Package 'LqG'

April 27, 2022

Type Package

Title Robust Group Variable Screening Based on Maximum Lq-Likelihood Estimation

Version 0.1.0

Author Mingcong Wu, Yang Li, Rong Li

Maintainer Rong Li <rong_li@ruc.edu.cn>

Description Produces a group screening procedure that is based on maximum Lq-likelihood estimation, to simultaneously account for the group structure and data contamination in variable screening. The methods are described in Li, Y., Li, R., Qin, Y., Lin, C., & Yang, Y. (2021) Robust Group Variable Screening Based on Maximum Lq-likelihood Estimation. Statistics in Medicine, 40:6818-6834.<doi:10.1002/sim.9212>.

License GPL-3

Encoding UTF-8

LazyData true

Imports base

Depends R (>= 3.5.0)

NeedsCompilation no

Repository CRAN

Date/Publication 2022-04-27 08:20:09 UTC

R topics documented:

`grsc.marg.MLqE` *Group Screening based on marginal Maximum Lq-likelihood Estimation*

Description

Group screening by ranking utility of each group. The group effect is defined based on the cumulation of the maximum Lq-likelihood estimate of the regression using only one predictor each time within the group.

Usage

```
grsc.marg.MLqE(
X,
Y,
n = dim(X)[1],
p = dim(X)[2],
q = 0.9,
m,
group,
eps = 1e-06,
d = n/log(n)
)
```

Arguments

`X`	A matrix of predictors.
`Y`	A vector of response.
`n`	A value of sample size
`p`	A value denoting the dimension of predictors
`q`	A value of distortion parameter of Lq function, default to `0.9`.
`m`	A number of the predictor groups
`group`	A vector of consecutive integers describing the grouping of the coefficients (see example below).
`eps`	The iteration coverage criterion, default to `1e-06`.
`d`	A value of the number of groups retained after screening, default to `n/log(n)`.

Details

grsc.marg.MLqE obtains the group effect of each group for subsequential group screening, based on the cumulative marginal MLqE coefficients within the group. It can work when both the correlation within groups and between groups are small. If group size equals to 1, individual screening is conducted.

Value

The grsc.marg.MLqE returns a list containing the following components:

beta.group	The vector of utility of each group, which is the criterion for the variable screening procedure.
group.screened	The vector of integers denoting the screened groups.

Examples

```
# This is an example of grsc.marg.MLqE with simulated data
data(LqG_SimuData)
X = LqG_SimuData$X
Y = LqG_SimuData$Y
n = dim(X)[1]
p = dim(X)[2]
m = 200
groups = rep(1:(p/5), each = 5)
result <- grsc.marg.MLqE(X = X,
                         Y = Y,
                         n = n,
                         p = p,
                         q = 0.9,
                         m = m,
                         group = groups,
                         eps = 1e-06,
                         d = 15)
result$beta.group
result$group.screened
```

---

grsc.MLqE *Group Screening based on Maximum Lq-likelihood Estimation*

Description

Group screening by ranking utility of each group. The group effect is defined based on the maximum Lq-likelihood estimates of the regression using each group of variables.

Usage

```
grsc.MLqE(
X,
Y,
n = dim(X)[1],
q = 0.9,
m,
group,
eps = 1e-06,
d = n/log(n)
)
```

Arguments

X	A matrix of predictors.
Y	A vector of response.
n	A value of sample size
q	A value of distortion parameter of Lq function, default to 0.9.
m	A number of the predictor groups
group	A vector of consecutive integers describing the grouping of the coefficients (see example below).
eps	The iteration coverage criterion, default to 1e-06.
d	A value of the number of groups retained after screening, default to n/log(n).

Details

grsc.MLqE obtains the group effect of each group for subsequential group screening, based on the maximum Lq-likelihood estimates of the regression using each group of variables. By inheriting the advantage of the MLqE in small or moderate sample situations, the method is more robust to heterogeneous data and heavy-tailed distributions. It can work when correlation is mild or large. If group size equals to 1, individual screening is conducted.

Value

The grsc.MLqE returns a list containing the following components:

beta.group	The vector of utility of each group, which is the criterion for the variable screening procedure.
group.screened	The vector of integers denoting the screened groups.

Examples

```
# This is an example of grsc.MLqE with simulated data
data(LqG_SimuData)
X = LqG_SimuData$X
Y = LqG_SimuData$Y
n = dim(X)[1]
m = 200
groups = rep(1:( dim(X)[2] / 5), each = 5)
result <- grsc.MLqE(X = X,
                    Y = Y,
                    n = n,
                    q = 0.9,
                    m = m,
                    group = groups,
                    eps = 1e-06,
                    d = 15)
result$beta.group
result$group.screened
```

LqG_SimuData *An Example of Simulated Data for LqG*

Description

The dataset LqG_SimuData contains n = 100 samples with p = 1000 predictors. The number of the groups m = 200.

Usage

```
LqG_SimuData
```

Format

A data list containing 100 samples

MLqE.est *Maximum Lq-likelihood Estimation*

Description

The iterative algorithm for MLqE of coefficients of regression using each group of variables.

Usage

```
MLqE.est(
X,
Y,
q = 0.9,
eps = 1e-06
)
```

Arguments

X	The matrix of the predictor group.
Y	The vector of response.
q	The value of distortion parameter of Lq function, default to 0.9.
eps	The iteration coverage criterion, default to 1e-06.

Details

The estimating equation of MLqE is a weighted version of that of the classical maximum likelihood estimation (MLE) where the distortion parameter q determines the similarity between the Lq function and the log function. When q = 1, MLqE is equivalent to MLE. The closer q is to 1, the more sensitive the MLqE is to outliers. As for the selection of q, there is presently no general method. However, MLqE is generally less sensitive to data contamination than MLE (to different degrees) when q is smaller than 1. Here, the default value of q is 0.9. Distortion parameter q can also be determined according to sample size n, choices of q_n with $|1 - q_n|$ between $\frac{1}{n}$ and $\frac{1}{\sqrt{n}}$ usually improves over the MLE.

Value

The `MLqE.est` returns a list containing the following components:

`t`	The integer specifying the number of the total iterations in the algorithm.
`beta_hat`	The vector of estimated coefficients.
`sigma_hat`	The value of the estimated variance.
`OMEGA_hat`	The matrix of the estimated weight.

Examples

```
# This is an example of grsc.marg.MLqE with simulated data
data(LqG_SimuData)
X = LqG_SimuData$X
Y = LqG_SimuData$Y
n = dim(X)[1]
p = dim(X)[2]
m = 200
groups = rep(1:( dim(X)[2] / 5), each = 5)
Xb = X[ , which( groups == 1)]
result = MLqE.est(Xb,
                  Y,
                  q = 0.9,
                  eps = 1e-06)
result$beta_hat
result$sigma_hat
result$OMEGA_hat
result$t
```

Index