Package 'LassoGEE'

April 16, 2020

Type Package

Title Generalized Estimating Equations with L_1 regularization in High-Dimension

Version 0.1.0

Author Yaguang Li, Xin Gao, Wei Xu

Maintainer Yaguang Li <liygcr7@gmail.com>

Description
Fits generalized estimating equations with L_1 regularization to longitudinal data with high dimensional covariates. Use a efficient iterative composite gradient descent algorithm (I-CGD).

License GPL (>= 2)

Depends R (>= 3.6.0)

Encoding UTF-8

LazyData true

Imports Rcpp (>= 1.0.4), RcppArmadillo, PGEE, MASS, mvtnorm

LinkingTo Rcpp, RcppArmadillo

RoxygenNote 7.1.0

R topics documented:

`LassoGEE` *Function to fit penalized GEE by I-CGD algorithm.*

Description

This function fits a L_1 penalized GEE model to longitudinal data by I-CGD algorithm.

Usage

```
LassoGEE(
  X,
  y,
  id,
  family = binomial("probit"),
  lambda,
  corstr = "independence",
  beta.ini = NULL,
  R = NULL,
  scale.fix = TRUE,
  scale.value = 1
)
```

Arguments

X	A design matrix of dimension (nm) * p.
y	A response vector of length m * n.
id	A vector for identifying subjects/clusters.
family	A family object: a list of functions and expressions for defining link and variance functions. Families supported here is same as in **PGEE** which are binomial, gaussian, gamma and poisson.
lambda	A numerical value for the penalization parameter.
corstr	A character string, which specifies the type of correlation structure. Structures supported in PGEE are "AR-1","exchangeable", "independence", and "unstructured". The default corstr type is "independence".
beta.ini	User specified initial values for regression parameters. The default value is NULL.
R	User specified correlation matrix. The default value is NULL.
scale.fix	A logical variable; if true, the scale parameter is fixed at the value of scale.value. The default value is TRUE.
scale.value	If scale.fix = TRUE, this assignes a numeric value to which the scale parameter should be fixed. The default value is 1.

Value

A list containing the following components:

betaest	return final estimation
beta_all_step	return estimate in each iteration
inner.count	iterative count in each stage
outer.iter	iterate number of outer loop

References

Li, Y., Gao, X., and Xu, W. (2020). Statistical consistency for generalized estimating equation with L_1 regularization.

Examples

```
## Not run:
set.seed(123)
p <- 256
s <- ceiling(p^{1/3})
n <- ceiling(10 * s * log(p))
m <- 4
# covariance matrix of p number of continuous covariates
X.sigma <- matrix(0, p, p)
{
  for (i in 1:p)
    X.sigma[i,] <- 0.5^(abs((1:p)-i))
}

# generate matrix of covariates
X <- as.matrix(rmvnorm(n*m, mean = rep(0,p), X.sigma))

# true regression parameter associated with the covariate
bt <- runif(s, 0.05, 0.5) # = rep(1/s,s)
beta.true <- c(bt,rep(0,p-s))
# intercept
beta_intercepts <- 0
# unstructure
tt <- runif(m*m,-1,1)
Rtmp <- t(matrix(tt, m,m))%*%matrix(tt, m,m)+diag(1,4)
R_tr <- diag(diag(Rtmp)^{-1/2})%*%Rtmp%*%diag(diag(Rtmp)^{-1/2})
diag(R_tr) = round(diag(R_tr))

# library(SimCorMultRes)
# simulation of clustered binary responses
simulated_binary_dataset <- rbin(clsize = m, intercepts = beta_intercepts,
                                 betas = beta.true, xformula = ~X, cor.matrix = R_tr,
                                 link = "probit")
lambda <- 0.2* s *sqrt(log(p)/n)
data = simulated_binary_dataset$simdata
y = data$y
X = data$X
id = data$id

ptm <- proc.time()
nCGDfit = LassoGEE(X = X, y = y, id = id, family = binomial("probit"),
                lambda = lambda, corstr = "unstructured")
proc.time() - ptm
betaest <- nCGDfit$betaest

## End(Not run)
```

Index