Package 'OPTS'

October 12, 2022

Type Package

Title Optimization via Subsampling (OPTS)

Version 0.1

Date 2022-05-20

Maintainer Mihai Giurcanu <giurcanu@uchicago.edu>

Author Mihai Giurcanu [aut, cre],
Marinela Capanu [aut, ctb],
Colin Begg [aut],
Mithat Gonen [aut]

Imports MASS, cvTools, changepoint

Description Subsampling based variable selection for low dimensional generalized linear models. The methods repeatedly subsample the data minimizing an information criterion (AIC/BIC) over a sequence of nested models for each subsample. Marinela Capanu, Mihai Giurcanu, Colin B Begg, Mithat Gonen, Subsampling based variable selection for generalized linear models.

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2022-05-25 07:50:08 UTC

R topics documented:

opts 2
opts_th 3

Index **5**

opts *Optimization via Subsampling (OPTS)*

Description

opts computes the OPTS MLE in low dimensional case.

Usage

```
opts(X, Y, m, crit = "aic", prop_split = 0.5, cutoff = 0.75, ...)
```

Arguments

X	n x p covariate matrix (without intercept)
Y	n x 1 binary response vector
m	number of subsamples
crit	information criterion to select the variables: (a) aic = minimum AIC and (b) bic = minimum BIC
prop_split	proportion of subsample size and sample size, default value = 0.5
cutoff	cutoff used to select the variables using the stability selection criterion, default value = 0.75
...	other arguments passed to the glm function, e.g., family = "binomial"

Value

opts returns a list:

betahat	OPTS MLE of regression parameter vector
Jhat	estimated set of active predictors (TRUE/FALSE) corresponding to the OPTS MLE
SE	standard error of OPTS MLE
freqs	relative frequency of selection for all variables

Examples

```
require(MASS)
P = 15
N = 100
M = 20
BETA_vector = c(0.5, rep(0.5, 2), rep(0.5, 2), rep(0, P - 5))
MU_vector = numeric(P)
SIGMA_mat = diag(P)

X <- mvrnorm(N, MU_vector, Sigma = SIGMA_mat)
linearPred <- cbind(rep(1, N), X)
Y <- rbinom(N, 1, plogis(linearPred))
```

```
# OPTS-AIC MLE
opts(X, Y, 10, family = "binomial")
```

---

opts_th *Threshold OPTimization via Subsampling (OPTS_TH)*

Description

opts_th computes the threshold OPTS MLE in low dimensional case.

Usage

```
opts_th(X, Y, m, crit = "aic", type = "binseg", prop_split = 0.5,
  prop_trim = 0.2, q_tail = 0.5, ...)
```

Arguments

X	n x p covariate matrix (without intercept)
Y	n x 1 binary response vector
m	number of subsamples
crit	information criterion to select the variables: (a) aic = minimum AIC and (b) bic = minimum BIC
type	method used to minimize the trimmed and averaged information criterion: (a) min = observed minimum subsampling trimmed average information, (b) sd = observed minimum using the 0.25sd rule (corresponding to OPTS-min in the paper), (c) pelt = PELT changepoint algorithm (corresponding to OPTS-PELT in the paper), (d) binseg = binary segmentation changepoint algorithm (corresponding to OPTS-BinSeg in the paper), (e) amoc = AMOC method.
prop_split	proportion of subsample size of the sample size; default value is 0.5
prop_trim	proportion that defines the trimmed mean; default value = 0.2
q_tail	quantiles for the minimum and maximum p-values across the subsample cutpoints used to define the range of cutpoints
...	other arguments passed to the glm function, e.g., family = "binomial"

Value

opts_th returns a list:

betahat	STOPES MLE of regression parameters
SE	SE of STOPES MLE
Jhat	set of active predictors (TRUE/FALSE) corresponding to STOPES MLE
cuthat	estimated cutpoint for variable selection

pval	marginal p-values from univariate fit
cutpoits	subsample cutpoints
aic_mean	mean subsample AIC
bic_mean	mean subsample BIC

Examples

```
require(MASS)
P = 15
N = 100
M = 20
BETA_vector = c(0.5, rep(0.5, 2), rep(0.5, 2), rep(0, P - 5))
MU_vector = numeric(P)
SIGMA_mat = diag(P)

X <- mvrnorm(N, MU_vector, Sigma = SIGMA_mat)
linearPred <- cbind(rep(1, N), X)
Y <- rbinom(N, 1, plogis(linearPred))

# Threshold OPTS-BinSeg MLE
opts_th(X, Y, M, family = "binomial")
```

Index

opts, 2
opts_th, 3