

# Package: SAGMM (via r-universe)

October 15, 2024

**Type** Package

**Title** Clustering via Stochastic Approximation and Gaussian Mixture Models

**Version** 0.2.4

**Date** 2019-06-29

**Author** Andrew T. Jones, Hien D. Nguyen

**Maintainer** Andrew T. Jones <andrewthomasjones@gmail.com>

**Description** Computes clustering by fitting Gaussian mixture models (GMM) via stochastic approximation following the methods of Nguyen and Jones (2018) <doi:10.1201/9780429446177>. It also provides some test data generation and plotting functionality to assist with this process.

**License** GPL-3

**Encoding** UTF-8

**Imports** Rcpp (>= 0.12.13), MixSim, mclust, stats, lowmemtkmeans

**LinkingTo** Rcpp, RcppArmadillo

**RoxygenNote** 6.1.1

**Suggests** testthat, ggplot2

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2019-06-29 05:50:12 UTC

## Contents

gainFactors . . . . .	2
generateSimData . . . . .	2
SAGMM . . . . .	3
SAGMMFit . . . . .	3

<b>Index</b>	<b>5</b>
--------------	----------

---

gainFactors	<i>Return Gamma, a sequence of gain factors</i>
-------------	---

---

**Description**

Generate a series of gain factors.

**Usage**

```
gainFactors(Number, Burnin)
```

**Arguments**

Number	Number of values required.
Burnin	Number of 'Burnin' values at the beginning of sequence.

**Value**

Gamma, a vector of gain factors.

**Examples**

```
g<-gainFactors(10^4, 2*10^3)
```

---

generateSimData	<i>Generate data for simulations to test the SAGMM package..</i>
-----------------	--

---

**Description**

This function is primarily a convenience wrapper for MixSim.

**Usage**

```
generateSimData(ngroups = 5, Dimensions = 5, Number = 10^4)
```

**Arguments**

ngroups	Number of mixture components. Default 5.
Dimensions	number of Dimensions. Default 5.
Number	number of samples. Default 10^4.

**Value**

List of results: X, Y, simobject.

**Examples**

```
sims<-generateSimData(ngroups=10, Dimensions=10, Number=10^4)
sims<-generateSimData()
```

SAGMM

*SAGMM: A package for Clustering via Stochastic Approximation and Gaussian Mixture Models.*

**Description**

The SAGMM package allows for computation of gaussian mixture models using stochastic approximation to increase efficiency with large data sets. The primary function `SAGMMFit` allows this to be performed in a relative flexible manner.

**Author(s)**

Andrew T. Jones and Hien D. Nguyen

**References**

Nguyen & Jones (2018). Big Data-Appropriate Clustering via Stochastic Approximation and Gaussian Mixture Models. In *Data Analytics* (pp. 79-96). CRC Press.

SAGMMFit

*Clustering via Stochastic Approximation and Gaussian Mixture Models (GMM)*

**Description**

Fit a GMM via Stochastic Approximation. See Reference.

**Usage**

```
SAGMMFit(X, Y = NULL, Burnin = 5, ngroups = 5, kstart = 10,
plot = FALSE)
```

**Arguments**

X	numeric matrix of the data.
Y	Group membership (if known). Where groups are integers in 1:ngroups. If provided ngroups can
Burnin	Ratio of observations to use as a burn in before algorithm begins.
ngroups	Number of mixture components. If Y is provided, and groups is not then is overridden by Y.
kstart	number of kmeans starts to initialise.
plot	If TRUE generates a plot of the clustering.

**Value**

	A list containing
Cluster	The clustering of each observation.
plot	A plot of the clustering (if requested).
l2	Estimate of $\text{Lambda}^2$
ARI1	Adjusted Rand Index 1 - using k-means
ARI2	Adjusted Rand Index 2 - using GMM Clusters
ARI3	Adjusted Rand Index 3 - using intialiation k-means
KM	Initial K-means clustering of the data.
pi	The cluster proportions (vector of length ngroups)
tau	tau matrix of conditional probabilities.
fit	Full output details from inner C++ loop.

**Author(s)**

Andrew T. Jones and Hien D. Nguyen

**References**

Nguyen & Jones (2018). Big Data-Appropriate Clustering via Stochastic Approximation and Gaussian Mixture Models. In Data Analytics (pp. 79-96). CRC Press.

**Examples**

```
sims<-generateSimData(ngroups=10, Dimensions=10, Number=10^4)
res1<-SAGMMFit(sims$X, sims$Y)
res2<-SAGMMFit(sims$X, ngroups=5)
```

# Index

`gainFactors`, [2](#)

`generateSimData`, [2](#)

SAGMM, [3](#)

SAGMM-package (SAGMM), [3](#)

SAGMMfit, [3](#)