Package 'crrp'

June 20, 2015

Type Package

Title Penalized Variable Selection in Competing Risks Regression

Version 1.0

Date 2015-06-19

Author Zhixuan Fu

Maintainer Zhixuan Fu <zhixuan.fu@yale.edu>

Depends survival, Matrix, cmprsk

Description In competing risks regression, the proportional subdistribution hazards(PSH) model is popular for its direct assessment of covariate effects on the cumulative incidence function. This package allows for penalized variable selection for the PSH model. Penalties include LASSO, SCAD, MCP, and their group versions.

License GPL (>= 2)

NeedsCompilation yes

Repository CRAN

Date/Publication 2015-06-20 00:56:59

R topics documented:

`crrp` *Penalized variable selection at the individual level in competing risks regression*

Description

Extends R package **ncvreg** to the proportional subdistribution hazards model. Penalties include LASSO, SCAD, and MCP. User-specified weights can be assigned to the penalty for each coefficient.

Usage

```
crrp(time, fstatus, X, failcode = 1, cencode = 0,
penalty = c("MCP", "SCAD", "LASSO"), gamma = switch(penalty, SCAD = 3.7, 2.7),
alpha = 1, lambda.min = 0.001, nlambda = 50, lambda, eps = 0.001,
max.iter = 1000, penalty.factor = rep(1, ncol(X)), weighted = FALSE)
```

Arguments

time	vector of failure/censoring times
fstatus	vector with a unique code for each failure type and a separate code for censored observations
X	design matrix; crrp standardizes X by default
failcode	code of fstatus that denotes the failure type of interest
cencode	code of fstatus that denotes censored observations
penalty	penalty to be applied to the model. Either "LASSO", "SCAD", or "MCP"
gamma	tuning parameter of the MCP/SCAD penalty. Default is 2.7 for MCP and 3.7 for SCAD
alpha	tuning parameter indicating contributions from the MCP/SCAD penalty and the L2 penalty. alpha=1 is equivalent to MCP/SCAD penalty, whereas alpha=0 would be equivalent to ridge regression. Default is 1
lambda.min	the smallest value for lambda. Default is .001
nlambda	number of lambda values. Default is 50
lambda	a user-specified sequence of lambda values. If not specified, a sequence of values of length nlambda is provided
eps	iteration stops when the relative change in any coefficient is less than eps. Default is 0.001
max.iter	maximum number of iterations. Default is 1000
penalty.factor	a vector of weights applied to the penalty for each coefficient. The length of the vector must be equal to the number of columns of X
weighted	if TRUE, weights must be provided by users. Default is FALSE

Details

The crrp function penalizes the partial likelihood of the proportional subdistribution hazards model from Fine and Gray(1999) with penalty LASSO, SCAD, and MCP. The coordinate algorithm is used for implementation. The criteria BIC and GCV are used to select the optimal tuning parameter.

Value

Return a list of class crrp with components

$beta	fitted coefficients matrix with nvars row and nlambda columns
$iter	number of iterations until convergence for each lambda
$lambda	sequence of tuning parameter values

$penalty	same as above
$gamma	same as above
$alpha	same as above
$loglik	log likelihood of the fitted model at each value of lambda
$GCV	generalized cross validation of the fitted model at each value of lambda
$BIC	Bayesian information criteria of the fitted model at each value of lambda
$SE	matrix of standard errors with nvars row and nlambda columns

Author(s)

Zhixuan Fu <zhixuan.fu@yale.edu>

References

- Breheny, P. and Huang, J. (2011) Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Ann. Appl. Statist.*, 5: 232-253.
- Fine J. and Gray R. (1999) A proportional hazards model for the subdistribution of a competing risk. *JASA* 94:496-509.
- Fu Z., Parikh C. and Zhou B.(2015). Penalized variable selection in competing risks regression. Manuscript submitted for publication.

See Also

gcrrp, cmprsk, ncvreg

Examples

```
#simulate competing risks data
set.seed(10)
ftime <- rexp(200)
fstatus <- sample(0:2,200,replace=TRUE)
cov <- matrix(runif(1000),nrow=200)
dimnames(cov)[[2]] <- c('x1','x2','x3','x4','x5')

#fit LASSO
fit <- crrp(ftime, fstatus, cov, penalty="LASSO")
#use BIC to select tuning parameters
beta <- fit$beta[, which.min(fit$BIC)]
beta.se <- fit$SE[, which.min(fit$BIC)]

#fit adaptive LASSO
weight <- 1/abs(crr(ftime, fstatus, cov)$coef)
fit2 <-crrp(ftime, fstatus, cov, penalty="LASSO", penalty.factor=weight, weighted=TRUE)
beta2 <- fit2$beta[, which.min(fit2$BIC)]
beta2.se <- fit2$SE[, which.min(fit2$BIC)]
```

gcrrp *Group penalized variable selection in competing risks regression*

Description

Extends R package **grpreg** to the proportional subdistribution hazards (PSH) model (Fine and Gray, 1999). Performs penalized variable selection at the group level. Penalties include group LASSO, adaptive group LASSO, group SCAD, and group MCP.

Usage

```
gcrrp(time, fstatus, X, failcode = 1, cencode = 0, group=1:ncol(X),
    penalty=c("gLASSO", "gMCP", "gSCAD"),gamma=switch(penalty, SCAD=3.7, 2.7),
      alpha=1, lambda.min=0.001, nlambda=50, lambda, eps=.001,
      max.iter=1000, weighted=FALSE)
```

Arguments

time	vector of failure/censoring times
fstatus	vector with a unique code for each failure type and a separate code for censored observations
X	design matrix; crrp standardizes and orthogonizes X by default
failcode	code of fstatus that denotes the failure type of interest
cencode	code of fstatus that denotes censored observations
group	vector of group indicator (see details)
penalty	penalty to be applied to the model. Either "gLASSO", "gSCAD", or "gMCP"
gamma	tuning parameter of the gMCP/gSCAD penalty. Default is 2.7 for group MCP and 3.7 for group SCAD.
alpha	tuning parameter indicating contributions from the MCP/SCAD penalty and the L2 penalty.
lambda.min	the smallest value for lambda. Default is .001
nlambda	number of lambda values. Default is 50
lambda	a user-specified sequence of lambda values. If not specified, a sequence of values of length nlambda is provided
eps	iteration stops when the relative change in any coefficient is less than eps. Default is 0.001
max.iter	maximum number of iterations. Default is 1000
weighted	Default is FALSE. If TRUE, it must be used with gLASSO to produce adaptive group LASSO penalty(see details)

Details

The group vector indicates the grouping of variables. For greatest efficiency, group should be a vector of consecutive integers, although unordered groups are also allowed.

Penalties include group LASSO, group SCAD, and group MCP. We also include adaptive group LASSO by putting weighted=TRUE. The gcrrp function calculates data-adaptive weights formulated by the maximum parital likelihood estimator(MPLE) of the PSH model. The weight for each group is the inverse of the norm of the corresponding sub-vector of MPLE. The algorithm employed is the group coordinate descent algorithm.

Value

Return a list of class gcrrp with components

$beta	fitted coefficients matrix with nvars row and nlambda columns
$iter	number of iterations until convergence for each lambda
$group	same as above
$lambda	sequence of tuning parameter values
$penalty	same as above.
$gamma	same as above.
$alpha	same as above.
$loglik	log likelihood of the fitted model at each value of lambda
$GCV	generalized cross validation of the fitted model at each value of lambda
$BIC	Bayesian information criteria of the fitted model at each value of lambda

Author(s)

Zhixuan Fu <zhixuan.fu@yale.edu>

References

- Breheny, P. and Huang, J. (2012) Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors. *Statistics and Computing*
- Fine J. and Gray R. (1999) A proportional hazards model for the subdistribution of a competing risk. *JASA* 94:496-509.
- Fu Z., Parikh C. and Zhou B.(2015). Penalized variable selection in competing risks regression. Manuscript submitted for publication.
- Huang J., Breheny, P. and Ma, S. (2012). A selective review of group selection in high dimensional models. *Statistical Science*, 27: 481-499.

See Also

crrp, cmprsk, grpreg

Examples

```
set.seed(10)
ftime <- rexp(200)
fstatus <- sample(0:2,200,replace=TRUE)
cov <- matrix(runif(2000),nrow=200)
dimnames(cov)[[2]] <- paste("x", 1:ncol(cov))
group <- c(1,1,2,2,2,3,4,4,5,5)
#fit gSCAD penalty
fit1 <- gcrrp(ftime, fstatus, cov, group=group, penalty="gSCAD")
beta1 <- fit1$beta[, which.min(fit1$BIC)]
#fit adaptive gLASSO
fit2 <- gcrrp(ftime, fstatus, cov, group=group, penalty="gLASSO", weighted=TRUE)
beta2 <- fit2$beta[, which.min(fit2$BIC)]
```

Index