Package ‘RPEClust’
November 6, 2019

Title Random Projection Ensemble Clustering Algorithm
Version 0.1.0
Date 2019-11-05
Author L. Anderlucci [aut], F. Fortunato [aut, cre], A. Montanari [ctb]
Maintainer Francesca Fortunato <francesca.fortunato3@unibo.it>
Depends R (>= 3.6.0), clusteval
Imports mclust, clue
Description Implements the methodology proposed by Anderlucci, Fortunato and Montanari (2019) <arXiv:1909.10832> for high-dimensional unsupervised classification. The random projection ensemble clustering algorithm applies a Gaussian Mixture Model to different random projections of the high-dimensional data and selects a subset of solutions accordingly to the Bayesian Information Criterion, computed here as discussed in Raftery and Dean (2006) <doi:10.1198/016214506000000113>. The clustering results obtained on the selected projections are then aggregated via consensus to derive the final partition.
License GPL-3
Encoding UTF-8
LazyData true
RoxygenNote 6.1.1
NeedsCompilation no
Repository CRAN
Date/Publication 2019-11-06 12:50:05 UTC

R topics documented:

  generateRP .......................................................... 2
  Meat ................................................................. 2
  RPGMClu ............................................................ 3

Index
**generateRP**  
*Generation of random matrices*

**Description**

This function generates $B$ random matrices of dimension $p$ by $d$ by using the Haar measure.

**Usage**

```r
generateRP(p, d, B)
```

**Arguments**

- `p`  
The original number of variables.
- `d`  
The reduced dimension.
- `B`  
The number of projections.

**Value**

A single matrix of dimension $p$ by $d*B$ containing $B$ random matrices of dimension $p$ by $d$.

**Examples**

```r
R <- generateRP(p=100, d=2, B=10)
dim(R)
```

---

**Meat**  
*Meat Data*

**Description**

This is the near-infrared spectroscopic meat data used in Murphy, Dean and Raftery (2009) <doi:10.1214/09-AOAS279> and originally collected by McElhinney, Downey and Fearn (1999) <doi:10.1255/jnirs.245>.

**Usage**

```r
data(Meat)
```

**Format**

A list with two components:

- `x`  Homogenized raw meat spectra. A matrix with 231 rows and 1050 columns.
- `y`  A vector containing the true class memberships.
**RPGMMClu**

**Source**

**References**
Murphy, Dean and Raftery (2010) <doi:10.1214/09-AOAS279>

**Examples**
data(Meat)
Meat$x[1:5,1:5]
Meat$y

---

**RPGMMClu**

**Random Projection Ensemble Clustering Algorithm**

**Description**
This function allows to run the RPEClu algorithm.

**Usage**

`RPGMMClu(x, true.cl = NULL, g, d = NULL, c = 10, B = 1000, B.star = 100, modelNames = NULL, diagonal = FALSE, ensmethod = "DWH", seed = 101, verb = FALSE)`

**Arguments**

- `x` A numeric high-dimensional matrix where rows correspond to observations and columns correspond to variables.
- `true.cl` A vector containing the true cluster membership. If supplied, the Adjusted Rand Index (ARI) of the predicted clustering is also returned. By default is set to NULL.
- `g` The number of clusters.
- `d` The dimension of the projected space. If is NULL (default option), then `d = ⌈c * log(g)⌉`.
- `c` The constant which governs the dimension of the projected space if `d` is not provided. The default is set to 10.
- `B` The number of generated projections; the default is 1000.
- `B.star` The number of base models to retain in the final ensemble; the default is 100.
- `modelNames` A vector of character strings indicating the models to be fitted in the EM phase of the GMM. The help file for `Mclust` describes the available options.
- `diagonal` A logical value indicating whether the conditional covariate matrix has a restricted form, i.e. it is diagonal. The default is FALSE.
ensmethod A character string specifying the method for computing the clustering consen-
sus. See the cl_consensus help file for available options. The default is DWH.

seed A single value indicating the initializing seed for random generations.

verb A logical controlling if the progress number of projections is displayed during
the fitting procedure. By default is FALSE.

Value

The output components are as follows:

ensemble A list including:
  label.vec The vector of labels predicted by the ensemble of size B.star.
  ari The corresponding ARI (if true.cl is not NULL).

individual A list including:
  label.vec The vectors of labels predicted by each base model.
  ari The corresponding ARI (if true.cl is not NULL).
  bic The BIC associated to each base model.
  bic.GMM The BIC associated to the Gaussian mixture fitted on each projected
  data.
  bic.reg The BIC for the linear regression of the \((p - d)\) last columns of \(Y^*\) on
  the first \(d\) ones.

References


Examples

data(Meat)
out.clu <- RPGMMClu(Meat$x, Meat$y, g=5, B=1000, B.star=100, verb=TRUE)

data <- sim_normal(n = rep(100, 2), p = 100, rho = rep(0.1, 2), delta = 0.5, sigma2 = 1, seed = 106)
out.clu <- RPGMMClu(data$x, data$y, g=2, B=10, B.star=5, verb=TRUE)
Index

*Topic datasets
   Meat, 2

c1_consensus, 4

generateRP, 2

Mclust, 3
Meat, 2

RPGMMClu, 3