Package ‘WLasso’

August 13, 2020

Type Package
Title Variable Selection for Highly Correlated Predictors
Version 1.0
Date 2020-08-07
Author Wencan Zhu [aut, cre],
          Celine Levy-Leduc [ctb],
          Nils Ternes [ctb]
Maintainer Wencan Zhu <wencan.zhu@agroparistech.fr>
Description It proposes a novel variable selection approach taking into account the correlations that may exist between the predictors of the design matrix in a high-dimensional linear model. Our approach consists in rewriting the initial high-dimensional linear model to remove the correlation between the predictors and in applying the generalized Lasso criterion. For further details we refer the reader to the paper <arXiv:2007.10768> (Zhu et al., 2020).
License GPL-2
Imports Matrix, genlasso, tibble, MASS, ggplot2
VignetteBuilder knitr
Suggests knitr, markdown
NeedsCompilation no
Depends R (>= 3.5.0)
Repository CRAN
Date/Publication 2020-08-13 09:10:10 UTC

R topics documented:

WLasso-package .................................................. 2
Sigma_Estimation ................................................ 3
top ................................................................. 4
top_thresh ...................................................... 5
Whitening_Lasso ................................................ 6
X ................................................................. 7
Y ................................................................. 7

Index 8
Description

It proposes a novel variable selection approach taking into account the correlations that may exist between the predictors of the design matrix in a high-dimensional linear model. Our approach consists in rewriting the initial high-dimensional linear model to remove the correlation between the predictors and in applying the generalized Lasso criterion. For further details we refer the reader to the paper <arXiv:2007.10768> (Zhu et al., 2020).

Details

The DESCRIPTION file:

- Package: WLasso
- Type: Package
- Title: Variable Selection for Highly Correlated Predictors
- Version: 1.0
- Date: 2020-08-07
- Authors@R: c(person("Wencan", "Zhu", email = "wencan.zhu@agroparistech.fr", role = c("aut", "cre")), person("Celine Levy-Leduc", "ctb"), person("Nils Ternes", "ctb"))
- Author: Wencan Zhu [aut, cre], Celine Levy-Leduc [ctb], Nils Ternes [ctb]
- Maintainer: Wencan Zhu <wencan.zhu@agroparistech.fr>
- Description: It proposes a novel variable selection approach taking into account the correlations that may exist between the predictors of the design matrix in a high-dimensional linear model. Our approach consists in rewriting the initial high-dimensional linear model to remove the correlation between the predictors and in applying the generalized Lasso criterion. For further details we refer the reader to the paper <arXiv:2007.10768> (Zhu et al., 2020).
- License: GPL-2
- Imports: Matrix, genlasso, tibble, MASS, ggplot2
- VignetteBuilder: knitr
- Suggests: knitr, markdown
- NeedsCompilation: no
- Packaged: 2020-08-07 12:09:08 UTC; wencan
- Depends: R (>= 3.5.0)

Index of help topics:

- Sigma_Estimation: Estimation of the correlation matrix
- WLasso-package: Variable Selection for Highly Correlated Predictors
- Whitening_Lasso: Whitening Lasso
- X: Example of a design matrix of a linear model
- Y: Example of a response variable of a linear model.
- top: Thresholding to zero of the smallest values
- top_thresh: Thresholding to a given threshold of the smallest values

This package consists of four functions: "Sigma_Estimation.R", "top.R", "top thresh.R" and "Whiten-
**Sigma_Estimation**

Estimation of the correlation matrix

**Description**

This function estimates the correlation matrix of the rows of a design matrix in linear models in the case where the correlation matrix has a block-wise structure.

**Usage**

```r
Sigma_Estimation(X)
```

**Arguments**

- `X` Design matrix of the linear model considered.

**Value**

- `mat` Estimation of the correlation matrix of the rows of `X`
- `alpha` Estimation of the coefficients of the blocks of the correlation matrix
- `group_act` Indices of the active or non active variables

**Author(s)**

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

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

**References**


**Examples**

```r
data(X)
Sigma_Estimation(X)
```
**top**

*Thresholding to zero of the smallest values*

**Description**

This function keeps only the K largest values of the vector `sorted_vect` and sets the others to zero.

**Usage**

```
top(x, thresh, sorted_vect)
```

**Arguments**

- **x** vector to threshold
- **thresh** threshold
- **sorted_vect** vector x sorted in descending order

**Value**

This function returns the thresholded vector.

**Author(s)**

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

**References**


**Examples**

```
x=sample(1:10,10)
sorted_vect=sort(x,decreasing=TRUE)
thresh=3
top(x,thresh,sorted_vect)
```
Thresholding to a given threshold of the smallest values

Description

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

Usage

```r
top_thresh(x, thresh, sorted_vect)
```

Arguments

- `x`: vector to threshold
- `thresh`: threshold
- `sorted_vect`: vector x sorted in descending order

Value

This function returns the thresholded vector.

Author(s)

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

References


Examples

```r
x <- sample(1:10, 10)
sorted_vect <- sort(x, decreasing = TRUE)
thresh <- 3
top_thresh(x, thresh, sorted_vect)
```
Whitening Lasso

Description

The function implements the approach described in the paper Zhu et al. (2020) given in the references.

Usage

Whitening_Lasso(X, Y, Sigma, gamma = 0.95, maxsteps = 2000)

Arguments

- **X**: Design matrix of the linear model.
- **Y**: Response variable of the linear model.
- **Sigma**: Correlation matrix of the rows of the design matrix. If not specified, the function Sigma_Estimation will be used to estimate this matrix.
- **gamma**: Parameter $\gamma$ defined in the paper Zhu et al. (2020) given in the references. Its default value is 0.95.
- **maxsteps**: Integer specifying the maximum number of steps for the generalized Lasso algorithm. Its default value is 2000.

Value

Returns a list with the following components

- **lambda**: different values of the parameter $\lambda$ considered.
- **beta**: matrix of the estimations of $\beta$ for all the $\lambda$ considered.
- **trans_mat**: transformation matrix which is the inverse of the square root of the estimation of the correlation matrix of the rows of the design matrix X.
- **beta.min**: estimation of $\beta$ which minimize the MSE.
- **mse**: MSE for all the $\lambda$ considered.

Author(s)

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

References

Examples

```r
data(X)
data(Y)
Sigma_est <- Sigma_Estimation(X)$mat
res_wlasso <- Whitening_Lasso(X=X, Y=Y, Sigma=Sigma_est, maxsteps=100, gamma=0.9)
beta_est <- res_wlasso$beta
```

<table>
<thead>
<tr>
<th>X</th>
<th>Example of a design matrix of a linear model</th>
</tr>
</thead>
</table>

Description

It contains an example of a design matrix of a linear model.

Usage

```r
data("X")
```

Format

The format is: num [1:50, 1:200] 0.855 1.32 1.018 -0.489 0.234 ...

Examples

```r
data(X)
```

<table>
<thead>
<tr>
<th>Y</th>
<th>Example of a response variable of a linear model.</th>
</tr>
</thead>
</table>

Description

It contains an example of a response variable of a linear model.

Usage

```r
data("Y")
```

Format

The format is: num [1:50, 1] 6.027 0.656 3.305 -1.922 -4.943 ...

Examples

```r
data(Y)
```
Index

* ~Whitening Lasso

Whitening_Lasso, 6

* ~thresholding

top, 4

* ~estimation of the correlation matrix

Sigma_Estimation, 3

* ~thresholding

top_thresh, 5

* datasets

X, 7
Y, 7

Sigma_Estimation, 3, 6

top, 4
top_thresh, 5

Whitening_Lasso, 6
WLasso (WLasso-package), 2
WLasso-package, 2

X, 7

Y, 7