

Package ‘PPLasso’

January 20, 2025

Type Package

Title Prognostic Predictive Lasso for Biomarker Selection

Version 2.0

Date 2023-02-26

Author Wencan Zhu [aut, cre],
Celine Levy-Leduc [ctb],
Nils Ternes [ctb]

Maintainer Wencan Zhu <wencan.zhu@agroparistech.fr>

Description We provide new tools for the identification of prognostic and predictive biomarkers. For further details we refer the reader to the paper: Zhu et al. Identification of prognostic and predictive biomarkers in high-dimensional data with PPLasso. BMC Bioinformatics. 2023 Jan 23;24(1):25.

License GPL-2

Imports genlasso, ggplot2, cvCovEst, glmnet, MASS

VignetteBuilder knitr

Suggests knitr, rmarkdown

NeedsCompilation no

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2023-02-27 09:12:35 UTC

Contents

PPLasso-package	2
Correction1Vect	3
Correction2Vect	4
ProgPredLasso	6
top	7
top_thresh	8

Index	9
--------------	----------

PPLasso-package

Prognostic Predictive Lasso for Biomarker Selection

Description

We provide new tools for the identification of prognostic and predictive biomarkers. For further details we refer the reader to the paper: Zhu et al. Identification of prognostic and predictive biomarkers in high-dimensional data with PPLasso. BMC Bioinformatics. 2023 Jan 23;24(1):25.

Details

The DESCRIPTION file:

```

Package:          PPLasso
Type:             Package
Title:            Prognostic Predictive Lasso for Biomarker Selection
Version:          2.0
Date:             2023-02-26
Authors@R:       c(person("Wencan", "Zhu", email = "wencan.zhu@agroparistech.fr", role = c("aut", "cre")), person("Ce
Author:          Wencan Zhu [aut, cre], Celine Levy-Leduc [ctb], Nils Ternes [ctb]
Maintainer:      Wencan Zhu <wencan.zhu@agroparistech.fr>
Description:     We provide new tools for the identification of prognostic and predictive biomarkers. For further details v
License:         GPL-2
Imports:         genlasso, ggplot2, cvCovEst, glmnet, MASS
VignetteBuilder: knitr
Suggests:        knitr, rmarkdown
NeedsCompilation: no
Packaged:        2023-02-26 08:33:08 UTC; Wencan
Depends:         R (>= 3.5.0)

```

Index of help topics:

```

Correction1Vect      Correction on two vectors
Correction2Vect      Correction on two vectors
PPLasso-package     Prognostic Predictive Lasso for Biomarker
                    Selection
ProgPredLasso       Identification of prognostic and predictive
                    biomarkers
top                  Thresholding to 0
top_thresh           Thresholding to a given threshold of the
                    smallest values

```

This package provide usfull tool for the identification of prognostics and predictive biomarkers.

Author(s)

Wencan Zhu [aut, cre], Celine Levy-Leduc [ctb], Nils Ternes [ctb]

Maintainer: Wencan Zhu <wencan.zhu@agroparistech.fr>

References

W. Zhu, C. Levy-Leduc, N. Ternes. "A variable selection approach for highly correlated predictors in high-dimensional genomic data". (2020)

Correction1Vect	<i>Correction on two vectors</i>
-----------------	----------------------------------

Description

For the estimation of β in Zhu et al. (2022), this function keeps only the M largest values coefficients set the others to 0.

Usage

```
Correction1Vect(X, Y, te = NULL, vector, top_grill. = c(1:length(vector)), delta = 0.95)
```

Arguments

X	Design matrix
Y	Response vector
te	treatment effects
vector	The vector on which we performe the thresholding
top_grill.	grill of the thresholding
delta	parameter δ in the thresholding

Value

This function returns the thresholded vector.

Author(s)

Wencan Zhu, Celine Levy-Leduc, Nils Ternes

Examples

```

vect_sample=sample(1:20,20)
X=t(sapply(c(1:10),FUN=function(x) rnorm(20)))
Y=rnorm(10)

Correction1Vect(X=X, Y=Y, vector=vect_sample)

## The function is currently defined as
function(X, Y, te=NULL, vector, top_grill.=c(1:length(vector)), delta=0.95){

  beta_interm <- sapply(top_grill., top, vect = vector)
  beta_te <- rbind(rep(te[1],length(top_grill.)), rep(te[2],length(top_grill.)), beta_interm)
  yhat <- as.matrix(X %*% beta_te)
  residuals <- sweep(yhat, 1, Y)
  mse_final_top <- colMeans(residuals^2)
  ratio_mse <- c()
  for (k in 1:(length(top_grill.) - 1)) {
    ratio_mse[k] <- round(mse_final_top[k + 1]/mse_final_top[k],6)
  }
  top_ratio <- min(which(ratio_mse >= delta))
  if (is.infinite(top_ratio)) {
    opt_final_top <- length(vector)
  }
  else {
    opt_final_top <- top_grill.[top_ratio]
  }

  return(round(top(vect = vector, thresh = opt_final_top), 6))
}

```

Correction2Vect

Correction on two vectors

Description

For the estimation of $\tilde{\beta}$ in Zhu et al. (2022), this function keeps only the K1 largest values of prognostic biomarkers coefficients and the k2 largest value of the presictive biomarkers coefficients and set the others to the smallest value among the k1 (k2) largest of prognostic (predictive part).

Usage

```

Correction2Vect(X, Y, te=NULL, vector_prog, vector_pred,
top_grill.=c(1:length(vector_prog)), delta=0.95, toZero=FALSE)

```

Arguments

X	Design matrix
Y	Response vector

te	treatment effects
vector_prog	Vector of prognostic biomarkers
vector_pred	Vector of predictive biomarkers
top_grill.	grill of the thresholding
delta	parameter δ in the thresholding
toZero	should the threshold to 0 or not

Value

This function returns the thresholded vector.

Author(s)

Wencan Zhu, Celine Levy-Leduc, Nils Ternes

Examples

```
x1=sample(1:10,10)
x2=sample(1:10,10)

X=t(sapply(c(1:10),FUN=function(x) rnorm(20)))
Y=rnorm(10)

Correction2Vect(X=X, Y=Y, vector_prog=x1, vector_pred=x2)

## The function is currently defined as
function(X, Y, te=NULL, vector_prog, vector_pred,
top_grill.=c(1:length(vector_prog)), delta=0.95, toZero=FALSE){

  if(toZero){
    matrix_top_fix <- sapply(top_grill., top, vect=vector_prog)
    matrix_top_opt <- sapply(top_grill., top, vect=vector_pred)
  } else {
    matrix_top_fix <- sapply(top_grill., top_thresh, vect=vector_prog)
    matrix_top_opt <- sapply(top_grill., top_thresh, vect=vector_pred)
  }

  opt_top_opt <- mse_fix <- c()
  for(j in 1:length(top_grill.)){
    fix_temp <- matrix_top_fix[,j]
    mse_temp <- c()
    yhat <- X*%c(te, fix_temp, matrix_top_opt[,1])

    mse_temp[1] <- sum((Y-yhat)^2)
    for(m in 2:length(top_grill.)){
      opt_temp <- matrix_top_opt[,m]
      threshed_vect <- c(te, fix_temp, opt_temp)
      yhat <- X*%threshed_vect
      mse_temp[m] <- sum((Y-yhat)^2)
    }
  }
}
```

```

ratio_mse <- round(mse_temp[m]/mse_temp[m-1], 6)
if(ratio_mse >= delta){
  opt_top_opt[j] <- top_grill.[m]
  mse_fix[j] <- mse_temp[m]
  break
}
}
if(m==length(top_grill.)){
  opt_top_opt[j] <- top_grill.[m]
  mse_fix[j] <- mse_temp[m]
}
if(j==1){
  ratio_final <- 0
} else {
  ratio_final <- mse_fix[j]/mse_fix[j-1]
}
if(ratio_final >= delta){
  opt_fix <- j
  opt_opt <- m
  break
}
}
}

if(exists("opt_fix")==FALSE){
  opt_fix <- ncol(matrix_top_fix)
  opt_opt <- ncol(matrix_top_opt)
}

return(c(matrix_top_fix[,opt_fix], matrix_top_opt[,opt_opt]))
}

```

ProgPredLasso

Identification of prognostic and predictive biomarkers

Description

The computes the regularization path of the Prognostic Predictive Lasso described in the paper Zhu et al. (2022) given in the references.

Usage

```
ProgPredLasso(X1, X2, Y=Y, cor_matrix=NULL, gamma=0.99, maxsteps=500, lambda='single')
```

Arguments

X1	Design matrix of patients characteristics with treatment 1
X2	Design matrix of patients characteristics with treatment 2

Y	Response variable
cor_matrix	Correlation matrix of biomarkers. If not specified, the function <code>cvCovEst</code> from package <code>cvCovEst</code> will be used to estimate this matrix.
gamma	Parameter γ defined in the paper Zhu et al. (2020) given in the references. Its default value is 0.99.
maxsteps	Integer specifying the maximum number of steps for the generalized Lasso algorithm. Its default value is 500.
lambda	Using single tuning parameter or both.

Value

Returns a list with the following components

lambda	different values of the parameter λ considered.
beta	matrix of the estimations of β for all the λ considered.
beta.min	estimation of β which minimize the MSE.
bic	BIC for all the λ considered.
mse	MSE for all the λ considered.

Author(s)

Wencan Zhu, Celine Levy-Leduc, Nils Ternes

top	<i>Thresholding to 0</i>
-----	--------------------------

Description

This function keeps only the K largest values of the vector and sets the others to 0.

Usage

```
top(vect, thresh)
```

Arguments

vect	vector to threshold
thresh	threshold

Value

This function returns the thresholded vector.

Author(s)

Wencan Zhu, Celine Levy-Leduc, Nils Ternes

Examples

```
x=sample(1:10,10)
thresh=3
top(x,thresh)

## The function is currently defined as
function(vect, thresh){
  sorted_vect <- sort(abs(vect),decreasing=TRUE)
  v<-sorted_vect[thresh]
  ifelse(abs(vect)>=v,vect,0)
}
```

top_thresh

Thresholding to a given threshold of the smallest values

Description

This function keeps only the K largest values of the vector and sets the others to the smallest value among the K largest.

Usage

```
top_thresh(vect,thresh)
```

Arguments

vect	vector to threshold
thresh	threshold

Value

This function returns the thresholded vector.

Author(s)

Wencan Zhu, Celine Levy-Leduc, Nils Ternes

Examples

```
x=sample(1:10,10)
thresh=3
top_thresh(x,thresh)

## The function is currently defined as
function (vect, thresh)
{
  sorted_vect <- sort(abs(vect),decreasing=TRUE)
  v = sorted_vect[thresh]
  ifelse(abs(vect) >= v, vect, v)
}
```


Index

* **~~thresholding**

Correction1Vect, [3](#)

top_thresh, [8](#)

Correction1Vect, [3](#)

Correction2Vect, [4](#)

cvCovEst, [7](#)

PPLasso (PPLasso-package), [2](#)

PPLasso-package, [2](#)

ProgPredLasso, [6](#)

top, [7](#)

top_thresh, [8](#)