

Package ‘IRCcheck’

April 9, 2021

Title Irrepresentable Condition Check

Version 1.0.0

Date 2021-04-01

Description Check the irrepresentable condition (IRC) in both L1-regularized regression <doi:10.1109/TIT.2006.883611> and Gaussian graphical models. The IRC requires that the important and unimportant variables are not correlated, at least not all that much, and it is necessary for consistent model selection. Exploring the IRC as a function of the number of variables, assumed sparsity, and effect size can provide valuable insights into the model selection properties of L1-regularization.

License MIT + file LICENSE

Encoding UTF-8

Imports glmnet, MASS, Rdpack, GGMncv, corpcor, parallel

RdMacros Rdpack,

RoxygenNote 7.1.1

NeedsCompilation no

Author Donald Williams [aut, cre]

Maintainer Donald Williams <drwilliams@ucdavis.edu>

Repository CRAN

Date/Publication 2021-04-09 09:00:05 UTC

R topics documented:

IRCcheck-package	2
constrained	2
gen_net	3
irc_ggm	4
irc_regression	5
symm_mat	7

Index	9
--------------	----------

 IRCcheck-package

IRCcheck: Irrepresentable Condition Check

Description

L1 regularization requires the IRC for consistent model selection, that is, with more data, the true model is converged upon. This package allows for checking the IRC in both regression and Gaussian graphical models.

Importantly, the IRC cannot be checked in real data. The primary use for this package is to explore the IRC in a *\emph{true}* model that may be used in a simulation study. Alternatively, it is very informative to simply look at the IRC as a function of sparsity and the number of variables, including the regularization path and false selections.

 constrained

Constrained Precision Matrix

Description

Compute the maximum likelihood estimate, given certain elements are constrained to zero (e.g., an adjacency matrix). This approach is described in Hastie et al. (2009).

Usage

```
constrained(Sigma, adj)
```

Arguments

Sigma	Covariance matrix
adj	Matrix with constraints. A zero indicates that element should be constrained to zero.

Value

A list containing the inverse covariance matrix and the covariance matrix.

Note

The algorithm is written in c++.

References

Hastie T, Tibshirani R, Friedman J (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Examples

```
# random adj
# 90 % sparsity (roughly)
p <- 20
adj <- matrix(sample(0:1, size = p^2, replace = TRUE,
                    prob = c(0.9, 0.1) ),
              nrow = p, ncol = p)

adj <- symm_mat(adj)

diag(adj) <- 1

# random correlation matrix
set.seed(1)
cors <- cov2cor(
  solve(
    rWishart(1, p + 2, diag(p))[, , 1])
)

# constrain to zero
net <- constrained(cors, adj = adj)
```

gen_net

Generate True Partial Correlation Matrix

Description

Generate True Partial Correlation Matrix

Usage

```
gen_net(p = 20, edge_prob = 0.3, lb = 0.05, ub = 0.3)
```

Arguments

p	number of variables (nodes)
edge_prob	connectivity
lb	lower bound for the partial correlations
ub	upper bound for the partial correlations

Value

A list with the true structure, adjacency matrix, and correlation matrix.

Note

The function checks for a valid matrix (positive definite), but sometimes this will still fail. For example, for larger p , to have large partial correlations this requires a sparse GGM (accomplished by setting `edge_prob` to a small value).

Examples

```
true_net <- gen_net(p = 10)
```

`irc_ggm`*Irrepresentable Condition: Gaussian Graphical Model*

Description

Check the IRC (or Incoherence condition) in Gaussian graphical Models, following Equation (8) in (Ravikumar et al. 2008).

Usage

```
irc_ggm(true_network, cores = 2)
```

Arguments

`true_network` A matrix of dimensions p by p , assumed to be a partial correlation matrix.
`cores` Integer. Number of cores for parallel computing (defaults to 2)

Value

infinity norm (greater than 1 the IRC is violated, with closer to zero better).

References

Ravikumar P, Raskutti G, Wainwright MJ, Yu B (2008). “Model Selection in Gaussian Graphical Models: High-Dimensional Consistency of l_1 -regularized MLE.” In *NIPS*, 1329–1336.

Examples

```
# generate network
net <- gen_net(p = 20, edge_prob = 0.3, lb = 0.05, ub = 0.3)

# check irc
irc_ggm(net$pcors)

# random adj
# 90 % sparsity (roughly)
p <- 20
```

```

adj <- matrix(sample(0:1, size = p^2, replace = TRUE,
                    prob = c(0.9, 0.1) ),
              nrow = p, ncol = p)

adj <- symm_mat(adj)

diag(adj) <- 1

# random correlation matrix
set.seed(1)
cors <- cov2cor(
  solve(
    rWishart(1, p + 2, diag(p))[, , 1])
)

# constrain to zero
net <- constrained(cors, adj = adj)

irc_ggm(net$wadj)

#' # random adj
# 50 % sparsity (roughly)
p <- 20
adj <- matrix(sample(0:1, size = p^2, replace = TRUE, prob = c(0.5, 0.5) ),
              nrow = p, ncol = p)

adj <- symm_mat(adj)
diag(adj) <- 1

# random correlation matrix
set.seed(1)
cors <- cov2cor(
  solve(
    rWishart(1, p + 2, diag(p))[, , 1])
)

# constrain to zero
net <- constrained(cors, adj = adj)

irc_ggm(net$wadj)

```

irc_regression

Irrepresentable Condition: Regression

Description

Check the IRC in multiple regression, following Equation (2) in (Zhao and Yu 2006).

Usage

```
irc_regression(X, which_nonzero)
```

Arguments

`X` A matrix of dimensions n (observations) by p (variables).
`which_nonzero` Numeric vector with the location of the nonzero relations (a.k.a., the active set).

Value

infinity norm (greater than 1 the IRC is violated)

Note

It is common to take 1 - the infinity norm, thereby indicating the IRC is violated when the value is negative.

References

Zhao P, Yu B (2006). “On Model Selection Consistency of Lasso.” *The Journal of Machine Learning Research*, 7, 2541–2563. ISSN 15324435, doi: [10.1109/TIT.2006.883611](https://doi.org/10.1109/TIT.2006.883611), 1305.7477, <https://doi.org/10.1109/TIT.2006.883611>.

Examples

```
# data
# note: irc_met (block diagonal; 1st 10 active)
cors <- rbind(
  cbind(matrix(.7, 10,10), matrix(0, 10,10)),
  cbind(matrix(0, 10,10), matrix(0.7, 10,10))
)
diag(cors) <- 1

X <- MASS::mvrnorm(2500, rep(0, 20), Sigma = cors, empirical = TRUE)

# check IRC
irc_regression(X, which_nonzero = 1:10)

# generate data
y <- X %*% c(rep(1,10), rep(0, 10)) + rnorm(2500)

fit <- glmnet::glmnet(X, y, lambda = seq(10, 0.01, length.out = 400))

# plot
plot(fit, xvar = "lambda")

# Example (more or less) from Zhao and Yu (2006)
# section 3.3
```

```

# number of predictors
p <- 2^4

# number active (q in Zhao and Yu 2006)
n_beta <- 4/8 * p

# betas
beta <- c(rep(1, n_beta), rep(0, p - n_beta))

check <- NA
for(i in 1:100){
  cors <- cov2cor(
    solve(
      rWishart(1, p , diag(p))[, ,1]
    )
  )

  # predictors
  X <- MASS::mvrnorm(500, rep(0, p), Sigma = cors, empirical = TRUE)

  check[i] <- irc_regression(X, which_nonzero = which(beta != 0))
}

# less than 1
mean(check < 1)

# or greater than 0
mean(1 - check > 0)

```

symm_mat

Symmetric Matrix Copy the upper triangular of a matrix into the lower triangular portion of the matrix.

Description

Symmetric Matrix

Copy the upper triangular of a matrix into the lower triangular portion of the matrix.

Usage

```
symm_mat(x)
```

Arguments

x A matrix of dimensions p by p .

Value

A matrix

Examples

```
adj <- matrix(sample(0:1, size = 25, replace = TRUE), 5, 5)
symm_mat(adj)
```


Index

constrained, [2](#)

gen_net, [3](#)

irc_ggm, [4](#)

irc_regression, [5](#)

IRCcheck-package, [2](#)

symm_mat, [7](#)