Package ‘EDOtrans’

November 10, 2021

Type Package
Title Euclidean Distance-Optimized Data Transformation
Version 0.2.0
Author Jorn Lotsch[aut,cre], Alfred Ultsch[aut]
Maintainer Jorn Lotsch <j.lotsch@em.uni-frankfurt.de>
Description A data transformation method which takes into account the special property of scale non-invariance with a breakpoint at 1 of the Euclidean distance.
License GPL-3
Encoding UTF-8
LazyData true
Imports methods, stats, utils, ABCanalysis, opGMMassessment
Depends R (>= 3.5.0)
NeedsCompilation no
Repository CRAN
Date/Publication 2021-11-10 14:10:02 UTC

R topics documented:

EDOtrans .......................... 2
FACSData ............................ 3
GMMartificialData .................. 3

Index .............................. 4
**Description**

The package provides the necessary functions for performing the EDO data transformation.

**Usage**

```r
EDOtrans(Data, Cls, PlotIt = FALSE, FitAlg = "normalmixEM", Criterion = "LR", MaxModes = 8, MaxCores = 2048, Seed)
```

**Arguments**

- **Data**  
  the data as a vector.
- **Cls**  
  the class information, if any, as a vector of similar length as instances in the data.
- **PlotIt**  
  whether to plot the fit directly.
- **FitAlg**  
  which fit algorithm to use: "ClusterRGMM" = GMM from ClusterR, "densityMclust" from mclust, "DO" from DistributionOptimization (slow), "MCMC" = NMixMCMC from mixAK, or "normalmixEM" from mixtools.
- **Criterion**  
  which criterion should be used to establish the number of modes from the best GMM fit: "AIC", "BIC", "FM", "GAP", "LR" (likelihood ratio test), "NbClust" (from NbClust), "SI" (Silverman).
- **MaxModes**  
  for automated GMM assessment: the maximum number of modes to be tried.
- **MaxCores**  
  for automated GMM assessment: the maximum number of processor cores used under Unix.
- **Seed**  
  seed parameter set internally.

**Value**

Returns a list of transformed data and class assignments.

- **DataEDO**  
  the EDO transformed data.
- **EDOfactor**  
  the factor by which each data point has been divided.
- **Cls**  
  the class information for each data instance.

**Author(s)**

Jorn Lotsch and Alfred Ultsch

**References**

## Examples

```r
## example 1
data(iris)
IrisEDOdata <- EDOtrans(Data = as.vector(iris[,1]), Cls = as.integer(iris$Species))
```

### Description

Example data of hematologic marker expression.

Data set of 4 flow cytometry-based lymphoma makers from 1559 cells from healthy subjects (class 1) and 1441 cells from lymphoma patients (class 2).

### Usage

```r
data("FACSdata")
```

### Details

Size 3000 x 4, stored in FACSdata$[FS, CDa, CDb, CDd] Original classes 2, stored in FACSdata$Cls

```
Examples
data(FACSdata)
str(FACSdata)
```

### GMMartificialData

Example data an artificial Gaussian mixture.

Dataset of 3000 instances with 3 variables that are Gaussian mixtures and belong to classes Cls = 1, 2, or 3, with different means and standard deviations and equal weights of 0.7, 0.3, and 0.1, respectively.

### Usage

```r
data("GMMartificialData")
```

### Details

Size 3000 x 3, stored in GMMartificialData$[Var1, Var2, Var3]

Classes 3, stored in GMMartificialData$Cls

```
Examples
data(GMMartificialData)
str(GMMartificialData)
```
Index

* EDOtrans
  EDOtrans, 2

* data transformation
  EDOtrans, 2

EDOtrans, 2

FACSdata, 3

GMMartificialData, 3