

# Package ‘opGMMassessment’

November 18, 2021

**Type** Package

**Title** Optimized Automated Gaussian Mixture Assessment

**Version** 0.3

**Author** Jorn Lotsch [aut,cre] (<<https://orcid.org/0000-0002-5818-6958>>),  
Sebastian Malkusch [aut] (<<https://orcid.org/0000-0001-6766-140X>>),  
Martin Maechler [ctb],  
Peter Rousseeuw [ctb],  
Anja Struyf [ctb],  
Mia Hubert [ctb],  
Kurt Hornik [ctb]

**Maintainer** Jorn Lotsch <j.lotsch@em.uni-frankfurt.de>

**Description** Necessary functions for optimized automated evaluation of the number and parameters of Gaussian mixtures in one-dimensional data. Various methods are available for parameter estimation and for determining the number of modes in the mixture.

**Depends** R (>= 3.5.0)

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**Imports** AdaptGauss, DataVisualizations, DistributionOptimization,  
cluster, mixtools, grDevices, methods, foreach, stats, utils,  
rlang, ggplot2, parallel, caTools, dplyr, mclust, mixAK,  
multimode, NbClust, ClusterR, doParallel

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2021-11-18 10:40:02 UTC

## R topics documented:

GMMplotGG . . . . .	2
Mixture3 . . . . .	3
OlfactionTDI . . . . .	3
opGMMassessment . . . . .	4

---

`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 Sebastian Malkusch

**References**

Lotsch, J., Malkusch S. (2021): `opGMMassessment` – an R Package for automated Gaussian mixture modeling.

**Examples**

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

---

Mixture3

*Example Gaussian mixture data.*


---

**Description**

Data set containing 1000 instances distributed according to a Gaussian mixture with  $m = [-10, 0, 10]$ ,  $s = [1, 2, 3]$ ,  $w = [0.07, 0.05, 0.88]$ .

**Usage**

```
data("Mixture3")
```

**Details**

Size 1000 x 1

**Examples**

```
data(Mixture3)
str(Mixture3)
```

---

OlfactionTDI

*Example data of olfactory subtest performance.*


---

**Description**

Data set containing the results of 3 odor subtests collected from 1,000 subjects and assigned to the clinical olfactory diagnostic categories of anosmia, hyposmia, or normosmia.

**Usage**

```
data("OlfactionTDI")
```

**Details**

Size 1000 x 3 , stored in `OlfactionTDI$[Threshold,Discrimination,Identification]` Original classes 3, stored in `OlfactionTDI$Class`

**Examples**

```
data(OlfactionTDI)
str(OlfactionTDI)
```

---

opGMMassessment	<i>Gaussian mixture assessment</i>
-----------------	------------------------------------

---

**Description**

The package provides the necessary functions for optimized automated evaluation of the number and parameters of Gaussian mixtures in one-dimensional data. It provides various methods for parameter estimation and for determining the number of modes in the mixture.

**Usage**

```
opGMMassessment(Data, FitAlg = "MCMC", Criterion = "LR",
  MaxModes = 8, MaxCores = getOption("mc.cores", 2L), PlotIt = FALSE, KS = TRUE, Seed)
```

**Arguments**

Data	the data as a vector.
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	the maximum number of modes to be tried.
MaxCores	the maximum number of processor cores used under Unix.
PlotIt	whether to plot the fit directly (plot will be stored nevertheless).
KS	perform a Kolmogorow-Smirnow test of the fit versus original distribution.
Seed	optional seed parameter set internally.

**Value**

Returns a list of Gaussian modes.

Cls	the classes to which the cases are assigned according to the Gaussian mode membership.
Means	means of the Gaussian modes.

SDs	standard deviations of the Gaussian modes.
Weights	weights of the Gaussian modes.
Boundaries	Bayesian boundaries between the Gaussian modes.
Plot	Plot of the obtained mixture.
KS	Results of the Kolmogorov-Smirnov test.

**Author(s)**

Jorn Lotsch and Sebastian Malkusch

**References**

Lotsch, J., Malkusch S. (2021): *opGMMassessment* – an R Package for automated Gaussian mixture modeling.

**Examples**

```
## example 1
data(iris)
opGMMassessment(Data = iris$Petal.Length,
  FitAlg = "normalmixEM",
  Criterion = "BIC",
  PlotIt = TRUE,
  MaxModes = 5,
  MaxCores = 1,
  Seed = 42)
```

# Index

- \* **Clustering**
    - opGMMassessment, 4
  - \* **GMMplotGG**
    - GMMplotGG, 2
  - \* **data visualization**
    - GMMplotGG, 2
  - \* **opGMMassessment**
    - opGMMassessment, 4
- GMMplotGG, 2
- Mixture3, 3
- OlfactionTDI, 3
- opGMMassessment, 4