Package ‘rIsing’

November 25, 2016

Type Package
Title High-Dimensional Ising Model Selection
Version 0.1.0
Description Fits an Ising model to a binary dataset using L1 regularized logistic regression and extended BIC. Also includes a fast lasso logistic regression function for high-dimensional problems. Uses the 'libLBFGS' optimization library by Naoaki Okazaki.
License GPL (>= 2)
Encoding UTF-8
LazyData true
Depends R (>= 3.1.0)
Imports Rcpp (>= 0.12.8), data.table (>= 1.9.6)
Suggests igraph, IsingSampler
LinkingTo Rcpp, RcppEigen (>= 0.3.2.9)
RoxygenNote 5.0.1
NeedsCompilation yes
Author Pratik Ramprasad [aut, cre], Jorge Nocedal [ctb, cph], Naoaki Okazaki [ctb, cph]
Maintainer Pratik Ramprasad <pratik.ramprasad@gmail.com>
Repository CRAN
Date/Publication 2016-11-25 08:43:07

R topics documented:

ising ................................................................. 2
logreg ............................................................... 3
rIsing ............................................................... 4

Index 5
ising

High-Dimensional Ising Model Selection

Description

Ising Model selection using L1-regularized logistic regression and extended BIC.

Usage

ising(x, gamma = 0.5, min_sd = 0, nlambda = 50,
lambda.min.ratio = 0.001, symmetrize = "mean")

Arguments

x The design matrix.

gamma (non-negative double) Parameter for the extended BIC (default 0.5). Higher gamma encourages sparsity. See references for more details.

min_sd (non-negative double) Columns of x with standard deviation less than this value will be excluded from the graph.

nlambda (positive integer) The number of parameters in the regularization path (default 50). A longer regularization path will likely yield more accurate results, but will take more time to run.

lambda.min.ratio (non-negative double) The ratio min(lambda) / max(lambda) (default 1e-3).

symmetrize The method used to symmetrize the output adjacency matrix. Must be one of "min", "max", "mean" (default), or FALSE. "min" and "max" correspond to the Wainwright min/max, respectively (see reference 1). "mean" corresponds to the coefficient-wise mean of the output adjacency matrix and its transpose. If FALSE, the output matrix is not symmetrized.

Value

A list containing the estimated adjacency matrix (Theta) and the optimal regularization parameter for each node (lambda), as selected by extended BIC.

References


logreg

Examples

```r
## Not run:
# simulate a dataset using IsingSampler
library(IsingSampler)
  n = 1e3
  p = 10
  Theta <- matrix(sample(c(-0.5,0.5), replace = TRUE, size = p*p), nrow = p, ncol = p)
  Theta <- Theta + t(Theta) # adjacency matrix must be symmetric
  diag(Theta) <- 0
  X <- unname(as.matrix(IsingSampler(n, graph = Theta, thresholds = 0, method = "direct")))
  m1 <- ising(X, symmetrize = "mean", gamma = 0.5, nlambda = 50)

# Visualize output using igraph
library(igraph)
  ig <- graph_from_adjacency_matrix(m1$Theta,  "undirected", weighted = TRUE, diag = FALSE)
  plot.igraph(ig, vertex.color = "skyblue")

## End(Not run)
```

logreg

L1 Regularized Logistic Regression

Description

L1 Regularized logistic regression using OWL-QN L-BFGS-B optimization.

Usage

```r
logreg(x, y, nlambda = 50, lambda.min.ratio = 0.001, lambda = NULL, 
    scale = TRUE, type = 2)
```

Arguments

- **x**: The design matrix.
- **y**: Vector of binary observations of length equal to nrow(x).
- **nlambda**: (positive integer) The number of parameters in the regularization path (default 50).
- **lambda.min.ratio**: (non-negative double) The ratio of max(lambda) / min(lambda) (default 1e-3).
- **lambda**: A user-supplied vector of regularization parameters. Under the default option (NULL), the function computes a regularization path using the input data.
- **scale**: (boolean) Whether to scale X before running the regression. The output parameters will always be rescaled. Use FALSE if X is already scaled.
- **type**: (integer 1 or 2) Type 1 aggregates the input data based on repeated rows in X. Type 2 (default) uses the data as is, and is generally faster. Use Type 1 if the data contains several repeated rows.
Value

A list containing the matrix of fitted weights (\(wmat\)), the vector of regularization parameters, sorted in decreasing order (\(\lambda\)), and the vector of log-likelihoods corresponding to \(\lambda\) (\(\logliks\)).

Examples

```r
# simulate some linear regression data
n <- 1e3
p <- 100
X <- matrix(rnorm(n*p), n,p)
wt <- sample(seq(0,9), p+1, replace = TRUE) / 10
z <- cbind(1,X) %*% wt + rnorm(n)
probs <- 1 / (1 + exp(-z))
y <- sapply(probs, function(p) rbinom(1,1,p))

m1 <- logreg(X, y)
m2 <- logreg(X, y, nlambda = 100, lambda.min.ratio = 1e-4, type = 1)
```

## Not run:

# Performance comparison
library(glmnet)
library(microbenchmark)
nlambda = 50; lambda.min.ratio = 1e-3
microbenchmark(
  logreg_type1 = logreg(X, y, nlambda = nlambda,
                        lambda.min.ratio = lambda.min.ratio, type = 1),
  logreg_type2 = logreg(X, y, nlambda = nlambda,
                        lambda.min.ratio = lambda.min.ratio, type = 2),
  glmnet = glmnet(X, y, family = "binomial",
                  nlambda = nlambda, lambda.min.ratio = lambda.min.ratio),
  times = 20L
)
## End(Not run)

rIsing: High-Dimensional Ising Model Selection.

Description

Fits an Ising model to a binary dataset using L1-regularized logistic regression and BIC. Also includes a fast lasso logistic regression function for high-dimensional problems. Uses the 'libLBFGS' optimization library by Naoki Okazaki.

rIsing functions

- `logreg`: L1-regularized logistic regression using OWL-QN L-BFGS-B optimization.
- `ising`: Ising Model selection using L1-regularized logistic regression and extended BIC.
Index

ising, 2
logreg, 3
rIsing, 4
rising-package (rIsing), 4