

Package ‘slfm’

March 15, 2020

Type Package

Title Fitting a Bayesian Sparse Latent Factor Model in Gene Expression Analysis

Version 1.0.1

Author Vinicius Mayrink [aut, cre],
Joao Duarte [aut]

Maintainer Vinicius Mayrink <vdinizm@gmail.com>

Depends R (>= 3.1.0)

Description Set of tools to find coherent patterns in gene expression (microarray) data using a Bayesian Sparse Latent Factor Model (SLFM) <DOI:10.1007/978-3-319-12454-4_15>. Considerable effort has been put to build a fast and memory efficient package, which makes this proposal an interesting and computationally convenient alternative to study patterns of gene expressions exhibited in matrices. The package contains the implementation of two versions of the model based on different mixture priors for the loadings: one relies on a degenerate component at zero and the other uses a small variance normal distribution for the spike part of the mixture.

Imports Rcpp (>= 0.11.0), coda, lattice

LinkingTo Rcpp, RcppArmadillo

Encoding UTF-8

License GPL-2

NeedsCompilation yes

RoxygenNote 7.1.0

Repository CRAN

Date/Publication 2020-03-15 22:20:02 UTC

R topics documented:

plot_matrix	2
process_matrix	3
slfm	3
slfm_list	5
slfm_package	6

plot_matrix	<i>plot_matrix</i>
-------------	--------------------

Description

Function to build a heat map displaying the values of given data matrix. This graph is useful for a visual inspection of the spatial distribution of the observations within the target matrix.

Usage

```
plot_matrix(  
  y,  
  standardize.rows = TRUE,  
  reorder.rows = TRUE,  
  reorder.cols = TRUE,  
  high.contrast = TRUE  
)
```

Arguments

<code>y</code>	data matrix to be evaluated.
<code>standardize.rows</code>	logical argument (default = TRUE) indicating whether to standardize the rows of <code>y</code> to build the image.
<code>reorder.rows</code>	logical argument (default = TRUE) indicating whether to reorder the rows of <code>y</code> to highlight a pattern.
<code>reorder.cols</code>	logical argument (default = TRUE) indicating whether to reorder the columns of <code>y</code> to highlight a pattern.
<code>high.contrast</code>	logical argument (default = TRUE) indicating whether to apply a transformation to increase contrast in the image of <code>y</code> .

See Also

[slfm](#), [process_matrix](#), [slfm_list](#)

process_matrix	<i>process_matrix</i>
----------------	-----------------------

Description

Function to apply a procedure to pre-process the gene expression data saved in matrices. This pre-processing step is required to allow a fair analysis of the results obtained from the SLFM model.

Usage

```
process_matrix(path, output_path, sample_size)
```

Arguments

path	path to the directory containing the set of matrices to be pre-processed.
output_path	path to the directory intended to accommodate the saved pre-processed matrices.
sample_size	number of matrices to be used on the principal component analysis.

See Also

[slfm](#), [slfm_list](#), [plot_matrix](#)

slfm	<i>slfm</i>
------	-------------

Description

Bayesian Sparse Latent Factor Model (SLFM) designed for the analysis of coherent patterns in gene expression data matrices. Details about the methodology being applied here can be found in *Duarte and Mayrink (2015)* and *Duarte and Mayrink (2019)*.

Usage

```
slfm(  
  x,  
  a = 2.1,  
  b = 1.1,  
  gamma_a = 1,  
  gamma_b = 1,  
  omega_0 = 0.01,  
  omega_1 = 10,  
  sample = 1000,  
  burnin = round(0.25 * sample),  
  lag = 1,  
  degenerate = FALSE  
)
```

Arguments

x	matrix with the pre-processed data.
a	positive shape parameter of the Inverse Gamma prior distribution (default = 2.1).
b	positive scale parameter of the Inverse Gamma prior distribution (default = 1.1).
gamma_a	positive 1st shape parameter of the Beta prior distribution (default = 1).
gamma_b	positive 2nd shape parameter of the Beta prior distribution (default = 1).
omega_0	prior variance of the spike mixture component (default = 0.01).
omega_1	prior variance of the slab mixture component (default = 10).
sample	sample size to be considered for inference after the burn in period (default = 1000).
burnin	size of the burn in period in the MCMC algorithm (default = sample/4).
lag	lag to build the chains based on spaced draws from the Gibbs sampler (default = 1).
degenerate	logical argument (default = FALSE) indicating whether to use the degenerate version of the mixture prior for the factor loadings.

Value

x: data matrix.
q_star: matrix of MCMC chains for q_star parameter.
alpha: summary table of MCMC chains for alpha parameter.
lambda: summary table of MCMC chains for lambda parameter.
sigma: summary table of MCMC chains for sigma parameter.
classification: classification of each alpha ('present', 'marginal', 'absent')

References

DOI:10.18637/jss.v090.i09 (Duarte and Mayrink; 2019)
DOI:10.1007/978-3-319-12454-4_15 (Duarte and Mayrink; 2015)

See Also

[process_matrix](#), [plot_matrix](#)

Examples

```
mat <- matrix(rnorm(2000), nrow = 20)
slfm(mat, sample = 1000)
```

slfm_list	<i>slfm_list</i>
-----------	------------------

Description

Function to fit the Bayesian Sparse Latent Factor Model to a group of data matrices within a directory. All matrices are supposed to have values representing the gene expression observed for different genes (rows) and different samples (columns).

Usage

```
slfm_list(
  path = ".",
  recursive = TRUE,
  a = 2.1,
  b = 1.1,
  gamma_a = 1,
  gamma_b = 1,
  omega_0 = 0.01,
  omega_1 = 10,
  sample = 1000,
  burnin = round(0.25 * sample),
  lag = 1,
  degenerate = FALSE
)
```

Arguments

path	path to the directory where the target data matrices are located.
recursive	logical argument (default = TRUE) indicating whether the function should look recursively inside folders.
a	positive shape parameter of the Inverse Gamma prior distribution (default = 2.1).
b	positive scale parameter of the Inverse Gamma prior distribution (default = 1.1).
gamma_a	positive 1st shape parameter of the Beta prior distribution (default = 1).
gamma_b	positive 2nd shape parameter of the Beta prior distribution (default = 1).
omega_0	prior variance of the spike mixture component (default = 0.01).
omega_1	prior variance of the slab mixture component (default = 10).
sample	sample size to be considered for inference after the burn in period (default = 1000).
burnin	size of the burn in period in the MCMC algorithm (default = sample/4).
lag	lag to build the chains based on spaced draws from the Gibbs sampler (default = 1).
degenerate	logical argument (default = FALSE) indicating whether to use the degenerate version of the mixture prior for the factor loadings.

See Also

[slfm](#), [process_matrix](#), [plot_matrix](#)

slfm_package

slfm: Fitting a Bayesian Sparse Latent Factor Model in Gene Expression Analysis

Description

Set of tools to find coherent patterns in gene expression (possibly microarray) data using a Bayesian Sparse Latent Factor Model (SLFM) <DOI:10.1007/978-3-319-12454-4_15> . Considerable effort has been put to build **slfm** fast and memory efficient, which makes this proposal an interesting and computationally convenient alternative to study patterns of gene expressions exhibited in matrices. The package contains the implementation of two versions of the model based on different mixture priors for the loadings: one relies on a degenerate component at zero and the other uses a small variance normal distribution for the spike part of the mixture. Additional functions are also available to handle data pre-processing procedures and to fit the model for a large number of probesets or genes. It includes functions to:

Details

* pre-process a set of matrices; * fit the available models to a set of matrices; * provide a detailed summarization of the model fit results.

Index

plot_matrix, [2](#), [3](#), [4](#), [6](#)
process_matrix, [2](#), [3](#), [4](#), [6](#)

slfm, [2](#), [3](#), [3](#), [6](#)
slfm_list, [2](#), [3](#), [5](#)
slfm_package, [6](#)