

# Package ‘EDOtrans’

March 4, 2021

**Type** Package

**Title** Euclidean Distance-Optimized Data Transformation

**Version** 0.1.0

**Author** Jorn Lotsch[aut,cre], Alfred Ultsch[aut]

**Maintainer** Jorn Lotsch <j.loetsch@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** ABCanalysis, AdaptGauss, DistributionOptimization, ClusterR,  
DataVisualizations, ggplot2, grDevices, methods, stats, rlang

**Depends** R (>= 3.5.0)

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2021-03-04 09:50:15 UTC

## R topics documented:

EDOtrans . . . . .	2
FlowcytometricData . . . . .	3
GMMartificialData . . . . .	3
GMMplotGG . . . . .	4
<b>Index</b>	<b>6</b>

EDOtrans

*Euclidean distance-optimized data transformation*

---

**Description**

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

**Usage**

```
EDOtrans(Data, Cls, Means, SDs, Weights, DO = FALSE, PlotGMM = FALSE)
```

**Arguments**

Data	the data as a vector.
Cls	the class information, if any, as a vector of similar length as instances in the data.
Means	a list of mean values for a Gaussian mixture.
SDs	a list of standard deviations for a Gaussian mixture.
Weights	a list of weights for a Gaussian mixture.
DO	whether to use the DistributionOptimization method for Gaussian mixture modeling.
PlotGMM	whether to plot the obtained Gaussian mixture model.

**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**

Lotsch, J. (2021): EDOtrans – an R Package for Euclidean distance-optimized data transformation.

**Examples**

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

---

FlowcytometricData     *Example data of hematologic marker expression.*

---

**Description**

Data set of 6 flow cytometry-based lymphoma makers from 55,843 cells from healthy subjects (class 1) and 55,843 cells from lymphoma patients (class 2).

**Usage**

```
data("FlowcytometricData")
```

**Details**

Size 111686 x 6 , stored in FlowcytometricData\$[Var\_1,Var\_2,Var\_3,Var\_4,Var\_5,Var\_6]

Original classes 2, stored in FlowcytometricData\$OrigCls GMM based classes, stored in FlowcytometricData\$[ClsVar\_1,ClsVar\_2]

Downsample membership classes 2, stored in FlowcytometricData\$Downsampled

**Examples**

```
data(FlowcytometricData)
str(FlowcytometricData)
```

---

GMMartificialData     *Example data an artificial Gaussioan mixture.*

---

**Description**

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**

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

**Details**

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

Classes 3, stored in GMMartificialData\$Cls

**Examples**

```
data(GMMartificialData)
str(GMMartificialData)
```

---

`GMMplotGG`*Plot of Gaussian mixtures*

---

**Description**

The function plots the components of a Gaussian mixture and superimposes them on a histogram of the data.

**Usage**

```
GMMplotGG(Data, Means, SDs, Weights, BayesBoundaries,  
SingleGausses = TRUE, Hist = FALSE, Bounds = TRUE, SumModes = TRUE, PDE = TRUE)
```

**Arguments**

<code>Data</code>	the data as a vector.
<code>Means</code>	a list of mean values for a Gaussian mixture.
<code>SDs</code>	a list of standard deviations for a Gaussian mixture.
<code>Weights</code>	a list of weights for a Gaussian mixture.
<code>BayesBoundaries</code>	a list of Bayesian boundaries for a Gaussian mixture.
<code>SingleGausses</code>	whether to plot the single Gaussian components as separate lines.
<code>Hist</code>	whether to plot a histogram of the original data.
<code>Bounds</code>	whether to plot the Bayesian boundaries for a Gaussian mixture as vertical lines.
<code>SumModes</code>	whether to plot the summed-up mixes.
<code>PDE</code>	whether to use the Pareto density estimation instead of the standard R density function.

**Value**

Returns a ggplot2 object.

`p1` the plot of Gaussian mixtures.

**Author(s)**

Jorn Lotsch and Alfred Ultsch

**References**

Lotsch, J. (2021): EDOtrans – an R Package for Euclidean distance-optimized data transformation.

**Examples**

```
## example 1
data(iris)
Means0 <- tapply(X = as.vector(iris[,1]), INDEX = as.integer(iris$Species), FUN = mean)
SDs0 <- tapply(X = as.vector(iris[,1]), INDEX = as.integer(iris$Species), FUN = sd)
Weights0 <- c(1/3, 1/3, 1/3)
GMM.Sepal.Length <- GMMplotGG(Data = as.vector(iris[,1]),
Means = Means0,
SDs = SDs0,
Weights = Weights0,
Hist = TRUE)
```

# Index

- \* **EDOtrans**

- EDOtrans, [2](#)

- GMMplotGG, [4](#)

- \* **GMMplotGG**

- GMMplotGG, [4](#)

- \* **data transformation**

- EDOtrans, [2](#)

- \* **data visualization**

- GMMplotGG, [4](#)

EDOtrans, [2](#)

FlowcytometricData, [3](#)

GMMartificialData, [3](#)

GMMplotGG, [4](#)