Package 'jewel'

January 7, 2021

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

Version 0.1.0

Author Anna Plaksienko, Claudia Angelini, Daniela De Canditiis

Maintainer Anna Plaksienko <anna@plaxienko.com>

Description jewel is a method for joint estimation of the graph of conditional dependencies between variables given multiple classes of data. Package also allows to estimate regularization parameter with Bayesian information criterion and cross-validation and to generate simulation data. Reference paper to be added.

Depends R (¿= 3.6.0)

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

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

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

NeedsCompilation no

R topics documented:

`estimateLambdaBIC` *Estimation of optimal lambda parameter for jewel method with Bayesian information criterion*

Description

Given a grid of lambda parameters, function evaluates Baeysian information criterion for each element. Optimal lambda is chosen as a BIC's minimum. Warm start is implemented.

Usage

```
estimateLambdaBIC(X, lambda)
```

Arguments

`X`	list of `K` data matrices of size `n_k` by `p` (`n_k` can be different for each class)
`lambda`	vector of grid lambda parameters for which function evaluates BIC

Value

The following list is returned

- lambda_opt - a number, optimal value of regularization parameter according to BIC procedure
- CV_err - a vector of cross-validation erros for each element of input vector `lambda`.

`estimateLambdaCV` *Estimation of optimal lambda parameter for jewel method based on cross-validation*

Description

Given a grid of lambda parameters, function performs cross-validation and finds the optimal one. Parallelization over folds and warm start are implemented.

Usage

```
estimateLambdaCV(X, lambda, k_folds = 5, verbose = TRUE, makePlot = TRUE)
```

Arguments

`X`	list of `K` data matrices of size `n_k` by `p` (`n_k` can be different for each class)
`lambda`	vector of grid lambda parameters over which cross-validation is performed
`k_folds`	number of folds in which data is divided. The default value is 5.
`verbose`	If verbose = FALSE, tracing information printing is disabled. The default value is TRUE.
`makePlot`	If makePlot = FALSE, plotting of CV error is disabled. The default value is TRUE.

Value

The following list is returned

- lambda_opt - a number, optimal value of regularization parameter according to cross-validation procedure
- CV_err - a vector of cross-validation errors for each element of input vector `lambda`

`evaluatePerformance` *Evaluation of performance of graph estimation methods if true graph is known*

Description

Function compares true and estimated simple graphs and calculates 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
`G_hat`	Estimated graph of the same size as `G`.

Value

vector of length 4 with TP, TN, FP, FN

`generateData` *Generate a "true" graph and corresponding data matrices*

Description

Function generates a scale-free graph with `p` vertices, `K` corresponding precision and covariance matrices, all of dize `p` by `p` and then for each `l`-th element of vector `n` generates `K` data matrices of size `n_l` by `p`. For the same underlying graph we can generate several datasets with different sample sizes.

Usage

```
generateData(
  K,
  n,
  p,
  power = 1,
  m = 1,
  c = 0.2,
  d = 0.8,
  makePlot = TRUE,
  verbose = TRUE
)
```

Arguments

K	Number of data matrices
n	Vector of the sample sizes for each desired set of data matrices. Can be a vector of one element.
p	Number of nodes in the true graph.
power	Power of preferential attachment for Barabasi-Albert algorithm for generation of scale-free graph.
m	Number of edges to add in each time step of Barabasi-Albert algorithm for generation of scale-free graph.
c	Entries in precision matrices are generated from the uniform distribution on the interval [-d,-c] + [c,d]. The default value is c = 0.2.
d	Entries in precision matrices are generated from the uniform distribution on the interval [-d,-c] + [c,d]. The default value is d = 0.8.
makePlot	If makePlot = FALSE, plotting of the generated true graph 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

- trueGraph - sparse adjacency matrix of the true graph
- data - list of lists, K data matrices of the size n_l by p for each sample size element l of the input vector n
- Sigma - list of K covariance matrices of the size p by p

---

generateLambdaGrid *Generate a sequence of lambda parameters*

Description

Function that generates a uniform in logarithmic space grid of tuning parameters. Lambda_max is max(1 / (n -1) max(X^TX)), lambda_min = lambda_max * eps.

Usage

```
generateLambdaGrid(X, n = 50, eps = 0.1, scale = TRUE)
```

Arguments

X	list of K data matrices of size n_k by p (n_k can be different for each class).
n	desired number of parameters. The default value is 50.
eps	lambda_min = lambda_max * eps. The default value is 0.1
scale	if TRUE and data is scaled, than resulting grid is independent of X and goes from eps to 1 uniformly in log-scale. The default value is TRUE.

Value

vector of regularization parameters of length n

`jewel` *Joint node-wise estimation of multiple Gaussian graphical models*

Description

Method for estimating the graph of conditional depndencies between the variables given multiple dataset (observations of variables under different conditiona or collected in distinct classes).

Usage

```
jewel(X, lambda, Theta = NULL, reltol = 0.01, maxIter = 10000, verbose = TRUE)
```

Arguments

`X`	List of `K` numeric data matrices of size `n_k` by `p` (`n_k` can be different for each class).
`lambda`	Tuning parameter which controls the sparsity of the resulting graph - bigger it is, less edges you get.
`Theta`	List of `K` starting regression coefficient matrices of size `p` by `p`. If not provided, initialized as all zeros.
`reltol`	Convergence threshold controlling relative error between iterations. The default value is 0.01.
`maxIter`	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

- EstAdjMat - adjacency matrix of the estimated graph
- Theta - list of `K` estimated covariance matrices of size `p` by `p`
- BIC - value of Bayesian information criterion
- residual - list of `K` resulting residuals X - X * Theta of size `n` by `p`