Package ‘SILGGM’

October 16, 2017

Type Package
Title Statistical Inference of Large-Scale Gaussian Graphical Model in Gene Networks
Version 1.0.0
Date 2017-10-15
Author Rong Zhang, Zhao Ren and Wei Chen
Maintainer Rong Zhang <roz16@pitt.edu>
Description Provides a general framework to perform statistical inference of each gene pair and global inference of whole-scale gene pairs in gene networks using the well known Gaussian graphical model (GGM) in a time-efficient manner. We focus on the high-dimensional settings where p (the number of genes) is allowed to be far larger than n (the number of subjects). Four main approaches are supported in this package: (1) the bivariate nodewise scaled Lasso (Ren et al (2015) <doi:10.1214/14-AOS1286>) (2) the de-sparsified nodewise scaled Lasso (Jankova and van de Geer (2017) <doi:10.1007/s11749-016-0503-5>) (3) the de-sparsified graphical Lasso (Jankova and van de Geer (2015) <doi:10.1214/15-EJS1031>) (4) the GGM estimation with false discovery rate control (FDR) using scaled Lasso or Lasso (Liu (2013) <doi:10.1214/13-AOS1169>). Windows users should install 'Rtools' before the installation of this package.
License GPL (>= 2)
Imports glasso, MASS, reshape, utils
Depends R (>= 3.0.0), Rcpp
LinkingTo Rcpp
NeedsCompilation yes
Repository CRAN
Date/Publication 2017-10-16 11:49:17 UTC

R topics documented:

SILGGM

Index
SILGGM

Statistical Inference of Large-Scale Gaussian Graphical Model in Gene Networks

Description

SILGGM is used to make statistical inference of conditional dependence among gene networks using the Gaussian graphical model (GGM). It includes four methods: (1) the bivariate nodewise scaled Lasso (B_NW_SL) (Ren et al., 2015) (2) the de-sparsified nodewise scaled Lasso (D-S_NW_SL) (Jankova and van de Geer, 2017) (3) the de-sparsified graphical Lasso (D-S.GL) (Jankova and van de Geer, 2015) and (4) the GGM estimation with false discovery rate control (FDR) using scaled Lasso or Lasso (GFC_SL or GFC_L) (Liu, 2013). This is an extensive and efficient package even for a high-dimensional setting.

Usage

SILGGM(x, method = NULL, lambda = NULL, global = FALSE, alpha = NULL, ndelta = NULL, true_graph = NULL, cytoscape_format = FALSE, csv_save = FALSE, directory = NULL)

Arguments

x
x is an n by p data matrix (n is the number of subjects and p is the number of genes, where p is allowed to be far larger than n).

method
Methods for statistical inference with 5 options: "B_NW_SL", "D-S_NW_SL", "D-S_GL", "GFC_SL" and "GFC_L". The default value is "D-S_NW_SL".

lambda
The value of a tuning parameter for a Lasso-type regularization approach. The default value is sqrt(2*log(p)/sqrt(n)) for method = "B_NW_SL", "D-S_NW_SL" or "GFC_SL" and sqrt(log(p)/n) for method = "D-S_GL". NOT applicable when method = "GFC_L".

global
If global = TRUE, the global inference of all gene pairs is performed. The default value is FALSE. ONLY applicable when method = "B_NW_SL", "D-S_NW_SL" or "D-S_GL".

alpha
A user-supplied sequence of pre-specified alpha levels for FDR control. The default is alpha = 0.05, 0.1 if no sequence is provided.

ndelta
The number of delta values decreased from 2 to 0 for selection of tuning parameters. The default value is 40. ONLY applicable when method = "GFC_L".

true_graph
The true graph structure in a study if available. The default value is NULL. This argument is particularly for global inference. If a true graph is available, both FDR(s) and the corresponding power(s) will be provided in the outputs. Otherwise, only FDR(s) and the associated threshold(s) for all absolute values of test statistics will be provided.

cytoscape_format
If cytoscape_format = TRUE, the outputs are shown in a table compatible with Cytoscape. The default value is FALSE.
csv_save

If `csv_save = TRUE`, the table in a Cytoscape format is saved to a directory as a `.csv` file. The file name is "Cytoscape_method.csv", where "method" depends on which method is used (e.g., the file name is "Cytoscape_D-S_NW_SL.csv" when method = "D-S_NW_SL"). The default value is FALSE.

directory

A user-specified directory to save the `.csv` files and ONLY applicable when `csv_save = TRUE`. If no directory is specified, the default value is NULL and a per-session temporary directory is generated in the program using the `tempdir()` function. However, the temporary directory and the saved files will be cleaned up after each R session ends. Therefore, a specified directory is HIGHLY recommended.

Details

In the original papers of the four methods, B_{NW\_SL}, D\_S-NW\_SL and D\_S-GL are developed for individual inference of each entry of a precision matrix, while GFC\_SL or GFC\_L is proposed particularly for simultaneous inference of all entries. However, GFC\_SL or GFC\_L essentially relies on p-values of all entries of a precision matrix, so implementations of the other three methods can also be extended to global inference under its FDR framework (Liu, 2013). Each method uses a Lasso-type regularization approach first, and then obtains an asymptotically efficient test statistic (e.g., z-score or a newly-constructed standardized test statistic) for each off-diagonal entry of a precision matrix under a certain sparseness condition. For individual inference of each gene pair, the package not only estimates the conditional dependence (each off-diagonal entry of a precision matrix) between each pair of genes but also provides the associated confidence interval, z-score and p-value. For global inference, it shows the FDR(s), the corresponding power(s) (if possible) and the decision(s) of the conditional dependence of each gene pair corresponding to the pre-specified alpha level(s) for FDR control. All of the outputs can be displayed in a table compatible with Cytoscape (Shannon et al., 2003), a popular and powerful software for network visualization. In addition, the table can be saved as a `.csv` file for a direct use in Cytoscape. The package performs each approach in a time-efficient manner and is able to accelerate the existing implementations to several orders of magnitudes without loss of accuracy.

Value

If `cytoscape_format = FALSE`, a list is returned including the following elements:

- `precision` A precision matrix including each gene pair. NOT applicable when method = "GFC\_SL" or "GFC\_L".
- `z_score_precision` A matrix of z-score for each off-diagonal entry of the precision matrix. NOT applicable when method = "GFC\_SL" or "GFC\_L".
- `p_precision` A matrix of p-value for each off-diagonal entry of the precision matrix. NOT applicable when method = "GFC\_SL" or "GFC\_L".
- `CI_low_precision` A matrix of lower value of 95% confidence interval for precision of the GGM. NOT applicable when method = "GFC\_SL" or "GFC\_L".
- `CI_high_precision` A matrix of higher value of 95% confidence interval for precision of the GGM. NOT applicable when method = "GFC\_SL" or "GFC\_L".
partialCor
A partial correlation matrix including each gene pair. NOT applicable when method = "GFC_SL" or "GFC_L".

z_score_partialCor
A matrix of z-score for each off-diagonal entry of the partial correlation matrix. ONLY applicable when method = "B_NW_SL".

p_partialCor
A matrix of p-value for each off-diagonal entry of the partial correlation matrix. ONLY applicable when method = "B_NW_SL".

CI_low_partialCor
A matrix of lower value of 95% confidence interval for partial correlation of the GGM. ONLY applicable when method = "B_NW_SL".

CI_high_partialCor
A matrix of higher value of 95% confidence interval for partial correlation of the GGM. ONLY applicable when method = "B_NW_SL".

t_stat
A matrix of newly-constructed standardized test statistic for each off-diagonal entry of the precision matrix. ONLY applicable when method = "GFC_SL" or "GFC_L".

FDR
The estimated FDR sequence for global inference of all off-diagonal entries of a precision matrix or all gene pairs based on the pre-specified alpha level(s).

threshold
The threshold sequence for absolute values of test statistics associated with the estimated FDR sequence.

power
The estimated power sequence for global inference of all off-diagonal entries of a precision matrix or all gene pairs associated with the estimated FDR sequence. ONLY applicable if true_graph is available.

global_decision
A list of p by p adjacency matrices of inferred graphs under the global inference corresponding to the sequence of pre-specified alpha levels. A value of 1 in the matrix means that there is conditional dependence (or an edge) between the gene pair, while a value of 0 means conditional independence (or no edge).

If cytoscape_format = TRUE, a list is returned including the following elements:

threshold
The threshold sequence for absolute values of test statistics associated with the estimated FDR sequence.

FDR
The estimated FDR sequence for global inference of all off-diagonal entries of a precision matrix or all gene pairs based on the pre-specified alpha level(s).

power
The estimated power sequence for global inference of all off-diagonal entries of a precision matrix or all gene pairs associated with the estimated FDR sequence. ONLY applicable if true_graph is available.

cytoscape_format_table
A table with Cytoscape format including all of the above possible outputs and can be saved to the directory shown in the directory argument as a .csv file.

Author(s)
Rong Zhang <roz16@pitt.edu>, Zhao Ren <zren@pitt.edu> and Wei Chen <wei.chen@chp.edu>
Maintainers: Rong Zhang <roz16@pitt.edu>
References


See Also

This package is based on the library *Rcpp*.

glasso in the package *glasso* is used when implementing the first step of `D_S_GL`.

Examples

```r
# Simulate a sparse precision matrix Omega
n <- 50
p <- 100
Omega.tmp <- matrix(0L, p, p)
diag(Omega.tmp) <- rep(1, p)
for(k in 1:(p/10)){
  i <- 10*(k-1)+1
  for(j in (10*(k-1)+2):((10*(k-1)+10)){
    Omega.tmp[i, j] <- 0.5
    Omega.tmp[j, i] <- 0.5
  }
}
eigenvalue <- eigen(Omega.tmp)$values
Omega <- Omega.tmp+(abs(min(eigenvalue))+0.05)*diag(p)
cov <- solve(Omega)

# Sample an n by p data matrix X based on it
library(MASS)
X <- mvrnorm(n, rep(0, p), cov)
```
# Run SILGGM
library(SILGGM)

# Use default method D-S_NW_SL without global inference
outlist1 <- SILGGM(x)

# Use method D-S_GL with global inference
# True graph is available
outlist2 <- SILGGM(x, method = "D-S_GL", global = TRUE, true_graph = Omega)

# Use method B_NW_SL without global inference
outlist3 <- SILGGM(x, method = "B_NW_SL")

# Use method GFC_SL or GFC_L
# True graph is available
outlist4 <- SILGGM(x, method = "GFC_SL", true_graph = Omega)
outlist5 <- SILGGM(x, method = "GFC_L", true_graph = Omega)

# Use method D-S_NW_SL with global inference
# True graph is available
# Show the outputs in a Cytoscape format
outlist6 <- SILGGM(x, method = "D-S_NW_SL", global = TRUE, true_graph = Omega, cytoscape_format = TRUE)

# Show the above outputs in a Cytoscape format table
# Save the table as a .csv file to a temporary directory
outlist7 <- SILGGM(x, method = "D-S_NW_SL", global = TRUE, true_graph = Omega, cytoscape_format = TRUE, csv_save = TRUE)
Index

glasso, 5
Rcpp, 5
SILGGM, 2