Package ‘cbird’

February 6, 2017

Type Package

Title Clustering of Multivariate Binary Data with Dimension Reduction via L1-Regularized Likelihood Maximization

Version 1.0

Date 2017-02-06

Author Michio Yamamoto

Maintainer Michio Yamamoto <michio.koko@gmail.com>

Description The clustering of binary data with reducing the dimensionality (CLUSBIRD) proposed by Yamamoto and Hayashi (2015) <doi:10.1016/j.patcog.2015.05.026>.

License GPL (>= 2)

URL http://michioyamamoto.com

Repository CRAN

NeedsCompilation yes

Date/Publication 2017-02-06 17:38:13

R topics documented:

`cbird` *Clustering of multivariate binary data with dimension reduction via L1-regularized likelihood maximization.*

Description

This function conducts the clustering of binary data with reducing the dimensionality (CLUSBIRD) proposed by Yamamoto and Hayashi (2015).

Usage

```
cbird(Y, N.comp, N.clust, lambda=0, N.ite=10000, N.random=1,
     show.random.ite=FALSE, eps=0.0001, mc.cores=1)
```

Arguments

`Y` Binary data matrix (N * D), where N denotes sample size and D denotes the number of binary variables (0 or 1).

`N.comp` The number of component to be extracted.

`N.clust` The number of mixture components, which corresponds to the number of clusters.

`lambda` A tuning parameter of an L1 penalty for loadings. A non-negative real value should be used as the value of lambda.

`N.ite` The number of maximum of iterations for the EM algorithm.

`N.random` The number of random sets of parameters for initial random starts.

`show.random.ite`
If `"TRUE"`, the number of each iteration is shown on the R console.

`eps` The criterion for the convergence of the alternating least-squares algorithm, which should be specified as a positive real value. If the difference between the values of penalized log likelihood functions of successive iteration is smaller than `eps`, then `cbird` makes a decision about the convergence of the algorithm.

`mc.cores` If `"parallel"` package has been installed, `"cbird"` adopts a multithread process for multiple initial random starts. If `"mc.cores"=1`, `"parallel"` package is not needed, and a single core process is conducted.

Value

`F` An estimated component score matrix for cluster centroids.

`A` An estimated loading matrix.

`mu` Estimated mean values in the subspace.

`U` The cluster assignment matrix (N * `N.clust`).

`g` The estimated mixture probability.

`n.ite` The number of iteration needed for convergence.

`loss` The value of log likelihood with L1 penalty.

`bic` The value of BIC.

`LL` The value of log likelihood.

`cluster` Estimated clusters where subjects were assigned to

`ptime` Time for calculation

Author(s)

Michio Yamamoto
`<michio.koko@gmail.com>`

References

Yamamoto, M. and Hayashi, K. (2015). Clustering of multivariate binary data with dimension reduction via L1-regularized maximization. Pattern Recognition, 48, 3959-3968.

Examples

```
##Random Binary Data (unmeaningful example)
##100 subjects and 20 variables
##Consider three mixture components in the data
set.seed(1)
Y <- matrix(rbinom(100 * 20, 1, 0.5), 100, 20)
out <- cbird(Y, 2, 3)

est <- EstScore(Y, out$A, out$mu)
```

`EstScore` *Estimate compontent scores for each subject using the result of cbird.*

Description

This function estimates components scores for each subject using the result of CLUSBIRD.

Usage

```
EstScore(X, A, mu, N.ite=10000, N.random=1, show.random.ite=FALSE,
oblique=FALSE, mc.cores=1)
```

Arguments

`X` Binary data matrix (N * D).

`A` Loading matrix (D * L) estimated by `cbird`.

`mu` A D-length mean vector estimated by `cbird`.

`N.ite` The number of maximum of iterations for the EM algorithm.

`N.random` The number of random sets of parameters for initial random starts.

`show.random.ite`
If `"TRUE"`, the number of each iteration is shown on the R console.

`oblique` If `"TRUE"`, the oblique component scores F are estimated. The default is `"FALSE"`.

`mc.cores` If `"parallel"` package has been installed, `"EstScore"` adopts a multithread process for multiple initial random starts. If `"mc.cores"=1`, `"parallel"` package is not needed, and a single core process is conducted.

Value

`F` An estimated component score matrix (N * D) containing scores for subjects.

`n.ite` The number of iteration needed for convergence.

`loss` The value of loss function used in ALS algorithm

Author(s)

Michio Yamamoto
`<michio.koko@gmail.com>`

References

Yamamoto, M. and Hayashi, K. (2015). Clustering of multivariate binary data with dimension reduction via L1-regularized maximization. Pattern Recognition, 48, 3959-3968.

Examples

```
##See the example of the function "cbird".
```

Index