Package 'rcbmm'

November 15, 2020

Type Package

Title Regularized Copula-Based Mixture Model: Estimation and Model Selection

Version 0.1.0

Author person(``Ben'', ``Barlow'', email = ``ben.j.barlow.1@gmail.com'', role = c(``aut'', ``cre''))

Maintainer Ben Barlow <ben.j.barlow.1@gmail.com>

Description Model-based clustering through regularized copula-based mixture models with arbitrary marginal distributions. Estimation via the expectation-conditional-maximization algorithm. Shrinkage driven methods used to adaptively select optimal correlation structure regulated by a tuning parameter. Model selection based on average silhouette width and BIC.

Encoding UTF-8

LazyData true

License MIT + file LICENSE

RoxygenNote 7.1.1

Suggests testthat

Imports cluster,
utils,
parallel,
stats,
copula,
mvtnorm,
mclust,
fitdistrplus,
Rcpp

LinkingTo Rcpp

R topics documented:

angles2rho 2
cm.step.1 2
cm.step.2 3
e.step 4
ecm 5
extract.copula.pars 7
extract.marginal.pars 7
fit.rcbmm 8
initialize.ecm 9

P2p.angles 11
rho2angles 11

Index **12**

`angles2rho` *Matrix conversion tool*

Description

A tool for converting a matrix of angles to a correlation matrix using Choleksy factor

Usage

```
angles2rho(theta)
```

Arguments

`theta`	A $(p-1)$ times $(p-1)$ matrix of angles.

Value

A p times p correlation matrix.

See Also

`rho2angles`

Examples

```
theta <- matrix(rep(0.5, 4), 2, 2)
angles2rho(theta)
```

`cm.step.1` *Conditional-maximization step 1 (M-step 2)*

Description

Implements the conditional-maximization step 1 of the ECM algorithm for regularized copula-based mixture model.

Usage

```
cm.step.1(
  x,
  K,
  z,
  mvdc,
  margins,
  trace = TRUE,
  restrictions = TRUE,
  variance_tolerance = 1e-05
)
```

Arguments

x	A numeric matrix or data frame of observations. Rows correspond to observations and columns correspond to variables.
K	The number of mixture components.
z	A numeric matrix representing the current value of the posterior probabilities of membership of the observations after the expectation step of the ECM algorithm. Columns are associated with a mixture component and rows are associated with observations. The current value of the probabilities should be computed via e.step.
mvdc	A list of objects of class mvdc. Each element of the list corresponds to a mixture component and contains the previous estimates for the component distribution's marginal and copula parameters.
margins	A character vector specifying the marginal distributions of the components in the mixture. The vector must have a length equal to the number of columns in x. Each element must be equal to "norm", "beta" or "gamma".
trace	A logical value indicating if an update regarding the step's progress should be displayed.
restrictions	A logical value indicating if the variance of each Beta marginal should be restricted
variance_tolerance	The lower bound for the variance of each Beta marginal, if restrictions = TRUE.

Value

A list of objects of class mvdc. Each element of the list corresponds to a mixture component and contains the updated estimates for the component distribution's marginal paramters and the same estimates as were parsed for the estimates of the copula parameter.

See Also

cm.step.2 ecm

---

cm.step.2 *Conditional-maximization step 2 (M-step 2)*

Description

Implements the conditional-maximization step 2 of the ECM algorithm for regularized copula-based mixture model.

Usage

```
cm.step.2(x, K, z, mvdc, margins, lambda, trace = T)
```

Arguments

x	A numeric matrix or data frame of observations. Rows correspond to observations and columns correspond to variables.
K	The number of mixture components.
z	A numeric matrix representing the current value of the posterior probabilities of membership of the observations after the expectation step of the last iteration of the ECM algorithm. Columns are associated with a mixture component and rows are associated with observations.
mvdc	A list of objects of class mvdc. Each element of the list corresponds to a mixture component and contains the current estimates for the component distribution's marginal and previous estimates for copula parameters. The estimates for the marginal parameters should have been updated by cm.step.1.
margins	A character vector specifying the marginal distributions of the components in the mixture. The vector must have a length equal to the number of columns in x. Each element must be equal to "norm", "beta" or "gamma".
lambda	A numeric value indicating the value of the tuning parameter for regularization.
trace	A logical value indicating if an update regarding the step's progress should be displayed.

Value

A list with the following elements

- mvdcA list of objects of class mvdc. Each element of the list corresponds to a mixture component and contains the updated estimates for the component distribution's copula paramter and the same estimates as were parsed for the estimates of the marginal parameters.
- penaltyThe shrinkage penalty to apply to the log-likelihood of the model, resulting from the estimates of the copula parameters and the value of lambda.

See Also

cm.step.1 ecm

---

e.step *Expectation step*

Description

Implements the expectation step of the ECM algorithm for regularized copula-based mixture model.

Usage

```
e.step(x, K, mixing_probs, mvdc, margins)
```

Arguments

x	A numeric matrix or data frame of observations. Rows correspond to observations and columns correspond to variables.
K	The number of mixture components.
mixing_probs	A numeric vector indicating the mixing proportions of the mixture model.
mvdc	A list of objects of class mvdc. Each element of the list corresponds to a mixture component and contains the current estimates for the component distribution's marginal and copula parameters.
margins	A character vector specifying the marginal distributions of the components in the mixture. The vector must have a length equal to the number of columns in x. Each element must be equal to "norm", "beta" or "gamma".

Value

A numeric matrix representing the posterior probabilities of membership of the observations after performing a single expectation step of the ECM algorithm. Columns are associated with a mixture component and rows are associated with observations.

See Also

ecm

---

ecm *Model estimation through ECM*

Description

Implements the expectation-conditional-maximization algorithm for a regularized copula-based mixture model given initial parameter values, starting with the expectation step.

Usage

```
ecm(
  x,
  K,
  lambda,
  start = NULL,
  margins,
  trace = TRUE,
  maxit = 1000,
  epsilon = 1e-06,
  dist_mat = NULL
)
```

Arguments

x	A numeric matrix or data frame of observations. Rows correspond to observations and columns correspond to variables.
K	An integer specifying the number of components for which a regularized copula-based mixture model should be fitted.
lambda	A numeric value indicating the value of the tuning parameter for regularization.
start	A list providing the starting values for ECM. The list is produced by fit.rcbmm.
margins	A character vector specifying the marginal distributions of the components in the mixture. The vector must have a length equal to the number of columns in x. Each element must be equal to "norm","beta" or "gamma".
trace	A logical value indicating if an update regarding the algorithm's progress should be displayed after each iteration.
maxit	A numeric value indicating the maximal number of ECM iterations.
epsilon	A numeric value specifying the tolerance associated with determining when convergence of the ECM algorithm has been achieved.
dist_mat	An object of type dist for calculating silhouette values.

Value

- KThe number of mixture components.
- lambdaThe value of the tuning parameter.
- zA numeric matrix representing the posterior probabilities of membership of the observations after the expectation step of the last iteration of the ECM algorithm. Columns are associated with a mixture component and rows are associated with observations.
- clustersA classification vector indicating the associated cluster of each observation. The classification corresponds to z.
- loglikA numeric vector displaying the penalized log-likelihood after each iteration of ECM.
- param_numberThe number of independent parameters associated with the model.
- BICThe BIC value of the model. Computed using the unpenalized log-likelihood after the last iteration of the ECM algorithm.
- mixing_probsThe mixing proportions associated with the model.
- mvdcA list of objects of class mvdc. Each element of the list corresponds to a mixture component.
- transformationA character value indicating the transformation when identifying starting values. The value is NULL unless lambda=0. See initialize.ecm.
- marginal_paramA list containing the marginal parameters of the model as estimated by ECM. Each element corresponds to a mixture component.
- copula_paramA list containing the copula parameters of the model as estimated by ECM. Each element corresponds to a mixture component.
- copula_param_anglesA list containing the copula parameters of the model re-expressed as angles.
- silhouetteSee information regarding silhouette package. Add reference here.

extract.copula.pars *Extraction of copula parameters from a model*

Description

A function for extracting the copula parameters parameterizing the mixture components of a regularized copula-based mixture model from a list of mvdc objects. The copula parameter for each mixture component can be extracted in terms of a correlation matrix or a matrix of angles.

Usage

```
extract.copula.pars(mvdc, as_angles = F)
```

Arguments

mvdc	A list of objects of class mvdc. Each element of the list corresponds to a mixture component and contains the estimates for the component distribution's marginal and copula parameters.
as_angles	A logical value indicating whether the copula parameter should be returned as a matrix of angles instead of a correlation matrix.

Value

A list of matrices corresponding to the copula parameter of each mixture component contained in mvdc.

See Also

extract.marginal.pars

extract.marginal.pars *Extraction of marginal parameters from a model*

Description

A function for extracting the marginal parameters parameterizing the mixture components of a regularized copula-based mixture model.

Usage

```
extract.marginal.pars(mvdc)
```

Arguments

mvdc	A list of objects of class mvdc. Each element of the list corresponds to a mixture component and contains the estimates for the component distribution's marginal and copula parameters.

Value

A list of lists corresponding to the marginal parameters of each mixture component contained in mvdc. Each sub-list in the list has a length corresponding to the number of marginal distributions defined by the model.

See Also

extract.copula.pars

---

fit.rcbmm *Model selection*

Description

The function selects the most appropriate model from a family of regularized copula-based mixture models arising from a varying number of components and a differing shirnkage parameter.

Usage

```
fit.rcbmm(
  x,
  lambda_grid,
  K = seq.int(2, 9),
  margins,
  maxit = 1000,
  epsilon = 1e-06,
  transform = TRUE,
  trace = FALSE
)
```

Arguments

x	A numeric matrix or data frame of observations. Rows correspond to observations and columns correspond to variables.
lambda_grid	An integer vector specifying the the values of the shrinkage parameter for which a regularized copula-based mixture model should be fitted. The default is lambda_grid = 0. If the vector parsed does not contain 0, then 0 is appended as starting values can only be obtained in the case no regularization is applied to the model.
K	An integer vector specifying the number of components for which a regularized copula-based mixture model should be fitted. The default is K=2:9.
margins	A character vector specifying the marginal distributions of the components in the mixture. The vector must have a length equal to the number of columns in x. Each element must be equal to "norm","beta" or "gamma".
maxit	A numeric value specifying the maximum number of iterations the ECM algorithm should run before being halted.
epsilon	A numeric value indicting the tolerance for convergence.
transform	A logical value indicating whether or not starting values for the case lambda = 0 should be obtained using the transformations SPH, PCS, PCR and SVD. The default is TRUE.
trace	A logical value indicating if an update regarding the step's progress should be displayed.

Value

- BICA matrix demonstrating the BIC values achieved by each model in the family. Each column corresponds to a given value of lambda and each row corresponds to a given number of components.
- SILA matrix demonstrating the average silhouette width achieved by each model in the family. Each column corresponds to a given value of lambda and each row corresponds to a given number of components.
- all_modelsA list of lists with each element containing information about a specific model fitted.
- selected_modelsA list containing the information about the optimal model for each number of mixture components in the family. The optimal model for each number of componets is selected by picking the lambda that results in the largest average silhouette width. See the help file for silhouette for details.
- final_modelThe model contained in `selected_models` that maximized BIC.

See Also

`ecm` `initialize.ecm`

`initialize.ecm` *Starting values for ECM*

Description

Identifies optimal starting values for ECM using the results of MBHAC applied to the data. The data undergoes transformations to enhance separation amongst groups prior to performing MBHAC.

Identifies optimal starting values for ECM using the results of MBHAC applied to the data. The data undergoes transformations to enhance separation amongst groups prior to performing MBHAC.

Usage

```
initialize.ecm(
  x,
  K,
  margins,
  transform = FALSE,
  hc_pairs = NULL,
  classification = NULL,
  trace = TRUE
)

initialize.ecm(
  x,
  K,
  margins,
  transform = FALSE,
  hc_pairs = NULL,
  classification = NULL,
  trace = TRUE
)
```

Arguments

x	A numeric matrix or data frame of observations. Rows correspond to observations and columns correspond to variables.
K	The number of mixture componets.
margins	A character vector specifying the marginal distributions of the components in the mixture. The vector must have a length equal to the number of columns in x. Each element must be equal to "norm", "beta" or "gamma".
transform	A logical value indicating whether or not starting values should be obtained using the transformations SPH, PCS, PCR and SVD. The default is TRUE.
hc_pairs	The results from MBHAC obtained from the function call hc() using the mclust package. If NULL, the function obtains the results by calling hc().
classification	A numeric vector representing a partitioning of the data x. If not NULL, the clustering identified by the vector takeß preference over the clustering identified by hc_pairs.
trace	A logical value indicating if an update regarding the initialization procedure's progress should be displayed.

Value

- mixing_probsA numeric vector indicating the starting values of the mixing proportions.
- mvdcA list of objects of class mvdc. Each element of the list corresponds to a mixture component and contains the starting values for the component distribution's marginal and copula parameters.
- transformationA character string indicating the transformation applied prior to performing MBHAC.
- loglikThe log-likelihood corresponding of the model given the starting values of the parameters and the data.

- mixing_probsA numeric vector indicating the starting values of the mixing proportions.
- mvdcA list of objects of class mvdc. Each element of the list corresponds to a mixture component and contains the starting values for the component distribution's marginal and copula parameters.
- transformationA character string indicating the transformation applied prior to performing MBHAC.
- loglikThe log-likelihood corresponding of the model given the starting values of the parameters and the data.

See Also

hc

hc

P2p.angles *Tool for handling angles*

Description

A tool for alternating between a matrix of angles and a vector of angles

Usage

```
P2p.angles(theta)

p2P.angles(angles)
```

Arguments

theta	A matrix of angles
angles	A vector of angles

Value

A vector of angles

A matrix of angles

rho2angles *Matrix conversion tool*

Description

A tool for converting a correlation matrix to a matrix of angles using Cholesky factor

Usage

```
rho2angles(rho_mat)
```

Arguments

rho_mat	A p times p correlation matrix to be converted to a matrix of angles.

Value

A $(p-1)$ times $(p-1)$ matrix of angles.

See Also

angles2rho

Index

angles2rho, 2, *11*

cm.step.1, 2, *4*
cm.step.2, *3*, 3

e.step, *3*, 4
ecm, *3–5*, 5, *9*
extract.copula.pars, 7, *8*
extract.marginal.pars, *7*, 7

fit.rcbmm, *6*, 8

hc, *10*

initialize.ecm, *6*, *9*, 9

P2p.angles, 11
p2P.angles *(*P2p.angles*)*, 11

rho2angles, *2*, 11