distribution.
Maintainer Mohammed Sedki <mohammed.sedki@u-psud.fr>
Imports methods, Rcpp (>= 0.11.1), parallel, mclust
LinkingTo Rcpp, RcppArmadillo
License GPL (>= 2)
ByteCompile true
LazyLoad yes
Collate 'ICLexact.R' 'RcppExports.R' 'VSLCM_Classes.R'
    'VSLCM_initialization.R' 'VarSelLCM.R' 'Print.R' 'Summary.R'
    'onAttach.R'
NeedsCompilation yes
Repository CRAN
Date/Publication 2015-03-19 23:47:39

R topics documented:

  VarSelLCM-package .................................................. 2
  print ............................................................... 3
  summary ........................................................... 4
  VarSelCluster ..................................................... 4
  VSLCMcriteria-class ............................................. 5
  VSLCMmodel-class ................................................. 6
Variable Selection for Model-Based Clustering using the Integrated Complete-Data Likelihood of a Latent Class Model

Description

Variable Selection for model-based clustering by using a mixture model of Gaussian distributions assuming conditional independence between variables. The algorithm carries out the model selection by optimizing the MICL criterion which has a closed form for such a distribution.

Details

Package: VarSelLCM
Type: Package
Version: 1.0
Date: 2014-11-25
License: GPL (>= 2)

The main functions to use is VarSelCluster. It carries out the model selection by maximizing the MICL criterion.

Author(s)

Matthieu Marbac and Mohammed Sedki
Maintainer: Mohammed Sedki <mohammed.sedki@upsud.fr>

References


Examples

# Variable selection
require(mclust)
require(VarSelLCM)
data(banknote)

results <- VarSelCluster(banknote[, -1], 2, parallel=FALSE)

summary(results)
print(results)

## Not run:
## To install multitest package : uncomment the following lines
## source("http://bioconductor.org/biocLite.R")
## biocLite("multtest")
require(mclust)
require(VarSelLCM)
require(multtest)
data(golub)

results <- VarSelCluster(t(golub), 2)

summary(results)

## End(Not run)

print

---

### Print an object of class VSLCMresults

**Description**

Print an object of class VSLCMresults

**Arguments**

- `x` An object of class `VSLCMresults`

**Value**

NULL.

**See Also**

print

**Examples**

```r
require(mclust)
require(VarSelLCM)
data(banknote)
results <- VarSelCluster(banknote[,-1], 2, parallel=FALSE)
print(results)
```
VarSelCluster

**summary**

*Produce result summary of a VSLCMresults class*

**Description**

Produce result summary of a VSLCMresults class

**Arguments**

- **object**: An object of class `VSLCMresults`

**Value**

NULL. Summaries to standard out.

**See Also**

`summary`

**Examples**

```r
require(mclust)
require(VarSelLCM)
data(banknote)
results <- VarSelCluster(banknote[,-1], 2, parallel=FALSE)
summary(results)
```

---

VarSelCluster

**Description**

This function performs the variable selection for a model-based clustering. The data distribution follow a mixture of Gaussian distributions assuming conditional independence between variables. The model selection is carried out by maximizing the MICL criterion.

**Usage**

```r
VarSelCluster(x, g, nbinit=50, parallel = TRUE)
```

**Arguments**

- **x**: A matrix, or data frame where rows correspond to observations and columns correspond to variables.
- **g**: An integer specifying the number of mixture components
- **nbinit**: The number of random initializations of the algorithm. The default is 50.
- **parallel**: A Boolean indicating if the code is parallelized or not. The default is TRUE.
Value

An object of class \texttt{VSLCMresults} providing the model maximizing the MICL criterion. Thus, it indicates which variables are relevant to the clustering. It also provides the maximum likelihood estimates associated to this model.

References


Examples

```r
require(mclust)
require(VarSelLCM)
data(banknote)
results <- VarSelCluster(banknote[,-1], 2, parallel=FALSE)

summary(results)
print(results)

## Not run:
## To install multitest package : uncomment the following lines
## source("http://bioconductor.org/biocLite.R")
## biocLite("multtest")
require(mclust)
require(VarSelLCM)
require(multtest)
data(golub)

results <- VarSelCluster(t(golub), 2)

summary(results)

## End(Not run)
```

---

\texttt{VSLCMcriteria-class} Class "VSLCMcriteria"

Description

This class contains the information criteria
Slots

- likelihood: Value of the log-likelihood
- BIC: Values of the BIC criterion
- ICL: Values of the ICL criterion
- MICL: Values of the MICL criterion

Examples

showClass("VSLCMcriteria")

---

VSLCMmodel-class Class "VSLCMmodel"

Description

This class contains the model information

Slots

- g: Number of components
- omega: Boolean vector indicating if each variable is irrelevant (1) or not (0) to the clustering

Examples

showClass("VSLCMmodel")

---

VSLCMparameters-class Class "VSLCMparameters"

Description

This class contains the parameters

Slots

- proportions: Vector of the class proportions
- means: Matrix of the means where row correspond to classes and column to variables
- variances: Matrix of the variances where row correspond to classes and column to variables

Examples

showClass("VSLCMparameters")
VSLCMpartitions-class

Class "VSLCMpartitions"

Description

This class contains the estimated partitions

Slots

zMAP: A vector indicating the class membership of each individual by using the MAP rule computed for the best model with its maximum likelihood estimates
zOPT: Partition maximizing the integrated complete-data likelihood of the selected model
tik: Fuzzy partition computed for the best model with its maximum likelihood estimates

Examples

showClass("VSLCMpartitions")

VSLCMresults-class

Class "VSLCMresults"

Description

This class contains the model information

Slots

data: A matrix, or data frame where rows correspond to observations and columns correspond to variables.
priors: Matrix where rows indicate the values of the hyper-parameters
criteria: An object of class VSLCMcriteria
partitions: An object of class VSLCMpartitions
model: An object of class VSLCMmodel
parameters: An object of class VSLCMparameters

Examples

showClass("VSLCMresults")
Index

*Topic classes
  VSLCMcriteria-class, 5
  VSLCMmodel-class, 6
  VSLCMparameters-class, 6
  VSLCMpartitions-class, 7
  VSLCMresults-class, 7
*Topic package
  VarSelLCM-package, 2

print. 3, 3
print, VSLCMresults-method (print), 3

summary, 4, 4
summary, VSLCMresults-method (summary), 4

VarSelCluster. 2, 4
VarSelLCM (VarSelLCM-package). 2
VarSelLCM-package, 2
VSLCMcriteria, 7
VSLCMcriteria-class, 5
VSLCMmodel, 7
VSLCMmodel-class, 6
VSLCMparameters, 7
VSLCMparameters-class, 6
VSLCMpartitions, 7
VSLCMpartitions-class, 7
VSLCMresults, 3–5
VSLCMresults-class, 7