Package ‘flamingos’

July 19, 2019

Type Package

Title Functional latent data models for clustering heterogeneous time-series ('FLaMingoS')

Version 0.1.0

Description flamingos is an opent sources toolbox (available in R and in Matlab) for the simultaneous clustering (or classification) and segmentation of heterogeneous functional data (i.e time series ore more generally longitudinal data), with original and flexible functional latent variable models, fitted by unsupervised algorithms, including EM algorithms.

URL https://github.com/fchamroukhi/FLaMingoS

License GPL (>= 3)

Depends R (>= 2.10)

Imports methods,
stats,
Rcpp

Suggests knitr,
rmarkdown

LinkingTo Rcpp,
RcppArmadillo

Collate flamingos-package.R
RcppExports.R
utils.R
kmeans.R
mkStochastic.R
FData.R
ParamMixHMM.R
ParamMixHMMR.R
ParamMixRHLP.R
StatMixHMM.R
StatMixHMMR.R
StatMixRHLP.R
ModelMixHMMR.R
ModelMixHMM.R
ModelMixRHLP.R
emMixHMM.R
emMixHMMR.R
emMixRHLP.R
cemMixRHLP.R
data-toydataset.R

VignetteBuilder knitr

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 6.1.1

R topics documented:

`flamingos-package` *FLaMingoS: Functional Latent datA Models for clusterING heterogeneOus time-Series*

Description

`flamingos` is a toolbox for the simultaneous clustering (or classification) and segmentation of heterogeneous functional data (i.e time-series ore more generally longitudinal data), with original and flexible functional latent variable models, fitted by unsupervised algorithms, including EM algorithms.

`flamingos` contains the following time series segmentation models:

- mixRHLP;
- mixHMM;
- mixHMMR;
- PWRM;
- MixReg;
- unsupMixReg.

For the advantages/differences of each of them, the user is referred to our mentioned paper references.

To learn more about `flamingos`, start with the vignettes: `browseVignettes(package = "flamingos")`

Author(s)

Maintainer: Florian Lecocq <florian.lecocq@outlook.com>

Authors:

- Faicel Chamroukhi <faicel.chamroukhi@unicaen.fr> (0000-0002-5894-3103)
- Marius Bartcus <marius.bartcus@gmail.com>

References

Chamroukhi, Faicel, and Hien D. Nguyen. 2019. *Model-Based Clustering and Classification of Functional Data.* Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. https://chamroukhi.com/papers/MBCC-FDA.pdf.

Chamroukhi, F. 2016. *Unsupervised Learning of Regression Mixture Models with Unknown Number of Components.* Journal of Statistical Computation and Simulation 86 (November): 2308–34. https://chamroukhi.com/papers/Chamroukhi-JSCS-2015.pdf.

Chamroukhi, Faicel. 2016. *Piecewise Regression Mixture for Simultaneous Functional Data Clustering and Optimal Segmentation.* Journal of Classification 33 (3): 374–411. https://chamroukhi.com/papers/Chamroukhi-PWRM-JournalClassif-2016.pdf.

Chamroukhi, F. 2015. *Statistical Learning of Latent Data Models for Complex Data Analysis.* Habilitation Thesis (HDR), Universite de Toulon.

Chamroukhi, F., H. Glotin, and A. Same. 2013. *Model-Based Functional Mixture Discriminant Analysis with Hidden Process Regression for Curve Classification.* Neurocomputing 112: 153–63. https://chamroukhi.com/papers/chamroukhi_et_al_neucomp2013a.pdf.

Chamroukhi, F., and H. Glotin. 2012. *Mixture Model-Based Functional Discriminant Analysis for Curve Classification.* In Proceedings of the International Joint Conference on Neural Networks (IJCNN), IEEE, 1-8. Brisbane, Australia. https://chamroukhi.com/papers/Chamroukhi-ijcnn-2012.pdf.

Same, A., F. Chamroukhi, Gerard Govaert, and P. Aknin. 2011. *Model-Based Clustering and Segmentation of Time Series with Changes in Regime.* Advances in Data Analysis and Classification 5 (4): 301–21. https://chamroukhi.com/papers/adac-2011.pdf.

Chamroukhi, F., A. Same, P. Aknin, and G. Govaert. 2011. *Model-Based Clustering with Hidden Markov Model Regression for Time Series with Regime Changes.* In Proceedings of the International Joint Conference on Neural Networks (IJCNN), IEEE, 2814–21. https://chamroukhi.com/papers/Chamroukhi-ijcnn-2011.pdf.

Chamroukhi, F., A. Same, G. Govaert, and P. Aknin. 2010. *A Hidden Process Regression Model for Functional Data Description. Application to Curve Discrimination.* Neurocomputing 73 (7-9): 1210–21. https://chamroukhi.com/papers/chamroukhi_neucomp_2010.pdf.

Chamroukhi, F. 2010. *Hidden Process Regression for Curve Modeling, Classification and Tracking.* Ph.D. Thesis, Universite de Technologie de Compiegne. https://chamroukhi.com/papers/FChamroukhi-Thesis.pdf.

See Also

Useful links:

- https://github.com/fchamroukhi/FLaMingoS cemMixRHLP *cemMixRHLP is used to fit a MixRHLP model.*

Description

cemMixRHLP is used to fit a MixRHLP model. The estimation method is performed by the Classification Expectation-Maximization algorithm (CEM algorithm).

Usage

```
cemMixRHLP(X, Y, G, K, p = 3, q = 1,
  variance_type = c("heteroskedastic", "homoskedastic"),
  init_kmeans = TRUE, n_tries = 1, max_iter = 100,
  threshold = 1e-05, verbose = FALSE, verbose_IRLS = FALSE)
```

Arguments

`X`	Numeric vector of length *m* representing the covariates/inputs $x_1, \ldots, x_m$.
`Y`	Matrix of size (n, m) representing the observed responses/outputs. `Y` consists of *n* functions of `X` observed at points $1, \ldots, m$.
`G`	The number of clusters (Number of RHLP models).
`K`	The number of regimes (RHLP components) for each cluster.
`p`	Optional. The order of the polynomial regression. By default, `p` is set at 3.
`q`	Optional. The dimension of the logistic regression. For the purpose of segmentation, it must be set to 1 (which is the default value).
`variance_type`	Optional character indicating if the model is "homoskedastic" or "heteroskedastic". By default the model is "heteroskedastic".
`init_kmeans`	Optional. A logical indicating whether or not the curve partition should be initialized by the K-means algorithm. Otherwise the curve partition is initialized randomly.
`n_tries`	Optional. Number of runs of the EM algorithm. The solution providing the highest log-likelihood will be returned. If `n_tries` > 1, then for the first run, parameters are initialized by uniformly segmenting the data into K segments, and for the next runs, parameters are initialized by randomly segmenting the data into K contiguous segments.
`max_iter`	Optional. The maximum number of iterations for the EM algorithm.
`threshold`	Optional. A numeric value specifying the threshold for the relative difference of log-likelihood between two steps of the EM as stopping criteria.
`verbose`	Optional. A logical value indicating whether or not values of the log-likelihood should be printed during EM iterations.
`verbose_IRLS`	Optional. A logical value indicating whether or not values of the criterion optimized by IRLS should be printed at each step of the EM algorithm.

Details

cemMixRHLP function implements the CEM algorithm. This function starts with an initialization of the parameters done by the method `initParam` of the class ParamMixRHLP, then it alternates between the E-Step, the C-Step (methods of the class StatMixRHLP), and the CM-Step (method of the class ParamMixRHLP) until convergence (until the relative variation of log-likelihood between two steps of the EM algorithm is less than the `threshold` parameter).

Value

EM returns an object of class ModelMixRHLP.

See Also

ModelMixRHLP, ParamMixRHLP, StatMixRHLP

---

emMixHMM — *emMixHMM is used to fit a MixHMM model.*

Description

emMixHMM is used to fit a MixHMM model. The estimation method is performed by the Expectation-Maximization algorithm.

Usage

```
emMixHMM(Y, K, R, variance_type = c("heteroskedastic", "homoskedastic"),
  order_constraint = TRUE, init_kmeans = TRUE, n_tries = 1,
  max_iter = 1000, threshold = 1e-06, verbose = FALSE)
```

Arguments

`Y`	Matrix of size (n, m) representing the observed responses/outputs. `Y` consists of *n* functions of `X` observed at points $1, \ldots, m$.
`K`	The number of clusters (Number of HMM models).
`R`	The number of regimes (HMM components) for each cluster.
`variance_type`	Optional character indicating if the model is "homoskedastic" or "heteroskedastic". By default the model is "heteroskedastic".
`order_constraint`	Optional. A logical indicating whether or not a mask of order one should be applied to the transition matrix of the Markov chain. For the purpose of segmentation, it must be set to `TRUE` (which is the default value).
`init_kmeans`	Optional. A logical indicating whether or not the curve partition should be initialized by the K-means algorithm. Otherwise the curve partition is initialized randomly.
`n_tries`	Optional. Number of runs of the EM algorithm. The solution providing the highest log-likelihood will be returned. If `n_tries` > 1, then for the first run, parameters are initialized by uniformly segmenting the data into K segments, and for the next runs, parameters are initialized by randomly segmenting the data into K contiguous segments.
`max_iter`	Optional. The maximum number of iterations for the EM algorithm.
`threshold`	Optional. A numeric value specifying the threshold for the relative difference of log-likelihood between two steps of the EM as stopping criteria.
`verbose`	Optional. A logical value indicating whether or not values of the log-likelihood should be printed during EM iterations.

Details

emMixHMM function implements the EM algorithm. This function starts with an initialization of the parameters done by the method `initParam` of the class ParamMixHMM, then it alternates between the E-Step (method of the class StatMixHMM) and the M-Step (method of the class ParamMixHMM) until convergence (until the relative variation of log-likelihood between two steps of the EM algorithm is less than the `threshold` parameter).

Value

EM returns an object of class ModelMixHMM.

See Also

ModelMixHMM, ParamMixHMM, StatMixHMM

`emMixHMMR` *emMixHMMR is used to fit a MixHMMR model.*

Description

emMixHMMR is used to fit a MixHMMR model. The estimation method is performed by the Expectation-Maximization algorithm.

Usage

```
emMixHMMR(X, Y, K, R, p = 3, variance_type = c("heteroskedastic",
  "homoskedastic"), order_constraint = TRUE, init_kmeans = TRUE,
  n_tries = 1, max_iter = 1000, threshold = 1e-06, verbose = FALSE)
```

Arguments

`X`	Numeric vector of length *m* representing the covariates/inputs $x_1, \dots, x_m$.
`Y`	Matrix of size (n, m) representing the observed responses/outputs. `Y` consists of *n* functions of `X` observed at points $1, \dots, m$.
`K`	The number of clusters (Number of HMMR models).
`R`	The number of regimes (HMMR components) for each cluster.
`p`	Optional. The order of the polynomial regression. By default, `p` is set at 3.
`variance_type`	Optional character indicating if the model is "homoskedastic" or "heteroskedastic". By default the model is "heteroskedastic".
`order_constraint`	Optional. A logical indicating whether or not a mask of order one should be applied to the transition matrix of the Markov chain. For the purpose of segmentation, it must be set to `TRUE` (which is the default value).
`init_kmeans`	Optional. A logical indicating whether or not the curve partition should be initialized by the K-means algorithm. Otherwise the curve partition is initialized randomly.

`n_tries` Optional. Number of runs of the EM algorithm. The solution providing the highest log-likelihood will be returned.

If `n_tries` > 1, then for the first run, parameters are initialized by uniformly segmenting the data into K segments, and for the next runs, parameters are initialized by randomly segmenting the data into K contiguous segments.

`max_iter` Optional. The maximum number of iterations for the EM algorithm.

`threshold` Optional. A numeric value specifying the threshold for the relative difference of log-likelihood between two steps of the EM as stopping criteria.

`verbose` Optional. A logical value indicating whether or not values of the log-likelihood should be printed during EM iterations.

Details

emMixHMMR function implements the EM algorithm. This function starts with an initialization of the parameters done by the method `initParam` of the class ParamMixHMMR, then it alternates between the E-Step (method of the class StatMixHMMR) and the M-Step (method of the class ParamMixHMMR) until convergence (until the relative variation of log-likelihood between two steps of the EM algorithm is less than the `threshold` parameter).

Value

EM returns an object of class ModelMixHMMR.

See Also

ModelMixHMMR, ParamMixHMMR, StatMixHMMR

`emMixRHLP` *emMixRHLP is used to fit a MixRHLP model.*

Description

emMixRHLP is used to fit a MixRHLP model. The estimation method is performed by the Expectation-Maximization algorithm.

Usage

```
emMixRHLP(X, Y, G, K, p = 3, q = 1,
  variance_type = c("heteroskedastic", "homoskedastic"),
  init_kmeans = TRUE, n_tries = 1, max_iter = 1000,
  threshold = 1e-05, verbose = FALSE, verbose_IRLS = FALSE)
```

Arguments

`X` Numeric vector of length *m* representing the covariates/inputs $x_1, \ldots, x_m$.

`Y` Matrix of size (n, m) representing the observed responses/outputs. `Y` consists of *n* functions of `X` observed at points $1, \ldots, m$.

`G` The number of clusters (Number of RHLP models).

`K` The number of regimes (RHLP components) for each cluster.

p	Optional. The order of the polynomial regression. By default, p is set at 3.
q	Optional. The dimension of the logistic regression. For the purpose of segmentation, it must be set to 1 (which is the default value).
variance_type	Optional character indicating if the model is "homoskedastic" or "heteroskedastic". By default the model is "heteroskedastic".
init_kmeans	Optional. A logical indicating whether or not the curve partition should be initialized by the K-means algorithm. Otherwise the curve partition is initialized randomly.
n_tries	Optional. Number of runs of the EM algorithm. The solution providing the highest log-likelihood will be returned. If n_tries > 1, then for the first run, parameters are initialized by uniformly segmenting the data into K segments, and for the next runs, parameters are initialized by randomly segmenting the data into K contiguous segments.
max_iter	Optional. The maximum number of iterations for the EM algorithm.
threshold	Optional. A numeric value specifying the threshold for the relative difference of log-likelihood between two steps of the EM as stopping criteria.
verbose	Optional. A logical value indicating whether or not values of the log-likelihood should be printed during EM iterations.
verbose_IRLS	Optional. A logical value indicating whether or not values of the criterion optimized by IRLS should be printed at each step of the EM algorithm.

Details

emMixRHLP function implements the EM algorithm. This function starts with an initialization of the parameters done by the method initParam of the class ParamMixRHLP, then it alternates between the E-Step (method of the class StatMixRHLP) and the M-Step (method of the class ParamMixRHLP) until convergence (until the relative variation of log-likelihood between two steps of the EM algorithm is less than the threshold parameter).

Value

EM returns an object of class ModelMixRHLP.

See Also

ModelMixRHLP, ParamMixRHLP, StatMixRHLP

FData-class *A Reference Class which represents functional data.*

Description

FData is a reference class which represents general independent and identically distributed (i.i.d.) functional objects. The data can be ordered by time (functional time series). In the last case, the field X represents the time.

Fields

X Numeric vector of length m representing the covariates/inputs.

Y Matrix of size (n, m) representing the observed responses/outputs. Y consists of n functions of X observed at points $1, \dots, m$.

`mkStochastic` *mkStochastic ensures that it is a stochastic vector, matrix or array.*

Description

mkStochastic ensures that it is a stochastic vector, matrix or array.

Usage

```
mkStochastic(M)
```

Arguments

`M`	A vector, matrix or array to transform.

Details

mkStochastic ensures that the giving argument is a stochastic vector, matrix or array, i.e., that the sum over the last dimension is 1.

Value

A vector, matrix or array for which the sum over the last dimension is 1.

`ModelMixHMM-class` *A Reference Class which represents a fitted MixHMM model.*

Description

ModelMixHMM represents an estimated MixHMM model.

Fields

`param` A ParamMixHMM object. It contains the estimated values of the parameters.

`stat` A StatMixHMM object. It contains all the statistics associated to the MixHMM model.

Methods

`plot(...)` Plot method

`...` Other graphics parameters.

`summary(digits = getOption("digits"))` Summary method.

`digits` The number of significant digits to use when printing.

See Also

ParamMixHMM, StatMixHMM

`ModelMixHMMR-class` *A Reference Class which represents a fitted MixHMMR model.*

Description

ModelMixHMMR represents an estimated MixHMMR model.

Fields

`param` A ParamMixHMMR object. It contains the estimated values of the parameters.

`stat` A StatMixHMMR object. It contains all the statistics associated to the MixHMMR model.

Methods

`plot(...)` Plot method

  `...` Other graphics parameters.

`summary(digits = getOption("digits"))` Summary method.

  `digits` The number of significant digits to use when printing.

See Also

ParamMixHMMR, StatMixHMMR

`ModelMixRHLP-class` *A Reference Class which represents a fitted MixRHLP model.*

Description

ModelMixRHLP represents an estimated MixRHLP model.

Fields

`param` A ParamMixRHLP object. It contains the estimated values of the parameters.

`stat` A StatMixRHLP object. It contains all the statistics associated to the MixRHLP model.

Methods

`plot(what = c("estimatedsignal", "regressors", "loglikelihood"), ...)` Plot method.

  `what` The type of graph requested:

  - `"estimatedsignal" =` Estimated signal (field `Ex` of class StatMixRHLP).
  - `"regressors" =` Polynomial regression components (fields `polynomials` and `pi_jgk` of class StatMixRHLP).
  - `"loglikelihood" =` Value of the log-likelihood for each iteration (field `stored_loglik` of class StatMixRHLP).

  `...` Other graphics parameters.

  By default, all the above graphs are produced.

`summary(digits = getOption("digits"))` Summary method.

  `digits` The number of significant digits to use when printing.

See Also

ParamMixRHLP, StatMixRHLP

`ParamMixHMM-class` *A Reference Class which contains parameters of a MixHMM model.*

Description

ParamMixHMM contains all the parameters of a MixHMM model.

Fields

`fData` FData object representing the sample (covariates/inputs `X` and observed responses/outputs `Y`).

`K` The number of clusters (Number of HMM models).

`R` The number of regimes (HMM components) for each cluster.

`variance_type` Character indicating if the model is homoskedastic (`variance_type = "homoskedastic"`) or heteroskedastic (`variance_type = "heteroskedastic"`). By default the model is heteroskedastic.

`alpha` Cluster weights. Matrix of dimension $(K, 1)$.

`prior` The prior probabilities of the Markov chains. `prior` is a matrix of dimension (R, K). The k-th column represents the prior distribution of the Markov chain asociated to the cluster k.

`trans_mat` The transition matrices of the Markov chains. `trans_mat` is an array of dimension (R, R, K).

`mask` Mask applied to the transition matrices `trans_mat`. By default, a mask of order one is applied.

`mu` Means. Matrix of dimension (R, K). The k-th column gives represents the k-th cluster and gives the means for the `R` regimes.

`sigma2` The variances for the `K` clusters. If MixHMM model is heteroskedastic (`variance_type = "heteroskedastic"`) then `sigma2` is a matrix of size (R, K) (otherwise MixHMM model is homoskedastic (`variance_type = "homoske` and `sigma2` is a matrix of size $(1, K)$).

`nu` The degree of freedom of the MixHMM model representing the complexity of the model.

Methods

`initGaussParamHmm(Y, k, R, variance_type, try_algo)` Initialize the means `mu` and `sigma2` for the cluster `k`.

`initParam(order_constraint = TRUE, init_kmeans = TRUE, try_algo = 1)` Method to initialize parameters `alpha`, `prior`, `trans_mat`, `mu` and `sigma2`.

If `init_kmeans = TRUE` then the curve partition is initialized by the K-means algorithm. Otherwise the curve partition is initialized randomly.

If `try_algo = 1` then `mu` and `sigma2` are initialized by segmenting the time series `Y` uniformly into `R` contiguous segments. Otherwise, `mu` and `sigma2` are initialized by segmenting randomly the time series `Y` into `R` segments.

`MStep(statMixHMM, order_constraint = TRUE)` Method which implements the M-step of the EM algorithm to learn the parameters of the MixHMM model based on statistics provided by the object `statMixHMM` of class StatMixHMM (which contains the E-step).

ParamMixHMMR-class *A Reference Class which contains parameters of a MixHMMR model.*

Description

ParamMixHMMR contains all the parameters of a MixHMMR model.

Fields

`fData` FData object representing the sample (covariates/inputs `X` and observed responses/outputs `Y`).

`K` The number of clusters (Number of HMMR models).

`R` The number of regimes (HMMR components) for each cluster.

`p` The order of the polynomial regression.

`variance_type` Character indicating if the model is homoskedastic (`variance_type = "homoskedastic"`) or heteroskedastic (`variance_type = "heteroskedastic"`). By default the model is heteroskedastic.

`alpha` Cluster weights. Matrix of dimension $(K, 1)$.

`prior` The prior probabilities of the Markov chains. `prior` is a matrix of dimension (R, K). The k-th column represents the prior distribution of the Markov chain asociated to the cluster k.

`trans_mat` The transition matrices of the Markov chains. `trans_mat` is an array of dimension (R, R, K).

`mask` Mask applied to the transition matrices `trans_mat`. By default, a mask of order one is applied.

`beta` Parameters of the polynomial regressions. `beta` is an array of dimension $(p + 1, R, K)$, with `p` the order of the polynomial regression. `p` is fixed to 3 by default.

`sigma2` The variances for the `K` clusters. If MixHMMR model is heteroskedastic (`variance_type = "heteroskedastic` then `sigma2` is a matrix of size (R, K) (otherwise MixHMMR model is homoskedastic (`variance_type = "homosk` and `sigma2` is a matrix of size

`nu` The degree of freedom of the MixHMMR model representing the complexity of the model.

`phi` A list giving the regression design matrices for the polynomial and the logistic regressions.

Methods

`initParam(order_constraint = TRUE, init_kmeans = TRUE, try_algo = 1)` Method to initialize parameters `alpha`, `prior`, `trans_mat`, `beta` and `sigma2`.

If `init_kmeans = TRUE` then the curve partition is initialized by the K-means algorithm. Otherwise the curve partition is initialized randomly.

If `try_algo = 1` then `beta` and `sigma2` are initialized by segmenting the time series `Y` uniformly into `R` contiguous segments. Otherwise, `beta` and `sigma2` are initialized by segmenting randomly the time series `Y` into `R` segments.

`initRegressionParam(Y, k, R, phi, variance_type, try_algo)` Initialize `beta` and `sigma2` for the cluster `k`.

`MStep(statMixHMMR, order_constraint = TRUE)` Method which implements the M-step of the EM algorithm to learn the parameters of the MixHMMR model based on statistics provided by the object `statMixHMMR` of class StatMixHMMR (which contains the E-step).

ParamMixRHLP-class *A Reference Class which contains parameters of a MixRHLP model.*

Description

ParamMixRHLP contains all the parameters of a MixRHLP model.

Fields

`fData` FData object representing the sample (covariates/inputs `X` and observed responses/outputs `Y`).

`G` The number of clusters (Number of RHLP models).

`K` The number of regimes (RHLP components) for each cluster.

`p` The order of the polynomial regression.

`q` The dimension of the logistic regression. For the purpose of segmentation, it must be set to 1.

`variance_type` Character indicating if the model is homoskedastic (`variance_type = "homoskedastic"`) or heteroskedastic (`variance_type = "heteroskedastic"`). By default the model is heteroskedastic.

`alpha` Cluster weights. Matrix of dimension $(1, G)$.

`W` Parameters of the logistic process. $\boldsymbol{W} = (\boldsymbol{w}_1, \dots, \boldsymbol{w}_G)$ is an array of dimension $(q + 1, K - 1, G)$, with $\boldsymbol{w}_g = (\boldsymbol{w}_{g,1}, \dots, \boldsymbol{w}_{g,K-1})$, $g = 1, \dots, G$, and `q` the order of the logistic regression. `q` is fixed to 1 by default.

`beta` Parameters of the polynomial regressions. $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_G)$ is an array of dimension $(p + 1, K, G)$, with $\boldsymbol{\beta}_g = (\boldsymbol{\beta}_{g,1}, \dots, \boldsymbol{\beta}_{g,K})$, $g = 1, \dots, G$, `p` the order of the polynomial regression. `p` is fixed to 3 by default.

`sigma2` The variances for the `G` clusters. If MixRHLP model is heteroskedastic (`variance_type = "heteroskedastic"` then `sigma2` is a matrix of size (K, G) (otherwise MixRHLP model is homoskedastic (`variance_type = "homoske` and `sigma2` is a matrix of size $(G, 1)$).

`nu` The degree of freedom of the MixRHLP model representing the complexity of the model.

`phi` A list giving the regression design matrices for the polynomial and the logistic regressions.

Methods

`CMStep(statMixRHLP, verbose_IRLS = FALSE)` Method which implements the M-step of the CEM algorithm to learn the parameters of the MixRHLP model based on statistics provided by the object `statMixRHLP` of class StatMixRHLP (which contains the E-step and the C-step).

`initParam(init_kmeans = TRUE, try_algo = 1)` Method to initialize parameters `alpha`, `W`, `beta` and `sigma2`.

If `init_kmeans = TRUE` then the curve partition is initialized by the K-means algorithm. Otherwise the curve partition is initialized randomly.

If `try_algo = 1` then `beta` and `sigma2` are initialized by segmenting the time series `Y` uniformly into `K` contiguous segments. Otherwise, `W`, `beta` and `sigma2` are initialized by segmenting randomly the time series `Y` into `K` segments.

`initRegressionParam(Xg, g, try_algo = 1)` Initialize the matrix of polynomial regression coefficients beta_g for the cluster `g`.

`MStep(statMixRHLP, verbose_IRLS = FALSE)` Method which implements the M-step of the EM algorithm to learn the parameters of the MixRHLP model based on statistics provided by the object `statMixRHLP` of class StatMixRHLP (which contains the E-step).

StatMixHMM-class *A Reference Class which contains statistics of a MixHMM model.*

Description

StatMixHMM contains all the statistics associated to a MixHMM model.

Fields

`tau_ik` Matrix of size (n, K) giving the posterior probabilities that the curve Y_i originates from the k-th HMM model.

`gamma_ikjr` Array of size (nm, R, K) giving the posterior probabilities that the observation Y_{ij} originates from the r-th regime of the k-th HMM model.

`loglik` Numeric. Log-likelihood of the MixHMM model.

`stored_loglik` Numeric vector. Stored values of the log-likelihood at each iteration of the EM algorithm.

`klas` Row matrix of the labels issued from `tau_ik`. Its elements are $klas(i) = k$, $i = 1, \dots, n$.

`z_ik` Hard segmentation logical matrix of dimension (n, K) obtained by the Maximum a posteriori (MAP) rule: $z\_ik = 1 \text{ if } z\_ik = \arg\max_s P(z_{is} = 1|\boldsymbol{Y_i}; \boldsymbol{\Psi}) = tau\_tk;\ 0$ otherwise, $k = 1, \dots, K$.

`smoothed` Matrix of size (m, K) giving the estimated mean series. The k-th column gives the estimated mean series of cluster k.

`mean_curve` To define.

`BIC` Numeric. Value of BIC (Bayesian Information Criterion).

`AIC` Numeric. Value of AIC (Akaike Information Criterion).

`ICL1` Numeric. Value of ICL (Integrated Completed Likelihood Criterion).

`log_w_k_fyi` Private. Only defined for calculations.

`exp_num_trans` Private. Only defined for calculations.

`exp_num_trans_from_l` Private. Only defined for calculations.

Methods

`computeStats(paramMixHMM)` Method used in the EM algorithm to compute statistics based on parameters provided by the object `paramMixHMM` of class ParamMixHMM.

`EStep(paramMixHMM)` Method used in the EM algorithm to update statistics based on parameters provided by the object `paramMixHMM` of class ParamMixHMM (prior and posterior probabilities).

`MAP()` MAP calculates values of the fields `z_ik` and `klas` by applying the Maximum A Posteriori Bayes allocation rule.
$z\_ik = 1 \text{ if } z\_ik = \arg\max_s P(z_{is} = 1|\boldsymbol{Y_i}; \boldsymbol{\Psi}) = tau\_tk;\ 0$ otherwise, $k = 1, \dots, K$.

See Also

ParamMixHMM

`StatMixHMMR-class` *A Reference Class which contains statistics of a MixHMMR model.*

Description

StatMixHMMR contains all the statistics associated to a MixHMMR model.

Fields

`tau_ik` Matrix of size (n, K) giving the posterior probabilities that the curve Y_i originates from the k-th HMM model.

`gamma_ikjr` Array of size (nm, R, K) giving the posterior probabilities that the observation Y_{ij} originates from the r-th regime of the k-th HMM model.

`loglik` Numeric. Log-likelihood of the MixHMMR model.

`stored_loglik` Numeric vector. Stored values of the log-likelihood at each iteration of the EM algorithm.

`klas` Row matrix of the labels issued from `tau_ik`. Its elements are $klas(i) = k$, $i = 1, \ldots, n$.

`z_ik` Hard segmentation logical matrix of dimension (n, K) obtained by the Maximum a posteriori (MAP) rule: $z\_ik = 1 \text{ if } z\_ik = \arg\max_s P(z_{is} = 1|\boldsymbol{Y_i}; \boldsymbol{\Psi}) = tau\_tk;\ 0$ otherwise, $k = 1, \ldots, K$.

`smoothed` Matrix of size (m, K) giving the estimated mean series. The k-th column gives the estimated mean series of cluster k.

`mean_curve` To define.

`BIC` Numeric. Value of BIC (Bayesian Information Criterion).

`AIC` Numeric. Value of AIC (Akaike Information Criterion).

`ICL1` Numeric. Value of ICL (Integrated Completed Likelihood Criterion).

`log_w_k_fyi` Private. Only defined for calculations.

`exp_num_trans` Private. Only defined for calculations.

`exp_num_trans_from_l` Private. Only defined for calculations.

Methods

`computeStats(paramMixHMMR)` Method used in the EM algorithm to compute statistics based on parameters provided by the object `paramMixHMMR` of class ParamMixHMMR.

`EStep(paramMixHMMR)` Method used in the EM algorithm to update statistics based on parameters provided by the object `paramMixHMMR` of class ParamMixHMMR (prior and posterior probabilities).

`MAP()` MAP calculates values of the fields `z_ik` and `klas` by applying the Maximum A Posteriori Bayes allocation rule.

$z\_ik = 1 \text{ if } z\_ik = \arg\max_s P(z_{is} = 1|\boldsymbol{Y_i}; \boldsymbol{\Psi}) = tau\_tk;\ 0$ otherwise

See Also

ParamMixHMMR

StatMixRHLP-class *A Reference Class which contains statistics of a MixRHLP model.*

Description

StatMixRHLP contains all the statistics associated to a MixRHLP model.

Fields

pi_jgk Array of size (nm, K, G) representing the logistic proportion for cluster g.

h_ig Matrix of size (n, G) giving the posterior probabilities (fuzzy segmentation matrix) that the curve Y_i originates from the g-th RHLP model.

c_ig Hard segmentation logical matrix of dimension (n, G) obtained by the Maximum a posteriori (MAP) rule: $c\_ig = 1$ if $c\_ig = \arg\max_s h\_is;\ 0$ otherwise, $g = 1, \dots, G$.

klas Column matrix of the labels issued from c_ig. Its elements are $klas(i) = g$, $i = 1, \dots, n$.

tau_ijgk Array of size (nm, K, G) giving the posterior probabilities that the observation Y_{ij} originates from the K-th regime of the g-th RHLP model.

polynomials Array of size (m, K, G) giving the values of the estimated polynomial regression components.

weighted_polynomials Array of size (m, K, G) giving the values of the estimated polynomial regression components weighted by the prior probabilities pi_jgk.

Ex Matrix of size *(m, G)*. Ex is the curve expectation (estimated signal): sum of the polynomial components weighted by the logistic probabilities pi_jgk.

loglik Numeric. Observed-data log-likelihood of the MixRHLP model.

com_loglik Numeric. Complete-data log-likelihood of the MixRHLP model.

stored_loglik Numeric vector. Stored values of the log-likelihood at each EM iteration.

stored_com_loglik Numeric vector. Stored values of the Complete log-likelihood at each EM iteration.

BIC Numeric. Value of BIC (Bayesian Information Criterion).

ICL Numeric. Value of ICL (Integrated Completed Likelihood).

AIC Numeric. Value of AIC (Akaike Information Criterion).

log_fg_xij Matrix of size (n, G) giving the values of the probability density function $f(y_i|h_i g = 1, \boldsymbol{x}, \boldsymbol{\Psi})$, $i = 1, \dots, n$.

log_alphag_fg_xij Matrix of size (n, G) giving the values of the logarithm of the joint probability density function $f(y_i,\ h_i = g|\boldsymbol{x}, \boldsymbol{\Psi})$, $i = 1, \dots, n$.

log_tau_ijgk Array of size (nm, K, G) giving the logarithm of tau_ijgk.

Methods

computeStats(paramMixRHLP) Method used in the EM algorithm to compute statistics based on parameters provided by the object paramMixRHLP of class ParamMixRHLP.

CStep(reg_irls) Method used in the CEM algorithm to update statistics.

EStep(paramMixRHLP) Method used in the EM algorithm to update statistics based on parameters provided by the object paramMixRHLP of class ParamMixRHLP (prior and posterior probabilities).

MAP() MAP calculates values of the fields c_ig and klas by applying the Maximum A Posteriori Bayes allocation rule.

$c\_ig = 1$ if $c\_ig = \arg\max_s h\_is;\ 0$ otherwise, $g = 1, \dots, G$.

See Also

ParamMixRHLP

`toydataset` *A dataset composed of simulated time series with regime changes.*

Description

A dataset composed of 30 simulated time series with regime changes.

Usage

```
toydataset
```

Format

A data frame with 350 rows and 31 variables:

x The covariate variable which is the time in that case.

y1 Times series with a wave form shape and for which a normally distributed random noise has been added.

y2 Same as `y1`.

y3 Same as `y1`.

y4 Same as `y1`.

y5 Same as `y1`.

y6 Same as `y1`.

y7 Same as `y1`.

y8 Same as `y1`.

y9 Same as `y1`.

y10 Same as `y1`.

y11 Time series generated as follows:

- First regime: 120 values of Normally distributed random numbers with mean 5.
- Second regime: 70 values of Normally distributed random numbers with mean 7.
- Third regime: 160 values of Normally distributed random numbers with mean 5.

y12 Same as `y11`.

y13 Same as `y11`.

y14 Same as `y11`.

y15 Same as `y11`.

y16 Same as `y11`.

y17 Same as `y11`.

y18 Same as `y11`.

y19 Same as `y11`.

y20 Same as `y11`.

y21 Time series generated as follows:

- First regime: 80 values of Normally distributed random numbers with mean 7.
- Second regime: 130 values of Normally distributed random numbers with mean 5.
- Third regime: 140 values of Normally distributed random numbers with mean 4.

y22 Same as `y21`.

y23 Same as `y21`.

y24 Same as `y21`.

y25 Same as `y21`.

y26 Same as `y21`.

y27 Same as `y21`.

y28 Same as `y21`.

y29 Same as `y21`.

y30 Same as `y21`.

Index