Package 'jewel'

September 19, 2022

Title Joint node-wise estimation of Gaussian graphical models from multiple datasets

Version 2.0.0

Author Anna Plaksienko, Claudia Angelini, Daniela De Canditiis

Maintainer Anna Plaksienko <anna@plaxienko.com>

Description jewel is a method to estimate networks of conditional dependencies (Gaussian graphical models) from multiple classes of data (similar but not exactly, i.e. measurements on different equipement, in different locations or for various subtypes). Package also allows to estimate regularization parameter with Bayesian information criterion and to generate simulation data. Reference paper to be added.

Depends R (>= 3.6.0)

Imports Matrix, matrixcalc, MASS, SMUT, igraph, parallel, purrr

URL https://github.com/annaplaksienko/jewel

License GPL-2

Encoding UTF-8

RoxygenNote 7.2.0

NeedsCompilation no

R topics documented:

`estimateLambdaBIC` *Estimate the optimal regularization parameter for jewel method with Bayesian information criterion*

Description

Function evaluates Baeysian information criterion (BIC) for each element of the grid. Optimal lambda is chosen as the one for which BIC's minimum is obtained. Warm start is implemented (matrices computed with the previous parameter are used as a starting point for the current element of the grid).

Usage

```
estimateLambdaBIC(X, lambda = NULL, tol = 1e-04, makePlot = TRUE)
```

Arguments

`X`	a list of `K` numeric data matrices of `n_k` samples and `p` variables (`n_k` can be different for each matrix).
`lambda`	an optional numeric vector of parameters for which the function evaluates BIC. Users are encouraged to tailor the grid to their specific data. If NULL, the default value is uniform in log sequence from 0.01 to 0.5.
`tol`	an optional number, convergence threshold controlling the relative error between iterations. The smaller it is, more precise is BIC. The default value is 0.0001.
`makePlot`	If makePlot = FALSE, plotting of BIC is disabled. The default value is TRUE.

Value

The following list is returned

- `lambda_opt` - a number, optimal value of regularization parameter according to BIC procedure;
- `BIC` - a numeric vector of BICs for each element of the grid of `lambda`.

`evaluatePerformance` *Evaluation of graph estimation method's performance if the true graph is known.*

Description

Function compares adjacency matrices of the true and estimated simple graphs and calculates the number of true positives (correctly estimated edges), true negatives (correctly estimated absence of edges), false positives (edges present in the estimator but not in the true graph) and false negatives (failure to identify an edge).

Usage

```
evaluatePerformance(G, G_hat)
```

Arguments

G	true graph's adjacency matrix.
G_hat	estimated graph's adjacency matrix. Must have the same dimensions as G.

Value

performance - a numeric vector of length 4 with TP, TN, FP, FN.

---

generateData_rewire *Generate a set of scale-free graphs and corresponding datasets (using the graphs as their Gaussian graphical models)*

Description

Function first generates K scale-free graphs with p vertices. They have the same order and degree distribution and share most of the edges, but some edges may vary (user can control how many). Function then generates corresponding precision and covariance matrices, all of the size p by p (see the paper for the details of the procedure). Then for each l-th element of vector n it generates K data matrices, each of the size n_l by p, i.e., for the same underlying graphs we can generate several sets of K datasets with different sample sizes.

Usage

```
generateData_rewire(
  K,
  p,
  n,
  power = 1,
  m = 1,
  perc = 0.05,
  int = NULL,
  makePlot = TRUE,
  verbose = TRUE
)
```

Arguments

K	number of graphs/data matrices.
p	number of nodes in the true graphs.
n	a numerical vector of the sample sizes for each desired set of K data matrices. Can be a vector of one element if the user wishes to obtain only one dataset of K matrices.

power	a number, power of preferential attachment for the Barabasi-Albert algorithm for the generation of the scale-free graph. Bigger number means more connected hubs. The default value is 1.
m	number of edges to add at each step of Barabasi-Albert algorithm for generation of the scale-free graph. Resulting graph has mp -(2m -1) edges. The default value is 1.
perc	a number, tuning parameter for the difference between graphs. Number of trials to perform in the rewiring procedure of the first graph is p * perc. Bigger the number, more different are the graphs.
int	a vector of two numbers, a and b. Entries of precision matrices are sampled from the uniform distribution on the interval [-b,-a] + [a,b]. The default values are a = 0.2,b = 0.8.
makePlot	If makePlot = FALSE, plotting of the generated graphs is disabled. The default value is TRUE.
verbose	If verbose = FALSE, tracing information printing is disabled. The default value is TRUE.

Value

The following list is returned

- Graphs – a list of adjacency matrices of the K generated graphs.
- CommomGraph - a matrix, common part (intersection) of the K generated graphs.
- Data - a list of lists, for each sample size of the input vector n one obtains K data matrices, each of the size n_1 by p.
- Sigma - a list of K covariance matrices of the size p by p.

---

jewel *Estimate Gaussian graphical models from multiple datasets*

Description

This function estimates Gaussian graphical models (i.e. networks of conditional dependencies, direct connections between variables) given several datasets. We assume that datasets contain measurements of the same variables collected under different conditions (different equipment, locations, even sub-types of disease).

Usage

```
jewel(
  X,
  lambda1,
  lambda2 = NULL,
  Theta = NULL,
  W = NULL,
```

```
  tol = 0.01,
  maxIter = 10000,
  verbose = TRUE
)
```

Arguments

X	a list of K numeric data matrices of n_k samples and p variables (n_k can be different for each matrix).
lambda1	a number, first regularization parameter (of the common penalty).
lambda2	an optional number, second regularization parameter (of the class-specific penalty). If NULL, set to lambda_2 = lambda_1 * 1.4
Theta	an optional list of K regression coefficient matrices of the size p by p. User-provided initialization can be used for warm-start procedures. If NULL, initialized as all zeros.
W	an optional list of K weights matrices of the size p by p. User-provided initialization can be used when some vertices are believed to be hubs. If NULL, initialized as all ones.
tol	an optional number, convergence threshold controlling the relative error between iterations. The default value is 0.01.
maxIter	an optional number, maximum allowed number of iterations. The default value is 10 000.
verbose	if verbose = FALSE, tracing information printing is disabled. The default value is TRUE.

Value

The following list is returned

- CommonG - an adjacency matrix of the common estimated graph (intersection of K estimated graphs).
- G_list - a list of K adjacency matrices for each estimated graph.
- Theta - a list of K estimated covariance matrices.
- BIC – a number, value of Bayesian information criterion for resulting graphs.