Package `LassoGEE`

November 6, 2020

Type Package

Title High-Dimensional Lasso Generalized Estimating Equations

Version 1.0

Author Yaguang Li, Xin Gao, Wei Xu

Maintainer Yaguang Li <liygcr7@gmail.com>

Description

Fits generalized estimating equations with L1 regularization to longitudinal data with high dimensional covariates. Use a efficient iterative composite gradient descent algorithm.

License GPL (>= 2)

URL <https://github.com/liygCR/LassoGEE>

Depends R (>= 3.6.0)

Encoding UTF-8

LazyData true

Imports Rcpp (>= 1.0.4), PGEE, MASS, mvtnorm, caret, SimCorMultRes

LinkingTo Rcpp, RcppArmadillo

RoxygenNote 7.1.1

NeedsCompilation yes

Repository CRAN

Date/Publication 2020-11-06 12:20:08 UTC

R topics documented:

cv.LassoGEE ......................................................... 2
IC ................................................................. 3
LassoGEE ......................................................... 4
print.cv.LassoGEE .................................................. 6
print.LassoGEE .................................................... 7

Index 9
cv.LassoGEE

Cross-validation for LassoGEE.

Description

Does k-fold cross-validation for LassoGEE to select tuning parameter value for longitudinal data with working independence structure.

Usage

cv.LassoGEE(
  X,
  y,
  id,
  family,
  method = c("CGD", "RWL"),
  scale.fix,
  scale.value,
  fold,
  lambda.vec,
  maxiter,
  tol
)

Arguments

- **X**: A design matrix of dimension \((nm) \times p\).
- **y**: A response vector of length \(m \times n\).
- **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.
- **method**: The algorithms that are available. "CGD" represents the I-CGD algorithm, and "RWL" represents re-weighted least square algorithm.
- **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 assigns a numeric value to which the scale parameter should be fixed. The default value is 1.
- **fold**: The number of folds used in cross-validation.
- **lambda.vec**: A vector of tuning parameters that will be used in the cross-validation.
- **maxiter**: The number of iterations that is used in the estimation algorithm. The default value is 50.
- **tol**: The tolerance level that is used in the estimation algorithm. The default value is 1e^-3.
IC

Value
An object class of cv.LassoGEE.

References

See Also
LassoGEE

Description
Information Criterion for a fitted LassoGEE object with the AIC, BIC, or GCV criteria.

Usage
IC(obj, criterion = c("BIC", "AIC", "GCV", "AICc", "EBIC"))

Arguments
obj A fitted LassoGEE object.
criterion The criterion by which to select the regularization parameter. One of "AIC", "BIC", "GCV", "AICc", or "EBIC"; default is "BIC".

Value
IC The calculated model selection criteria

References
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 or re-weighted least square algorithm.

Usage
LassoGEE(
  X,
  y,
  id,
  family = binomial("probit"),
  lambda,
  corstr = "independence",
  method = c("CGD", "RWL"),
  beta.ini = NULL,
  R = NULL,
  scale.fix = TRUE,
  scale.value = 1,
  maxiter = 50,
  tol = 0.001,
  silent = TRUE,
  Mv = NULL,
  verbose = TRUE
)

Arguments
X A design matrix of dimension $(nm) \times p$.
y A response vector of length $m \times n$.
id A vector for identifying subjects/clusters.
family A family object representing one of the built-in families. Families supported here are the same as in PGEE, e.g., binomial, gaussian, gamma and poisson, and the corresponding link functions are supported, e.g., identity, and probit.
lambda A user supplied value for the penalization parameter.
corstr A character string that indicates the correlation structure among the repeated measurements of a subject. Structures supported in LassoGEE are "AR1", "exchangeable", "unstructured", and "independence". The default corstr type is "independence".
method The algorithms that are available. "CGD" represents the I-CGD algorithm, and "RWL" represents re-weighted least square algorithm.
beta.ini User specified initial values for regression parameters. The default value is NULL.
R
scale.fix A logical variable. The default value is TRUE, then the value of the scale parameter is fixed to scale.value.
scale.value If scale.fix = TRUE, a numeric value will be assigned to the fixed scale parameter. The default value is 1.
maxiter The maximum number of iterations used in the algorithm. The default value is 50.
tol The tolerance level used in the algorithm. The default value is 1e-3.
silent A logical variable; if false, the iteration counts at each iteration of CGD are printed. The default value is TRUE.
Mv If either "stat_M_dep", or "non_stat_M_dep" is specified in corstr, then this assigns a numeric value for Mv. Otherwise, the default value is NULL.
verbose A logical variable; Print the out loop iteration counts. The default value is TRUE.

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

See Also
cv.LassoGEE

Examples
# required R package
library(mvtnorm)
library(SimCorMultRes)
#
set.seed(123)
p <- 200
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})%^%diag(diag(Rtmp)^{-1/2})

# 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

---

print.cv.LassoGEE  

### print a cross-validated LassoGEE object

**Description**

Print a summary of the results of cross-validation for a LassoGEE model.

**Usage**

```r
## S3 method for class 'cv.LassoGEE'
print(x, digits = NULL, ...)
```
Arguments

x  fitted 'cv.LassoGEE' object
digits  significant digits in printout
...  additional print arguments

Details

A summary of the cross-validated fit is produced. print.cv.LassoGEE(object) will print the summary for a sequence of lambda.

References


See Also

LassoGEE, and cv.LassoGEE methods.
See Also

LassoGEE, and cv.LassoGEE methods.
Index

* models
  print.cv.LassoGEE, 6
  print.LassoGEE, 7
* regression
  print.cv.LassoGEE, 6
  print.LassoGEE, 7

cv.LassoGEE, 2
IC, 3
LassoGEE, 4

print.cv.LassoGEE, 6
print.LassoGEE, 7