Package ‘fdacluster’

May 25, 2023

Title Joint Clustering and Alignment of Functional Data

Version 0.2.2


License GPL (>= 3)

Encoding UTF-8

LazyData true

LazyDataCompression xz

LinkingTo Rcpp, RcppArmadillo, nloptr

RoxygenNote 7.2.3

Suggests fda, funData, future, knitr, rmarkdown, testthat (>= 3.0.0)

Imports cli, cluster, dbscan, dplyr, fdsrvf, forcats, furrr, ggplot2, lpSolve, nloptr, progressr, purrr, Rcpp, rlang, tibble, tidyr, withr

Depends R (>= 3.5.0)


Config/testthat/edition 3


**VignetteBuilder** knitr

**NeedsCompilation** yes

**Author** Aymeric Stamm [aut, cre] (<https://orcid.org/0000-0002-8725-3654>),
Laura Sangalli [ctb],
Piercesare Secchi [ctb],
Simone Vantini [ctb],
Valeria Vitelli [ctb],
Alessandro Zito [ctb]

**Maintainer** Aymeric Stamm <aymeric.stamm@cnrs.fr>

**Repository** CRAN

**Date/Publication** 2023-05-25 10:20:02 UTC

**R topics documented:**

- autoplot.caps ........................................... 2
- autoplot.mcaps ........................................... 3
- caps ....................................................... 4
- compare_caps ............................................ 5
- diagnostic_plot ......................................... 7
- fdadbscan ................................................. 8
- fdadist ................................................... 10
- fdahclust ................................................. 11
- fdakmeans ............................................... 14
- plot.caps ............................................... 17
- plot.mcaps ............................................... 18
- sim30_caps ............................................... 19
- sim30_mcaps ............................................. 19
- simulated30 ............................................. 20
- simulated30_sub ......................................... 20
- simulated90 ............................................. 21

**Index** .................................................. 22

---

**autoplot.caps**

*Visualizes the result of a clustering strategy stored in a caps object with ggplot2*

**Description**

This function creates a visualization of the result of the k-mean alignment algorithm and invisibly returns the corresponding `ggplot2::ggplot` object which enable further customization of the plot. The user can choose to visualize either the amplitude information data in which case original and aligned curves are shown or the phase information data in which case the estimated warping functions are shown.
## Usage

```r
## S3 method for class 'caps'
autoplot(object, type = c("amplitude", "phase"), ...)
```

### Arguments

- **object**: An object of class `caps`.
- **type**: A string specifying the type of information to display. Choices are "amplitude" for plotting the original and aligned curves which represent amplitude information data or "phase" for plotting the corresponding warping functions which represent phase information data. Defaults to "amplitude".
- **...**: Not used.

### Value

A `ggplot2::ggplot` object invisibly.

### Examples

```r
ggplot2::autoplot(sim30_caps, type = "amplitude")
ggplot2::autoplot(sim30_caps, type = "phase")
```

---

**autoplot.mcaps**

*Visualizes results of multiple clustering strategies using ggplot2*

### Description

This is an S3 method implementation of the `ggplot2::autoplot()` generic for objects of class `mcaps` to visualize the performances of multiple `caps` objects applied on the same data sets either in terms of WSS or in terms of silhouette values.

### Usage

```r
## S3 method for class 'mcaps'
autoplot(
  object,
  validation_criterion = c("wss", "silhouette"),
  what = c("mean", "distribution"),
  ...
)
```
Arguments

object An object of class mcaps.
validation_criterion A string specifying the validation criterion to be used for the comparison. Choices are "wss" or "silhouette". Defaults to "wss".
what A string specifying the kind of information to display about the validation criterion. Choices are "mean" (which plots the mean values) or "distribution" (which plots the boxplots). Defaults to "mean".

Value

An object of class ggplot2::ggplot.

Examples

p <- ggplot2::ggplot(sim30_mcaps)

caps Class for clustering with amplitude and phase separation

Description

The k-means algorithm with joint amplitude and phase separation produces a number of outputs. This class is meant to encapsulate them into a single object for providing dedicated S3 methods for e.g. plotting, summarizing, etc. The name of the class stems from Clustering with Amplitude and Phase Separation.

Usage

as_caps(x)
is_caps(x)

Arguments

x A list coercible into an object of class caps.

Details

An object of class caps is a list with the following components:

- original_curves: A numeric matrix of shape $N \times L \times M$ storing a sample with the $N\times L$-dimensional original curves observed on grids of size $M$. 


compare_caps

- aligned_curves: A numeric matrix of shape $N \times L \times M$ storing a sample with the $N$ $L$-dimensional aligned curves observed on grids of size $M$.
- center_curves: A numeric matrix of shape $K \times L \times M$ storing the $K$ centers which are $L$-dimensional curves observed on a grid of size $M$;
- warplings: A numeric matrix of shape $N \times M$ storing the estimated warping functions for each of the $N$ curves evaluated on the within-cluster common grids of size $M$;
- grids: A numeric matrix of size $K \times M$ storing the within-cluster common grids of size $M$ on which curves are evaluated;
- n_clusters: An integer value storing the number of clusters;
- memberships: An integer vector of length $N$ storing the cluster ID which each curve belongs to;
- distances_to_center: A numeric vector of length $N$ storing the distance of each curve to the center of its cluster;
- silhouettes: A numeric vector of length $N$ storing the silhouette values of each observation;
- n_iterations: An integer value storing the number of iterations performed until convergence;
- call_name: A string storing the name of the function that was used to produce the k-means alignment results;
- call_args: A list containing the exact arguments that were passed to the function call_name that produced this output.

Value

The function `as_caps()` returns an object of class `caps`. The function `is_caps()` returns a boolean which evaluates to TRUE is the input object is of class `caps`.

Examples

```r
as_caps(sin30_caps)
is_caps(sin30_caps)
```

---

**compare_caps**

Generates results of multiple clustering strategies

**Description**

This function searches for clusters in the input data set using different strategies and generates an object of class `mcaps` which stores multiple objects of class `caps`. This is a helper function to facilitate comparison of clustering methods and choice of an optimal one.
Usage

\begin{verbatim}
compare_caps(
x, y,
n_clusters = 1:5,
metric = c("l2", "pearson"),
clustering_method = c("kmeans", "hclust-complete", "hclust-average", "hclust-single"),
warping_class = c("affine", "dilation", "none", "shift", "srsf"),
centroid_type = c("mean", "medoid", "lowess", "poly"),
cluster_on_phase = FALSE
)
\end{verbatim}

Arguments

\textbf{x} \quad A numeric vector of length \( M \) or a numeric matrix of shape \( N \times M \) or an object of class \texttt{funData::funData}. If a numeric vector or matrix, it specifies the grid(s) of size \( M \) on which each of the \( N \) curves have been observed. If an object of class \texttt{funData::funData}, it contains the whole functional data set and the \( y \) argument is not used.

\textbf{y} \quad Either a numeric matrix of shape \( N \times M \) or a numeric array of shape \( N \times L \times M \) or an object of class \texttt{fda::fd}. If a numeric matrix or array, it specifies the \( N \)-sample of \( L \)-dimensional curves observed on grids of size \( M \). If an object of class \texttt{fda::fd}, it contains all the necessary information about the functional data set to be able to evaluate it on user-defined grids.

\textbf{n_clusters} \quad An integer vector specifying a set of clustering partitions to create. Defaults to 1:5.

\textbf{metric} \quad A string specifying the metric used to compare curves. Choices are "l2" or "pearson". Defaults to "l2". Used only when \texttt{warping_class} \neq "srsf". For the boundary-preserving warping class, the L2 distance between the SRSFs of the original curves is used.

\textbf{clustering_method} \quad A character vector specifying one or more clustering methods to be fit. Choices are "kmeans", "hclust-complete", "hclust-average" or "hclust-single". Defaults to all of them.

\textbf{warping_class} \quad A character vector specifying one or more classes of warping functions to use for curve alignment. Choices are "affine", "dilation", "none", "shift" or "srsf". Defaults to all of them.

\textbf{centroid_type} \quad A character vector specifying one or more ways to compute centroids. Choices are "mean" or "medoid". Defaults to all of them.

\textbf{cluster_on_phase} \quad A boolean specifying whether clustering should be based on phase variation or amplitude variation. Defaults to \texttt{FALSE} which implies amplitude variation.

Value

An object of class \texttt{mcaps} which is a \texttt{tibble::tibble} storing the objects of class \texttt{caps} in correspondence of each combination of possible choices from the input arguments.
Examples

```r
# Extracts 15 out of the 30 simulated curves in `simulated30_sub` data set
idx <- c(1:5, 11:15, 21:25)
x <- simulated30_sub$x[idx, ]
y <- simulated30_sub$y[idx, , ]

# Runs an HAC with complete linkage with affine alignment, searching for up
# to 2 clusters and using the sample mean as centroid type:
out <- compare_caps(
  x = x,
  y = y,
  n_clusters = 2,
  warping_class = "affine",
  clustering_method = "hclust-complete",
  centroid_type = "mean"
)

# Then visualize the results
# Either with ggplot via ggplot2::autoplot(out)
# or using graphics::plot()
# You can visualize the WSS values:
plot(out, validation_criterion = "wss")
# Or the average silhouette values:
plot(out, validation_criterion = "silhouette")
```

diagnostic_plot  

*Diagnostic plot for the result of a clustering strategy stored in a* caps *object*

Description

This function plots the values of the distance to center and silhouette for each observation. Observations are ordered within cluster by decreasing value of silhouette.

Usage

```r
diagnostic_plot(x)
```

Arguments

- `x` An object of class caps.

Value

An object of class ggplot2::ggplot.
Examples

diagnostic_plot(sim30_caps)

fdadbscan

Performs density-based clustering for functional data with amplitude and phase separation

Description

This function extends DBSCAN to functional data. It includes the possibility to separate amplitude and phase information.

Usage

fdadbscan(
  x,
  y,
  warping_class = c("affine", "dilation", "none", "shift", "srsf"),
  centroid_type = "mean",
  metric = c("L2", "pearson"),
  cluster_on_phase = FALSE,
  use_verbose = TRUE,
  warping_options = c(0.15, 0.15),
  maximum_number_of_iterations = 100L,
  number_of_threads = 1L,
  parallel_method = 0L,
  distance_relative_tolerance = 0.001,
  use_fence = FALSE,
  check_total_dissimilarity = TRUE,
  compute_overall_center = FALSE
)

Arguments

x A numeric vector of length $M$ or a numeric matrix of shape $N \times M$ or an object of class funData::funData. If a numeric vector or matrix, it specifies the grid(s) of size $M$ on which each of the $N$ curves have been observed. If an object of class funData::funData, it contains the whole functional data set and the $y$ argument is not used.

y Either a numeric matrix of shape $N \times M$ or a numeric array of shape $N \times L \times M$ or an object of class fda::fd. If a numeric matrix or array, it specifies the $N$-sample of $L$-dimensional curves observed on grids of size $M$. If an object of class fda::fd, it contains all the necessary information about the functional data set to be able to evaluate it on user-defined grids.

warping_class A string specifying the warping class. Choices are "affine", "dilation", "none", "shift" or "srsf". Defaults to "affine". The SRSF class is the only class which is boundary-preserving.
centoid_type A string specifying the type of centroid to compute. Choices are "mean", "medoid", "lowess" or "poly". Defaults to "mean". If LOWESS approximation is chosen, the user can append an integer between 0 and 100 as in "lowess20". This number will be used as the smoother span. This gives the proportion of points in the plot which influence the smooth at each value. Larger values give more smoothness. The default value is 10%. If polynomial approximation is chosen, the user can append an positive integer as in "poly3". This number will be used as the degree of the polynomial model. The default value is 4L.

metric A string specifying the metric used to compare curves. Choices are "l2" or "pearson". Defaults to "l2". Used only when warping_class != "srsf". For the boundary-preserving warping class, the L2 distance between the SRSFs of the original curves is used.

cluster_on_phase A boolean specifying whether clustering should be based on phase variation or amplitude variation. Defaults to FALSE which implies amplitude variation.

use_verbose A boolean specifying whether the algorithm should output details of the steps to the console. Defaults to TRUE.

warping_options A numeric vector supplied as a helper to the chosen warping_class to decide on warping parameter bounds. This is used only when warping_class != "srsf".

maximum_number_of_iterations An integer specifying the maximum number of iterations before the algorithm stops if no other convergence criterion was met. Defaults to 100L.

number_of_threads An integer specifying the number of threads used for parallelization. Defaults to 1L. This is used only when warping_class != "srsf".

parallel_method An integer value specifying the type of desired parallelization for template computation. If 0L, templates are computed in parallel. If 1L, parallelization occurs within a single template computation (only for the medoid method as of now). Defaults to 0L. This is used only when warping_class != "srsf".

distance_relative_tolerance A numeric value specifying a relative tolerance on the distance update between two iterations. If all observations have not sufficiently improved in that sense, the algorithm stops. Defaults to 1e-3. This is used only when warping_class != "srsf".

use_fence A boolean specifying whether the fence algorithm should be used to robustify the algorithm against outliers. Defaults to FALSE. This is used only when warping_class != "srsf".

check_total_dissimilarity A boolean specifying whether an additional stopping criterion based on improvement of the total dissimilarity should be used. Defaults to TRUE. This is used only when warping_class != "srsf".

compute_overall_center A boolean specifying whether the overall center should be also computed. Defaults to FALSE. This is used only when warping_class != "srsf".
Value

An object of class \texttt{caps}.

Examples

\begin{verbatim}
# Extracts 15 out of the 30 simulated curves in `simulated30_sub` data set
idx <- c(1:5, 11:15)
x <- simulated30_sub$x[idx, ]
y <- simulated30_sub$y[idx, , ]

# Runs an HAC with affine alignment, searching for 2 clusters
out <- fdadbscan(
  x = x,
  y = y,
  warping_class = "affine"
)

# Then visualize the results
# Either with ggplot2 via ggplot2::autoplot(out)
# or using graphics::plot()
# You can visualize the original and aligned curves with:
plot(out, type = "amplitude")
# Or the estimated warping functions with:
plot(out, type = "phase")
\end{verbatim}

\textbf{fdadist} \hspace{1cm} Computes the distance matrix for functional data with amplitude and phase separation

Description

This function computes the matrix of pairwise distances between curves a functional data sample. This can be achieved with or without phase and amplitude separation, which can be done using a variety of warping classes.

Usage

\begin{verbatim}
fdadist(
  x,
  y = NULL,
  warping_class = c("affine", "dilation", "none", "shift", "srsf"),
  metric = c("l2", "pearson"),
  cluster_on_phase = FALSE,
  labels = NULL
)
\end{verbatim}
Arguments

x  A numeric vector of length \( M \) or a numeric matrix of shape \( N \times M \) or an object of class `funData::funData`. If a numeric vector or matrix, it specifies the grid(s) of size \( M \) on which each of the \( N \) curves have been observed. If an object of class `funData::funData`, it contains the whole functional data set and the \( y \) argument is not used.

\( y \)  Either a numeric matrix of shape \( N \times M \) or a numeric array of shape \( N \times L \times M \) or an object of class `fda::fd`. If a numeric matrix or array, it specifies the \( N \)-sample of \( L \)-dimensional curves observed on grids of size \( M \). If an object of class `fda::fd`, it contains all the necessary information about the functional data set to be able to evaluate it on user-defined grids.

warping_class  A string specifying the warping class. Choices are "affine", "dilation", "none", "shift" or "srsf". Defaults to "affine". The SRSF class is the only class which is boundary-preserving.

metric  A string specifying the metric used to compare curves. Choices are "l2" or "pearson". Defaults to "12". Used only when `warping_class` != "srsf". For the boundary-preserving warping class, the L2 distance between the SRSFs of the original curves is used.

cluster_on_phase  A boolean specifying whether clustering should be based on phase variation or amplitude variation. Defaults to `FALSE` which implies amplitude variation.

labels  A character vector specifying curve labels. Defaults to `NULL` which uses sequential numbers as labels.

Value

A `stats::dist` object storing the distance matrix between the input curves using the metric specified through the argument `metric` and the warping class specified by the argument `warping_class`.

Examples

```r
idx <- c(1:5, 11:15, 21:25)
D <- fdadist(simulated30_sub$x[idx, ], simulated30_sub$y[idx, , ])
```

Description

This function extends hierarchical agglomerative clustering to functional data. It includes the possibility to separate amplitude and phase information.
fdahclust

Usage

`fdahclust(
  x,
  y = NULL,
  n_clusters = 1L,
  warping_class = c("affine", "dilation", "none", "shift", "srsf"),
  centroid_type = "mean",
  metric = c("l2", "pearson"),
  linkage_criterion = c("complete", "average", "single", "ward.D2"),
  cluster_on_phase = FALSE,
  useVerbose = TRUE,
  warping_options = c(0.15, 0.15),
  maximum_number_of_iterations = 100L,
  number_of_threads = 1L,
  parallel_method = 0L,
  distance_relative_tolerance = 0.001,
  use_fence = FALSE,
  check_total_dissimilarity = TRUE,
  compute_overall_center = FALSE
)

Arguments

x
A numeric vector of length \( M \) or a numeric matrix of shape \( N \times M \) or an object of class `funData::funData`. If a numeric vector or matrix, it specifies the grid(s) of size \( M \) on which each of the \( N \) curves have been observed. If an object of class `funData::funData`, it contains the whole functional data set and the \( y \) argument is not used.

y
Either a numeric matrix of shape \( N \times M \) or a numeric array of shape \( N \times L \times M \) or an object of class `fda::fd`. If a numeric matrix or array, it specifies the \( N \)-sample of \( L \)-dimensional curves observed on grids of size \( M \). If an object of class `fda::fd`, it contains all the necessary information about the functional data set to be able to evaluate it on user-defined grids.

n_clusters
An integer value specifying the number of clusters. Defaults to 1L.

warping_class
A string specifying the warping class. Choices are "affine", "dilation", "none", "shift" or "srsf". Defaults to "affine". The SRSF class is the only class which is boundary-preserving.

centroid_type
A string specifying the type of centroid to compute. Choices are "mean", "medoid", "lowess" or "poly". Defaults to "mean". If LOWESS approximation is chosen, the user can append an integer between 0 and 100 as in "lowess20". This number will be used as the smoother span. This gives the proportion of points in the plot which influence the smooth at each value. Larger values give more smoothness. The default value is 10%. If polynomial approximation is chosen, the user can append an positive integer as in "poly3". This number will be used as the degree of the polynomial model. The default value is 4L.

metric
A string specifying the metric used to compare curves. Choices are "l2" or "pearson". Defaults to "l2". Used only when `warping_class` != "srsf". For
the boundary-preserving warping class, the L2 distance between the SRSFs of
the original curves is used.

**linkage_criterion**

A string specifying which linkage criterion should be used to compute
distances between sets of curves. Choices are "complete" for complete link-
age, "average" for average linkage and "single" for single linkage. See
`stats::hclust()` for more details. Defaults to "complete".

**cluster_on_phase**

A boolean specifying whether clustering should be based on phase variation or
amplitude variation. Defaults to FALSE which implies amplitude variation.

**use_verbose**

A boolean specifying whether the algorithm should output details of the steps to
the console. Defaults to TRUE.

**warping_options**

A numeric vector supplied as a helper to the chosen warping_class to decide on
warping parameter bounds. This is used only when warping_class != "srsf".

**maximum_number_of_iterations**

An integer specifying the maximum number of iterations before the algorithm
stops if no other convergence criterion was met. Defaults to 100L.

**number_of_threads**

An integer value specifying the number of threads used for parallelization. De-
defaults to 1L. This is used only when warping_class != "srsf".

**parallel_method**

An integer value specifying the type of desired parallelization for template com-
putation. If 0L, templates are computed in parallel. If 1L, parallelization occurs
within a single template computation (only for the medoid method as of now).
Defaults to 0L. This is used only when warping_class != "srsf".

**distance_relative_tolerance**

A numeric value specifying a relative tolerance on the distance update between
two iterations. If all observations have not sufficiently improved in that sense,
the algorithm stops. Defaults to 1e-3. This is used only when warping_class
!= "srsf".

**use_fence**

A boolean specifying whether the fence algorithm should be used to robus-
tify the algorithm against outliers. Defaults to FALSE. This is used only when
warping_class != "srsf".

**check_total_dissimilarity**

A boolean specifying whether an additional stopping criterion based on im-
provement of the total dissimilarity should be used. Defaults to TRUE. This is
used only when warping_class != "srsf".

**compute_overall_center**

A boolean specifying whether the overall center should be also computed. De-
defaults to FALSE. This is used only when warping_class != "srsf".

**Details**

The number of clusters is required as input because, with functional data, once hierarchical cluster-
ing is performed, curves within clusters need to be aligned to their corresponding centroid.
Value

An object of class \texttt{caps}.

Examples

```
#----------------------------------
# Extracts 15 out of the 30 simulated curves in `simulated30_sub` data set
idx <- c(1:5, 11:15, 21:25)
x <- simulated30_sub$x[idx, ]
y <- simulated30_sub$y[idx, , ]

#----------------------------------
# Runs an HAC with affine alignment, searching for 2 clusters
out <- fdahclust(
  x = x,
  y = y,
  n_clusters = 2,
  warping_class = "affine"
)

#----------------------------------
# Then visualize the results
# Either with ggplot2 via ggplot2::autoplot(out)
# or using graphics::plot()
# You can visualize the original and aligned curves with:
plot(out, type = "amplitude")
# Or the estimated warping functions with:
plot(out, type = "phase")
```

\texttt{fdakmeans} \hspace{1cm} \textit{Performs k-means clustering for functional data with amplitude and phase separation}

Description

This function provides implementations of the k-means clustering algorithm for functional data, with possible joint amplitude and phase separation. A number of warping class are implemented to achieve this separation.

Usage

```
fdakmeans(
  x,
  y = NULL,
  n_clusters = 1L,
  seeds = NULL,
  seeding_strategy = c("kmeans++", "exhaustive-kmeans++", "exhaustive", "hclust"),
  warping_class = c("affine", "dilation", "none", "shift", "srsf"),
  centroid_type = "mean",
)
metric = c("l2", "pearson"),
c cluster_on_phase = FALSE,
use_ verbose = TRUE,
warping_ options = c(0.15, 0.15),
maximum_number_of_iterations = 100L,
number_of_threads = 1L,
parallel_method = 0L,
distance_relative_tolerance = 0.001,
use_fence = FALSE,
check_total_dissimilarity = TRUE,
compute_overall_center = FALSE,
add_silhouettes = TRUE,
expand_domain = TRUE
)

Arguments

x A numeric vector of length M or a numeric matrix of shape N × M or an object of class funData::funData. If a numeric vector or matrix, it specifies the grid(s) of size M on which each of the N curves have been observed. If an object of class funData::funData, it contains the whole functional data set and the y argument is not used.

y Either a numeric matrix of shape N × M or a numeric array of shape N × L × M or an object of class fda::fd. If a numeric matrix or array, it specifies the N-sample of L-dimensional curves observed on grids of size M. If an object of class fda::fd, it contains all the necessary information about the functional data set to be able to evaluate it on user-defined grids.

n_clusters An integer value specifying the number of clusters. Defaults to 1L.

seeds An integer value or vector specifying the indices of the initial centroids. If an integer vector, it is interpreted as the indices of the initial centroids and should therefore be of length n_clusters. If an integer value, it is interpreted as the index of the first initial centroid and subsequent centroids are chosen according to the k-means++ strategy. It can be NULL in which case the argument seeding_strategy is used to automatically provide suitable indices. Defaults to NULL.

seeding_strategy A character string specifying the strategy for choosing the initial centroids in case the argument seeds is set to NULL. Choices are "kmeans++", "exhaustive-kmeans++" which performs an exhaustive search over the choice of the first centroid, "exhaustive" which tries on all combinations of initial centroids or "hclust" which first performs hierarchical clustering using Ward’s linkage criterion to identify initial centroids. Defaults to "kmeans++", which is the fastest strategy.

warping_class A string specifying the warping class. Choices are "affine", "dilation", "none", "shift" or "srsf". Defaults to "affine". The SRSF class is the only class which is boundary-preserving.

centroid_type A string specifying the type of centroid to compute. Choices are "mean", "medoid", "lowess" or "poly". Defaults to "mean". If LOWESS approximation is chosen, the user can append an integer between 0 and 100 as in "lowess20". This
number will be used as the smoother span. This gives the proportion of points in the plot which influence the smooth at each value. Larger values give more smoothness. The default value is 10%. If polynomial approximation is chosen, the user can append an positive integer as in "poly3". This number will be used as the degree of the polynomial model. The default value is 4L.

**metric**
A string specifying the metric used to compare curves. Choices are "l2" or "pearson". Defaults to "l2". Used only when `warping_class != "srsf"`. For the boundary-preserving warping class, the L2 distance between the SRSFs of the original curves is used.

**cluster_on_phase**
A boolean specifying whether clustering should be based on phase variation or amplitude variation. Defaults to `FALSE` which implies amplitude variation.

**use_verbose**
A boolean specifying whether the algorithm should output details of the steps to the console. Defaults to `TRUE`.

**warping_options**
A numeric vector supplied as a helper to the chosen `warping_class` to decide on warping parameter bounds. This is used only when `warping_class != "srsf"`.

**maximum_number_of_iterations**
An integer specifying the maximum number of iterations before the algorithm stops if no other convergence criterion was met. Defaults to 100L.

**number_of_threads**
An integer value specifying the number of threads used for parallelization. Defaults to 1L. This is used only when `warping_class != "srsf"`.

**parallel_method**
An integer value specifying the type of desired parallelization for template computation. If 0L, templates are computed in parallel. If 1L, parallelization occurs within a single template computation (only for the medoid method as of now). Defaults to 0L. This is used only when `warping_class != "srsf"`.

**distance_relative_tolerance**
A numeric value specifying a relative tolerance on the distance update between two iterations. If all observations have not sufficiently improved in that sense, the algorithm stops. Defaults to 1e-3. This is used only when `warping_class != "srsf"`.

**use_fence**
A boolean specifying whether the fence algorithm should be used to robustify the algorithm against outliers. Defaults to `FALSE`. This is used only when `warping_class != "srsf"`.

**check_total_dissimilarity**
A boolean specifying whether an additional stopping criterion based on improvement of the total dissimilarity should be used. Defaults to `TRUE`. This is used only when `warping_class != "srsf"`.

**compute_overall_center**
A boolean specifying whether the overall center should be also computed. Defaults to `FALSE`. This is used only when `warping_class != "srsf"`.

**add_silhouettes**
A boolean specifying whether silhouette values should be computed for each observation for internal validation of the clustering structure. Defaults to `TRUE`.
expand_domain  A boolean specifying how to define the within-cluster common grids. When set to FALSE, the intersection of the individual domains is used. When set to TRUE, the union is used and mean imputation is performed to fill in the missing values. Defaults to TRUE.

Value

An object of class caps.

Examples

#----------------------------------
# Extracts 15 out of the 30 simulated curves in `simulated30_sub` data set
idx <- c(1:5, 11:15, 21:25)
x <- simulated30_sub$x[idx, ]
y <- simulated30_sub$y[idx, , ]

#----------------------------------
# Runs a k-means clustering with affine alignment, searching for 2 clusters
out <- fdakmeans(
  x = x,
  y = y,
  n_clusters = 2,
  warping_class = "affine"
)

#----------------------------------
# Then visualize the results
# Either with ggplot2 via ggplot2::autoplot(out)
# or using graphics::plot()
# You can visualize the original and aligned curves with:
plot(out, type = "amplitude")
# Or the estimated warping functions with:
plot(out, type = "phase")

plot.caps  
Plots the result of a clustering strategy stored in a caps object

Description

This function creates a visualization of the result of the k-mean alignment algorithm without returning the plot data as an object. The user can choose to visualize either the amplitude information data in which case original and aligned curves are shown or the phase information data in which case the estimated warping functions are shown.

Usage

## S3 method for class 'caps'
plot(x, type = c("amplitude", "phase"), ...)
Arguments

x An object of class `caps`.

type A string specifying the type of information to display. Choices are "amplitude" for plotting the original and aligned curves which represent amplitude information data or "phase" for plotting the corresponding warping functions which represent phase information data. Defaults to "amplitude".

Examples

```
plot(sim30_caps, type = "amplitude")
plot(sim30_caps, type = "phase")
```

Description

This is an S3 method implementation of the `graphics::plot()` generic for objects of class `mcaps` to visualize the performances of multiple `caps` objects applied on the same data sets either in terms of WSS or in terms of silhouette values.

Usage

```
## S3 method for class 'mcaps'
plot(
  x, 
  validation_criterion = c("wss", "silhouette"), 
  what = c("mean", "distribution"), 
  ... 
)
```

Arguments

x An object of class `mcaps`.

validation_criterion A string specifying the validation criterion to be used for the comparison. Choices are "wss" or "silhouette". Defaults to "wss".

what A string specifying the kind of information to display about the validation criterion. Choices are "mean" (which plots the mean values) or "distribution" (which plots the boxplots). Defaults to "mean".

Examples

```
plot(sim30_mcaps)
```
**Description**

An object of class `caps` storing the result of the `fdakmeans()` function applied on the data set `simulated30` using the affine warping class and the Pearson metric and searching for 2 clusters.

**Usage**

`sim30_caps`

**Format**

An object of class `caps`.

---

**Description**

An object of class `mcaps` storing the result of the `compare_caps()` function applied on the data set `simulated30_sub` for comparing the clustering structures found by the `fdakmeans()` function with mean centroid type used with various classes of warping functions and varying number of clusters.

**Usage**

`sim30_mcaps`

**Format**

An object of class `mcaps` which is effectively a `tibble::tibble` with 5 columns and as many rows as there are clustering strategies to compare. The 5 column-variables are:

- `n_clusters`: The number of clusters;
- `clustering_method`: The clustering method;
- `warping_class`: The class of warping functions used for curve alignment;
- `centroid_type`: The type of centroid used to compute a cluster representative;
- `caps_obj`: The result of the corresponding clustering strategy as objects of class `caps`. 
**simulated30**  
*Simulated data for examples*

---

**Description**

A data set containing 30 simulated uni-dimensional curves.

**Usage**

`simulated30`

**Format**

A list with abscissas `x` and values `y`:

- `x` Matrix 30x200;
- `y` Array 30x1x200.

---

**simulated30_sub**  
*Simulated data for examples*

---

**Description**

A data set containing 30 simulated uni-dimensional curves.

**Usage**

`simulated30_sub`

**Format**

A list with abscissas `x` and values `y`:

- `x` Matrix 30x30;
- `y` Array 30x1x30.
simulated90

Simulated data from the CSDA paper

Description

A data set containing 90 simulated uni-dimensional curves.

Usage

simulated90

Format

A list with abscissas $x$ and values $y$:

- $x$ Vector of size 100;
- $y$ Matrix of size 90x100.
Index

* datasets
  sim30_caps, 19
  sim30_mcaps, 19
  simulated30, 20
  simulated30_sub, 20
  simulated90, 21

as_caps(caps), 4
as_caps(), 5
autoplot.caps, 2
autoplot.mcaps, 3

caps, 3, 4, 4, 5–7, 10, 14, 17–19
compare_caps, 5
compare_caps(), 19
diagnostic_plot, 7

fda::fd, 6, 8, 11, 12, 15
fdadbscan, 8
fdadist, 10
fdahclust, 11
fdakmeans, 14
fdakmeans(), 19
funData::funData, 6, 8, 11, 12, 15

ggplot2::autoplot(), 3
ggplot2::ggplot, 2–4, 7
graphics::plot(), 18

is_caps(caps), 4
is_caps(), 5

plot.caps, 17
plot.mcaps, 18

sim30_caps, 19
sim30_mcaps, 19
simulated30, 19, 20
simulated30_sub, 19, 20
simulated90, 21

stats::dist, 11
stats::hclust(), 13
tibble::tibble, 6, 19

22