Package ‘ManlyMix’

April 25, 2018

Version 0.1.14
Date 2018-04-25
Title Manly Mixture Modeling and Model-Based Clustering
Depends R (>= 3.0.0)
LazyLoad yes
LazyData no
Description The utility of this package includes finite mixture modeling and model-based clus-
tering through Manly mixture models by Zhu and Mel-
ykov (2016) <DOI:10.1016/j.csda.2016.01.015>. It also provides capabilities for for-
ward and backward model selection procedures.
License GPL (>= 2)
Author Xuwen Zhu [aut, cre],
Volodymyr Melnykov [aut],
Michael Hutt [ctb, cph] (NM optimization in c),
Stephen Moshier [ctb, cph] (eigen calculations in c),
Rouben Rostamian [ctb, cph] (memory allocation in c)
Maintainer Xuwen Zhu <xuwen.zhu@louisville.edu>
NeedsCompilation yes
Repository CRAN
Date/Publication 2018-04-25 20:09:41 UTC

R topics documented:

- ManlyMix-package .................................................. 2
- acidity ................................................................ 3
- ais ..................................................................... 4
- bankruptcy ............................................................. 5
- ClassAgree ............................................................ 6
- Manly.EM .............................................................. 6
- Manly.Kmeans ......................................................... 8
- Manly.model .......................................................... 10
- Manly.overlap ......................................................... 11
ManlyMix-package

Description

The utility of this package includes finite mixture modeling and model-based clustering based on Manly mixtures as well as forward and backward model selection procedures.

Details

<table>
<thead>
<tr>
<th>Package</th>
<th>ManlyMix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Package</td>
</tr>
<tr>
<td>Version</td>
<td>0.1.7</td>
</tr>
<tr>
<td>Date</td>
<td>2016-12-01</td>
</tr>
<tr>
<td>License</td>
<td>GPL (&gt;= 2)</td>
</tr>
<tr>
<td>LazyLoad</td>
<td>no</td>
</tr>
</tbody>
</table>

Function 'Manly.sim' simulates Manly mixture datasets.
Function 'Manly.overlap' estimates the pairwise overlaps for a Manly mixture.
Function 'Manly.EM' runs the EM algorithm for Manly mixture models.
Function 'Manly.select' runs forward and backward model selection procedures.
Function 'Manly.Kmeans' runs k-means model with Manly transformation.
Function 'Manly.var' produces the variance-covariance matrix of the parameter estimates from Manly mixture model.
Function 'Manly.plot' produces the density plot or contour plot of Manly mixture.
Function 'Manly.model' incorporates all Manly mixture related functionality.

Author(s)

Xuwen Zhu and Volodymyr Melnykov.
Maintainer: Xuwen Zhu <xuwen.zhu@louisville.edu>

References


Examples

```r
set.seed(123)
K <- 3; p <- 4
X <- as.matrix(iris[, -5])
id.true <- rep(1:K, each = 50)

# Obtain initial memberships based on the K-means algorithm
id.km <- kmeans(X, K)$cluster

# Run the CEM algorithm for Manly K-means model
la <- matrix(0.1, K, p)
C <- Manly.Kmeans(X, id = id.km, la = la)

# Run the EM algorithm for a Gaussian mixture model based on K-means solution
G <- Manly.EM(X, id = id.km)
id.G <- G$id

# Run FORWARD SELECTION ('silent' is on)
F <- Manly.select(X, model = G, method = "forward", silent = TRUE)

# Run the EM algorithm for a full Manly mixture model based on Gaussian mixture solution
la <- matrix(0.1, K, p)
M <- Manly.EM(X, id = id.G, la = la)

# Run BACKWARD SELECTION ('silent' is off)
B <- Manly.select(X, model = M, method = "backward")

BICs <- c(G$bic, M$bic, F$bic, B$bic)
names(BICs) <- c("Gaussian", "Manly", "Forward", "Backward")
BICs
```

### acidity

#### Description

Acidity index measured in a sample of 155 lakes in the Northeastern United States. The data are on the log scale.

#### Usage

```r
data(acidity)
```
Format

A data vector with 155 observations on the acidity index.

Details

The data was first analysed by Crawford et al. (1994).

References


Examples

```r
data(acidity)
```

---

ais  
*Australian Institute of Sport data*

Description

Data on 102 male and 100 female athletes collected at the Australian Institute of Sport, courtesy of Richard Telford and Ross Cunningham.

Usage

```r
data(ais)
```

Format

A data frame with 202 observations on the following 13 variables.

- **sex** Factor with levels: female, male;
- **sport** Factor with levels: B_Ball, Field, Gym, Netball, Row Swim, T_400m, Tennis, T_Sprnt, W_Polo;
- RCC  Red cell count;
- WCC  White cell count;
- Hc   Hematocrit;
- Hg   Hemoglobin;
- Fe   Plasma ferritin concentration;
- BMI  Body Mass Index;
- SSF  Sum of skin folds;
- Bfat Body fat percentage;
Details
The data have been made publicly available in connection with the book by Cook and Weisberg (1994).

References

Examples
data(ais)

data(bankruptcy)

Description
The data set contain the ratio of retained earnings (RE) to total assets, and the ratio of earnings before interests and taxes (EBIT) to total assets of 66 American firms recorded in the form of ratios. Half of the selected firms had filed for bankruptcy.

Usage
data(bankruptcy)

Format
A data frame with the following variables:
- Y The status of the firm: 0 bankruptcy or 1 financially sound;
- RE Ratio of retained earnings to total assets;
- EBIT Ratio of earnings before interests and taxes to total assets

References

Examples
data(bankruptcy)
ClassAgree  

\textit{Calculates the confusion matrix and number of misclassifications}

\textbf{Description}

Calculates the confusion matrix and number of misclassifications.

\textbf{Usage}

\texttt{ClassAgree(\textit{est.id, trueid})}

\textbf{Arguments}

\begin{itemize}
  \item \texttt{est.id}  \hspace{1cm} estimated membership vector
  \item \texttt{trueid}  \hspace{1cm} true membership vector
\end{itemize}

\textbf{Value}

\begin{itemize}
  \item \texttt{ClassificationTable}  \hspace{1cm} confusion table between true and estimated partitions
  \item \texttt{MisclassificationNum}  \hspace{1cm} number of misclassifications
\end{itemize}

\textbf{Examples}

\begin{verbatim}
set.seed(123)
K <- 3; p <- 4
X <- as.matrix(iris[, -5])
id.true <- rep(1:K, each = 50)

# Obtain initial memberships based on the K-means algorithm
id.km <- kmeans(X, K)$cluster

ClassAgree(id.km, id.true)
\end{verbatim}

---

\textbf{Manly.EM}

\textit{EM algorithm for Manly mixture model}

\textbf{Description}

Runs the EM algorithm for a Manly mixture model with specified initial membership and transformation parameters.
Usage

Manly.EM(X, id = NULL, la = NULL, tau = NULL, Mu = NULL, S = NULL,
tol = 1e-5, max.iter = 1000)

Arguments

- **x**: dataset matrix (n x p)
- **id**: initial membership vector (length n)
- **la**: initial transformation parameters (K x p)
- **tau**: initial vector of mixing proportions (length K)
- **Mu**: initial matrix of mean vectors (K x p)
- **S**: initial array of covariance matrices (p x p x K)
- **tol**: tolerance level
- **max.iter**: maximum number of iterations

Details

Runs the EM algorithm for a Manly mixture model for a provided dataset. Manly mixture model assumes that a multivariate Manly transformation applied to each component allows to reach near-normality. A user has a choice to specify either initial id vector 'id' and transformation parameters 'la' or initial mode parameters 'la', 'tau', 'Mu', and 'S'. In the case when transformation parameters are not provided, the function runs the EM algorithm without any transformations, i.e., it is equivalent to the EM algorithm for a Gaussian mixture model. If some transformation parameters have to be excluded from the consideration, in the corresponding positions of matrix 'la', the user has to specify value 0. Notation: n - sample size, p - dimensionality of the dataset X, K - number of mixture components.

Value

- **la**: matrix of the estimated transformation parameters (K x p)
- **tau**: vector of mixing proportions (length K)
- **Mu**: matrix of the estimated mean vectors (K x p)
- **S**: array of the estimated covariance matrices (p x p x K)
- **gamma**: matrix of posterior probabilities (n x K)
- **id**: estimated membership vector (length n)
- **ll**: log likelihood value
- **bic**: Bayesian Information Criterion
- **iter**: number of EM iterations run
- **flag**: convergence flag (0 - success, 1 - failure)

See Also

Manly.select
**Examples**

```r
set.seed(123)

K <- 3; p <- 4
X <- as.matrix(iris[, -5])
id.true <- rep(1:K, each = 50)

# Obtain initial memberships based on the K-means algorithm
id.km <- kmeans(X, K)$cluster

# Run the EM algorithm for a Gaussian mixture model based on K-means solution
A <- Manly.EM(X, id.km)
id.Gauss <- A$id

ClassAgree(id.Gauss, id.true)

# Run the EM algorithm for a Manly mixture model based on Gaussian mixture solution
la <- matrix(0.1, K, p)
B <- Manly.EM(X, id.Gauss, la)
id.Manly <- B$id

ClassAgree(id.Manly, id.true)
```

---

**Manly.Kmeans**

\[ \text{\textit{k-means algorithm with Manly transformation}} \]

---

**Description**

Runs the CEM algorithm for k-means clustering with specified initial membership and transformation parameters.

**Usage**

```r
Manly.Kmeans(X, id = NULL, la = NULL, Mu = NULL, S = NULL, initial = "k-means", K = NULL, nstart = 100, method = "ward.D", tol = 1e-5, max.iter = 1000)
```

**Arguments**

- `X`: dataset matrix (n x p)
- `id`: initial membership vector (length n)
- `la`: initial transformation parameters (K x p)
- `Mu`: initial matrix of mean vectors (K x p)
- `S`: initial vector of variances (K)
Manly.Kmeans

initial  initialization strategy of the EM algorithm ("k-means" - partition obtained by k-means clustering, "hierarchical" - partition obtained by hierarchical clustering)
K  number of clusters for the k-means initialization
nstart  number of random starts for the k-means initialization
method  linkage method for the hierarchical initialization
tol  tolerance level
max.iter  maximum number of iterations

Details
Runs the CEM algorithm for k-means clustering with Manly transformation for a provided dataset. The model assumes that a multivariate Manly transformation applied to each component allows to reach near-normality. A user has a choice to specify either initial id vector 'id' and transformation parameters 'la' or initial mode parameters 'la', 'Mu', and 'S'. In the case when transformation parameters are not provided, the function runs the EM algorithm without any transformations, i.e., it is equivalent to the EM algorithm for a k-means model. If some transformation parameters have to be excluded from the consideration, in the corresponding positions of matrix 'la', the user has to specify value 0. Notation: n - sample size, p - dimensionality of the dataset X, K - number of mixture components.

Value
la  matrix of the estimated transformation parameters (K x p)
Mu  matrix of the estimated mean vectors (K x p)
S  array of the estimated covariance matrices (K)
id  estimated membership vector (length n)
iter  number of EM iterations run
flag  convergence flag (0 - success, 1 - failure)

See Also
Manly.EM

Examples

set.seed(123)
K <- 3; p <- 4
X <- as.matrix(iris[, -5])
id.true <- rep(1:K, each = 50)

# Obtain initial memberships based on the traditional K-means algorithm
id.km <- kmeans(X, K)$cluster

# Run the CEM algorithm for k-means with Manly transformation based on traditional k-means solution
la <- matrix(0.1, K, p)
B <- Manly.Kmeans(X, id.km, la)
id.Manly <- B$id

ClassAgree(id.Manly, id.true)

---

**Description**

Runs all the functionality of a Manly mixture model.

**Usage**

```r
Manly.model(X, K = 1:5, Gaussian = FALSE, initial = "k-means",
nstart = 100, method = "ward.D", short.iter = 5,
select = "none", silent = TRUE, plot = FALSE, var1 = NULL,
var2 = NULL, VarAssess = FALSE, conf.CI = NULL, overlap = FALSE, N = 1000,
tol = 1e-5, max.iter = 1000, ...
```

**Arguments**

- `X`: dataset matrix (n x p)
- `K`: number of components tested
- `Gaussian`: whether Gaussian mixture models are run or not
- `initial`: initialization strategy of the EM algorithm ("k-means" - partition obtained by k-means clustering, "hierarchical" - partition obtained by hierarchical clustering, "emEM" - parameters estimated by the emEM algorithm)
- `nstart`: number of random starts for the k-means or the emEM initialization
- `method`: linkage method for the hierarchical initialization
- `short.iter`: number of short emEM iterations to run
- `select`: control to run Manly.select or not ("none" - do not run Manly.select, "forward" - run forward selection, "backward" - run backward selection)
- `silent`: control the output from Manly.select
- `plot`: control to construct the density or contour plot or not
- `var1`: x-axis variable for contour plot or variable for density plot
- `var2`: y-axis variable for contour plot
- `VarAssess`: run the variability assessment of the Manly mixture model or not
- `conf.CI`: specify the confidence level of parameter estimates
- `overlap`: estimate the overlap of Manly mixture components or not
- `N`: number of Monte Carlo simulations to run in the Manly.overlap function
- `tol`: tolerance level
- `max.iter`: maximum number of iterations
- `...`: further arguments related to `Manly.plot`
Details

Wrapper function that incorporates all functionality associated with Manly mixture modeling.

Value

<table>
<thead>
<tr>
<th>Model</th>
<th>best mixture model obtained</th>
</tr>
</thead>
<tbody>
<tr>
<td>VarAssess</td>
<td>estimated variance-covariance matrix for model parameter estimates</td>
</tr>
<tr>
<td>Overlap</td>
<td>estimated overlap of Manly mixture components</td>
</tr>
</tbody>
</table>

See Also

Manly.EM

Examples

```r
set.seed(123)

K <- 3; p <- 4
X <- as.matrix(iris[, -5])
id.true <- rep(1:K, each = 50)

Obj <- Manly.model(X, K = 1:5, initial = "emEM", nstart = 1, short.iter = 5)
```

---

**Manly.overlap**

Estimates the overlap for a Manly mixture

Description

Estimates the pairwise overlap matrix for a Manly mixture by simulating samples based on user-specified parameters.

Usage

`Manly.overlap(tau, Mu, S, la, N = 1000)`

Arguments

<table>
<thead>
<tr>
<th>la</th>
<th>matrix of transformation parameters (K x p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tau</td>
<td>vector of mixing proportions (length K)</td>
</tr>
<tr>
<td>Mu</td>
<td>matrix of mean vectors (K x p)</td>
</tr>
<tr>
<td>S</td>
<td>array of covariance matrices (p x p x K)</td>
</tr>
<tr>
<td>N</td>
<td>number of samples simulated</td>
</tr>
</tbody>
</table>
**Details**

Estimates the pairwise overlap matrix for a Manly mixture. Overlap is defined as sum of two misclassification probabilities.

**Value**

- OmegaMap: matrix of misclassification probabilities (K x K); OmegaMap[i,j] is the probability that X coming from the i-th component is classified to the j-th component.
- BarOmega: value of average overlap.
- MaxOmega: value of maximum overlap.

**References**


**Examples**

```r
set.seed(123)
# sets the number of components, dimensionality and sample size
K <- 3
p <- 2

# sets the mixture parameters
tau <- c(0.25, 0.3, 0.45)
Mu <- matrix(c(4.5,4,5,7,8,5.5), 3)
lA <- matrix(c(0.2,0.5,0.3,0.25,0.35,0.4), 3)
S <- array(NA, dim = c(p,p,K))
S[,1] <- matrix(c(0.4,0,0,0.4), 2)
S[,2] <- matrix(c(1,-0.2,-0.2,0.6), 2)
S[,3] <- matrix(c(2,-1,-1,2), 2)

# computes the overlap
A <- Manly.overlap(tau, Mu, S, la)
print(A)
```

---

*Manly.plot*  
*Density plot or contour plot for Manly mixture model*

**Description**

Provides a contour plot or a density plot for the fitted data with Manly mixture model.
Usage

Manly.plot(X, var1 = NULL, var2 = NULL, model = NULL, x.slice = 100,
y.slice = 100, x.mar = 1, y.mar = 1, col = "lightgrey", ...)

Arguments

X       dataset matrix (n x p)
var1    x-axis variable for contour plot or variable for density plot
var2    y-axis variable for contour plot
model   fitted Manly mixture model
x.slice number of slices in the first variable sequence in the contour
y.slice number of slices in the second variable sequence in the contour
x.mar   value to be subtracted/added to the smallest/largest observation in the x-axis
y.mar   value to be subtracted/added to the smallest/largest observation in the y-axis
col     color of the contour lines
...     further arguments related to contour and hist

Details

Provides the contour plot or density plot for the fitted data by Manly mixture model.

See Also

Manly.EM

Examples

set.seed(123)
K <- 2; p <- 2
X <- as.matrix(faithful)

# Obtain initial memberships based on the K-means algorithm
id.km <- kmeans(X, K)$cluster

# Run the EM algorithm for a Manly mixture model based on K-means solution
la <- matrix(0.1, K, p)
B <- Manly.EM(X, id.km, la)

Manly.plot(X, model = B, var1 = 1, x.mar = 1, y.mar = 2,
xaxs="i", yaxs="i", xaxt="n", yaxt="n", xlab="",
ylab = "", nlevels = 10, drawlabels = FALSE,
lwd = 3.2, col = "lightgrey", pch = 19)
Manly.select

Manly transformation selection

Description

Runs forward or backward model selection procedures for finding the optimal model in terms of BIC.

Usage

`Manly.select(X, model, method, tol = 1e-5, max.iter = 1000, silent = FALSE)`

Arguments

- `X` (dataset matrix (n x p))
- `model` (list containing parameters of the initial model)
- `method` (model selection method (options 'forward' and 'backward'))
- `tol` (tolerance level)
- `max.iter` (maximum number of iterations)
- `silent` (output control)

Details

Runs Manly forward and backward model selection procedures for a provided dataset. Forward and backward selection can be started from any ManlyMix object provided in 'model'. Manly transformation parameters are provided in matrix 'model$la’. If some transformations are not needed for specific components, zeros have to be specified in corresponding position. When all transformation parameters are set to zero, Manly mixture model degenerates to a Gaussian mixture model. Notation: n - sample size, p - dimensionality of the dataset X, K - number of mixture components.

Value

- `la` (matrix of the estimated transformation parameters (K x p))
- `tau` (vector of mixing proportions (length K))
- `Mu` (matrix of the estimated mean vectors (K x p))
- `S` (array of the estimated covariance matrices (p x p x K))
- `gamma` (matrix of posterior probabilities (n x K))
- `id` (estimated membership vector (length n))
- `ll` (log likelihood value)
- `bic` (Bayesian Information Criterion)
- `iter` (number of EM iterations run)
- `flag` (convergence flag (0 - success, 1 - failure))
**Manly.sim**

**Simulates Manly mixture dataset**

**Description**

Simulates Manly mixture dataset given the mixture parameters and sample size.

**Usage**

Manly.sim(n, la, tau, Mu, S)
Manly.var

Arguments

- **n**: sample size
- **la**: matrix of transformation parameters (K x p)
- **tau**: vector of mixing proportions (length K)
- **mu**: matrix of mean vectors (K x p)
- **S**: array of covariance matrices (p x p x K)

Details

Simulates a Manly mixture dataset. Manly mixture data points are computed from back-transforming Gaussian distributed data points using user-specified transformation parameters 'la'.

Value

- **X**: the simulated Manly mixture dataset
- **id**: the simulated membership of the data

Examples

```r
set.seed(123)

# sets the number of components, dimensionality and sample size
K <- 3
p <- 2
n <- 1000

# sets the parameters to simulate data from
tau <- c(0.25, 0.3, 0.45)
mu <- matrix(c(12, 4, 12, 4, 10), 3)
la <- matrix(c(1.2, 0.5, 1, 0.5, 0.5, 0.7), 3)
S <- array(NA, dim = c(p, p, K))
S[,,1] <- matrix(c(4, 0, 0, 4), 2)
S[,,2] <- matrix(c(5, -1, -1, 3), 2)
S[,,3] <- matrix(c(2, -1, -1, 2), 2)

# use function Manly.sim to simulate dataset with membership
A <- Manly.sim(n, la, tau, mu, S)

# plot the data
plot(A$x, col = A$id)
```

Description

Runs the variability assessment for a Manly mixture model.
usage

\texttt{Manly.var(X, model = NULL, conf.CI = NULL)}

arguments

\begin{itemize}
\item \texttt{X} \hspace{1cm} \text{dataset matrix (n x p)}
\item \texttt{model} \hspace{1cm} \text{Manly mixture model}
\item \texttt{conf.CI} \hspace{1cm} \text{confidence level, say 95 percent confidence}
\end{itemize}

details

Returns the estimated variance-covariance matrix and confidence intervals for model parameter estimates.

value

\begin{itemize}
\item \texttt{V} \hspace{1cm} \text{variance-covariance matrix.}
\item \texttt{CI} \hspace{1cm} \text{confidence intervals for each parameter.}
\end{itemize}

see also

\texttt{Manly.EM}

examples

\begin{verbatim}
set.seed(123)

# Use iris dataset
K <- 3; p <- 4
X <- as.matrix(iris[, -5])

# Use k-means clustering result
# all skewness parameters set to be 0.1 as the initialization of the EM algorithm
id.km <- kmeans(X, K)$cluster
la <- matrix(0.1, K, p)

# Run the EM algorithm with Manly mixture model
M.EM <- Manly.EM(X, id.km, la)

# Run the variability assessment
Manly.var(X, M.EM, conf.CI = 0.95)
\end{verbatim}
Wheat kernel Data

Description

The examined group comprised kernels belonging to three different varieties of wheat: Kama, Rosa and Canadian, 70 elements each, randomly selected for the experiment. High quality visualization of the internal kernel structure was detected using a soft X-ray technique. Studies were conducted using combine harvested wheat grain originating from experimental fields, explored at the Institute of Agrophysics of the Polish Academy of Sciences in Lublin.

Usage

data(seeds)

Format

A data frame with 210 observations on the following 7 variables.

- V1 Area A;
- V2 Perimeter P;
- V3 Compactness;
- V4 Length of kernel;
- V5 Width of kernel;
- V6 Asymmetry coefficient;
- V7 Length of kernel groove;
- V8 Seed species: 1, 2, 3

References


Examples

data(seeds)
Index

*Topic **CEM algorithm**
  Manly.Kmeans, 8

*Topic **Contour plot**
  Manly.plot, 12

*Topic **Data simulation**
  Manly.sim, 15

*Topic **EM algorithm**
  Manly.EM, 6
  Manly.model, 10
  Manly.plot, 10, 12
  Manly.select, 14
  Manly.sim, 15
  Manly.var, 16

*Topic **Forward and backward selection**
  Manly.EM, 6
  Manly.select, 14

*Topic **Manly mixture**
  Manly.EM, 6
  Manly.model, 10
  Manly.plot, 12
  Manly.select, 14
  Manly.sim, 15
  Manly.var, 16

*Topic **Manly transformation**
  Manly.Kmeans, 8

*Topic **Overlap**
  Manly.overlap, 11

*Topic **Variability assessment**
  Manly.var, 16

*Topic **datasets**
  acidity, 3
  ais, 4
  bankruptcy, 5
  seeds, 18

*Topic **k-means**
  Manly.Kmeans, 8

bankruptcy, 5
ClassAgree, 6
contour, 13
hist, 13
Manly.EM, 6
Manly.Kmeans, 8
Manly.model, 10
Manly.overlap, 11
Manly.plot, 10, 12
Manly.select, 14
Manly.sim, 15
Manly.var, 16
ManlyMix-package, 2
seeds, 18