

Package ‘clogitLasso’

June 27, 2018

Type Package

Title Sparse Conditional Logistic Regression for Matched Studies

Version 1.1

Date 2018-06-21

Author Marta Avalos, Helene Pouyes, Marius Kwemou and Binbin Xu

Maintainer Marta Avalos <marta.avalos-fernandez@u-bordeaux.fr>

Description Fit a sequence of conditional logistic regression models with lasso, for small to large sized samples. Avalos, M., Pouyes, H., Grandvalet, Y., Oriols, L., & Lagarde, E. (2015) <doi:10.1186/1471-2105-16-S6-S1>.

Imports lassoshooting (>= 0.1.5),foreach

Repository CRAN

BugReports <https://github.com/mavalosf/clogitLasso>

NeedsCompilation no

LazyLoad yes

License GPL (>= 2)

Date/Publication 2018-06-27 21:34:49 UTC

RoxygenNote 6.0.1

R topics documented:

clogitLasso	2
cv.clogitLasso	3
plot.clogitLasso	5
Index	6

clogitLasso	<i>fit lasso for conditional logistic regression for matched case-control studies</i>
-------------	---

Description

Fit a sequence of conditional logistic regression with lasso penalty, for small to large sized samples

Usage

```
clogitLasso(X, y, strata, fraction = NULL, nbfraction = 100,
  nopenalize = NULL, BACK = TRUE, standardize = FALSE, maxit = 100,
  maxitB = 500, thr = 1e-10, tol = 1e-10, epsilon = 1e-04,
  trace = TRUE, log = TRUE, adaptive = FALSE, separate = FALSE,
  ols = FALSE, p.fact = NULL, remove = FALSE)
```

Arguments

X	Input matrix, of dimension nobs x nvars; each row is an observation vector
y	Binary response variable, with 1 for cases and 0 for controls
strata	Vector with stratum membership of each observation
fraction	Sequence of lambda values
nbfraction	The number of lambda values - default is 100
nopenalize	List of coefficients not to penalize starting at 0
BACK	If TRUE, use Backtracking-line search -default is TRUE
standardize	Logical flag for x variable standardization, prior to fitting the model sequence.
maxit	Maximum number of iterations of outer loop - default is 100
maxitB	Maximum number of iterations in Backtracking-line search - default is 100
thr	Threshold for convergence in lassoshooting. Default value is 1e-10. Iterations stop when max absolute parameter change is less than thr
tol	Threshold for convergence-default value is 1e-10
epsilon	ratio of smallest to largest value of regularisation parameter at which we find parameter estimates
trace	If TRUE the algorithm will print out information as iterations proceed -default is TRUE
log	If TRUE, fraction are spaced uniformly on the log scale
adaptive	If TRUE adaptive lasso is fitted-default is FALSE
separate	If TRUE, the weights in adaptive lasso are build separately using univariate models. Default is FALSE, weights are build using multivariate model
ols	If TRUE, weights less than 1 in adaptive lasso are set to 1. Default is FALSE
p.fact	Weights for adaptive lasso
remove	If TRUE, invariable covariates are removed-default is FALSE

Details

The sequence of models implied by fraction is fit by IRLS (iteratively reweighted least squares) algorithm. by coordinate descent with warm starts and sequential strong rules

Value

An object of type clogitLasso which is a list with the following components:

beta	nbfraction-by-ncol matrix of estimated coefficients. First row has all 0s
fraction	A sequence of regularisation parameters at which we obtained the fits
nz	A vector of length nbfraction containing the number of nonzero parameter estimates for the fit at the corresponding regularisation parameter
arg	List of arguments

Author(s)

Marta Avalos, Helene Pouyes, Marius Kwemou and Binbin Xu

References

Avalos, M., Pouyes, H., Grandvalet, Y., Orriols, L., & Lagarde, E. (2015). *Sparse conditional logistic regression for analyzing large-scale matched data from epidemiological studies: a simple algorithm*. BMC bioinformatics, 16(6), S1. doi: [10.1186/1471210516S6S1](https://doi.org/10.1186/1471210516S6S1).

Examples

```
## Not run:
# generate data
y <- rep(c(1,0), 100)
X <- matrix(rnorm(20000, 0, 1), ncol = 100) # pure noise
strata <- sort(rep(1:100, 2))

# 1:1
fitLasso <- clogitLasso(X,y,strata,log=TRUE)

## End(Not run)
```

cv.clogitLasso	<i>Cross-validation of clogitLasso object</i>
----------------	---

Description

Cross-validation of clogitLasso object

Usage

```
cv.clogitLasso(objclogitLasso, K = 10, gpe = NULL)
```

Arguments

objclogitLasso An objet of type clogitLasso
K The number of folds used in cross validation
gpe A list of group defined by the user.

Value

An object of type cv.clogitLasso with the following components:

lambda Vector of regularisation parameter
mean_cv vector of mean deviances for each value of the regularisation parameter
beta Vector of estimated coefficients with optimal regularisation parameter
lambdaopt Optimal regularisation parameter

Author(s)

Marta Avalos, Helene Pouyes, Marius Kwemou and Binbin Xu

References

Avalos, M., Pouyes, H., Grandvalet, Y., Orriols, L., & Lagarde, E. (2015). *Sparse conditional logistic regression for analyzing large-scale matched data from epidemiological studies: a simple algorithm*. BMC bioinformatics, 16(6), S1. doi: [10.1186/1471210516S6S1](https://doi.org/10.1186/1471210516S6S1).

Examples

```
## Not run:  
# generate data  
y <- rep(c(1,0), 100)  
X <- matrix (rnorm(20000, 0, 1), ncol = 100) # pure noise  
strata <- sort(rep(1:100, 2))  
  
# fitLasso <- clogitLasso(X,y,strata,log=TRUE)  
  
# Cross validation  
cv.fit <- cv.clogitLasso(fitLasso)  
  
## End(Not run)
```

plot.clogitLasso *Plot coefficients from a clogitLasso object*

Description

Plot the parameter profile associated clogitLasso object

Usage

```
## S3 method for class 'clogitLasso'
plot(x, logLambda = TRUE, add.legend = FALSE,
     add.labels = TRUE, lty = 1:ncol(x$beta), col = 1:ncol(x$beta), ...)
```

Arguments

x	an objet of type clogitLasso
logLambda	Set to TRUE if the horizontal axis is on log scale
add.legend	Take the value TRUE if legend should be printed in top right hand corner
add.labels	set to TRUE if labels are to be added to curves at leftmost side
lty	Same to lty parameter of plot function
col	Same to col parameter of plot function
...	additional arguments to plot function

Author(s)

Marta Avalos, Helene Pouyes, Marius Kwemou and Binbin Xu

References

Avalos, M., Pouyes, H., Grandvalet, Y., Orriols, L., & Lagarde, E. (2015). *Sparse conditional logistic regression for analyzing large-scale matched data from epidemiological studies: a simple algorithm*. BMC bioinformatics, 16(6), S1. doi: [10.1186/1471210516S6S1](https://doi.org/10.1186/1471210516S6S1).

Examples

```
## Not run:
# generate data
y <- rep(c(1,0), 100)
X <- matrix (rnorm(20000, 0, 1), ncol = 100) # pure noise
strata <- sort(rep(1:100, 2))

fitLasso <- clogitLasso(X,y,strata,log=TRUE)
# plot
plot(fitLasso)

## End(Not run)
```

Index

`clogitLasso`, [2](#)

`cv.clogitLasso`, [3](#)

`plot.clogitLasso`, [5](#)