

Package ‘BicMix’

July 11, 2019

Type Package

Title BicMix

Version 0.0-0

Date 2017-10-20

Author Chuan Gao <chuan.gao.cornell@gmail.com>

Maintainer Chuan Gao <chuan.gao.cornell@gmail.com>

Description BicMix is a bayesian statistical method that decomposes a normally distributed matrix into the product of a sparse loading matrix, and a sparse factor matrix. The method employs a three parameter beta distribution that act as a spike and slab prior to induce sparsity to the two matrices. The method further consider the loadings and factors to be drawn from a mixture of sparse and dense component, which allows both the sparse components and the dense hidden confounding to be modeled. The method is especially suitable for deriving context informed functional gene modules, where the context is inferred from the factor matrix, and the sparse gene modules are found on the corresponding loadings. Please refer to our paper for more details (<http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004791>).

License GPL (>= 2)

Depends R (>= 2.14)

Collate BicMixWrapper.R

NeedsCompilation yes

RoxygenNote 6.1.1

R topics documented:

BicMixR	2
gen_BicMix_data	3
Index	5

BicMixR

An algorithm for decomposing a high dimensional matrix into the product of a sparse loading matrix, and a sparse factor matrix.

Description

An algorithm for decomposing a high dimensional matrix into the product of a sparse loading matrix, and a sparse factor matrix.

Usage

```
BicMixR(y = y, nf = 100, a = 0.5, b = 0.5, itr = 5001,
        rsd = NULL, out_itr = 500, out_dir = NULL, x_method = NULL,
        tol = NULL, qnorm = TRUE)
```

Arguments

y	matrix to be decomposed, no missing values are allowed
nf	the number of factors for the algorithm to start with, will be shrank to a smaller number reflecting the number of factors needed to explain the variance, default to 100
a	paramater one for the three parameter beta distribution, default to 0.5 to recapitulate horseshoe
b	paramater two for the three parameter beta distribution, default to 0.5 to recapitulate horseshoe
itr	The maximum number of iterations the algorithm is allowed to run, default to 5000
rsd	random seed for initializing the parameter values, default to be randomly drawn
out_itr	Iteration number out_itr, the algorithm will write temporary results into the specified directory (see below) every out_itr number of iterations. default to 500
out_dir	Directory where the algorithm will write temporary results into at the specified iteration number(see above)
x_method	whether induce sparsity on the X matrix, take values either "sparse" or "dense". default to "sparse"
tol	tolerance threshold for convergence, default to 1e-5
qnorm	whether to qq-normalize the gene expression matrix, default to TRUE

Value

lam: the sparse loading matrix

ex: the factor matrix

z: a vector indicating whether the corresponding loading is sparse (value of 1)

o: a vector indicating whether the corresponding factor is sparse (value of 1)

nf: the number of factors learned by the model

exx: the expected value of the covariance matrix, $E(XX^T)$

Author(s)

Chuan Gao <chuan.gao.cornell@gmail.com>

References

<http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004791>

Examples

```
library(BicMix)
## simulate data, the parameter std specifies the standard error of non-zero entries in the
## loading and factor matrices, where a normal distribution of mean zero
## is assumed for these values.
data = gen_BicMix_data(std=2)
## Visualize the loading matrix
image(t(data$lam),x=1:ncol(data$lam),y=1:nrow(data$lam),xlab="Loadings",ylab="Samples")
## Visualize the factor matrix
image(t(data$ex),x=1:ncol(data$ex),y=1:nrow(data$ex),xlab="Samples",ylab="Factors")
## run algorithm on the simulated data
system("mkdir results")
result = BicMixR(data$y,nf=100,a=0.5,b=0.5,itr=5000,out_dir="results",tol=1e-5,x_method="sparse",rsd=123)
## calculate a correlation matrix of the estimated loading matrix
## and the true loading matrix. Ideally, there should be one and
## only one big correlation value for a given row and column of the
## correlation matrix if the recovered sparse loadings and the true sparse loadings
cor.est.real = cor(result$lam[,result$z==1],data$lams)
## visualize the correlation matrix
image(cor.est.real,x=1:nrow(cor.est.real),y=1:ncol(cor.est.real),
xlab="Recovered loadings",ylab="True loadings")
```

gen_BicMix_data	<i>Simulate matrix with dimension of 500 x 20. Number of loadings and factors is set to 15, where 10 loadings and 10 factors are sparse. The sparse loadings and factors contain mostly zeros, and random blocks of nonzero values generated from $N(0,std)$. The dense loadings and factors are also generated from $N(0,std)$. The error matrix is generated from $N(0,1)$.</i>
-----------------	--

Description

Simulate matrix with dimension of 500 x 20. Number of loadings and factors is set to 15, where 10 loadings and 10 factors are sparse. The sparse loadings and factors contain mostly zeros, and random blocks of nonzero values generated from $N(0,std)$. The dense loadings and factors are also generated from $N(0,std)$. The error matrix is generated from $N(0,1)$.

Usage

```
gen_BicMix_data(std = 2)
```

Arguments

std	standard deviation for the normal distribution of the non-zero entries of the sparse components
-----	---

Value

a list containing the following

- lams: the sparse loadings
- lamd: the dense loadings
- lam: the loading matrix combining both the sparse and dense loading
- exs: the sparse factors matrix
- exd: the dense factors matrix
- ex: the factors matrix combining both the sparse and dense factors
- y: the y matrix calculated as $y = \text{lam} * \text{ex} + \text{err}$

Index

BicMixR, [2](#)

gen_BicMix_data, [3](#)