Package ‘tilting’

December 26, 2016

Type  Package
Title  Variable Selection via Tilted Correlation Screening Algorithm
Version  1.1.1
Date  2016-12-22
Author  Haeran Cho [aut, cre], Piotr Fryzlewicz [aut]
Maintainer  Haeran Cho <haeran.cho@bristol.ac.uk>
Description  Implements an algorithm for variable selection in high-dimensional linear regression using the "tilted correlation", a new way of measuring the contribution of each variable to the response which takes into account high correlations among the variables in a data-driven way.
Depends  R (>= 2.14.0), mvtnorm
License  GPL (>= 2)
LazyLoad  yes
NeedsCompilation  no
Repository  CRAN
Date/Publication  2016-12-26 12:25:13

R topics documented:

<table>
<thead>
<tr>
<th>Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>tilting-package</td>
<td>2</td>
</tr>
<tr>
<td>col.norm</td>
<td>3</td>
</tr>
<tr>
<td>get.thr</td>
<td>3</td>
</tr>
<tr>
<td>lse.beta</td>
<td>4</td>
</tr>
<tr>
<td>projection</td>
<td>5</td>
</tr>
<tr>
<td>select.model</td>
<td>5</td>
</tr>
<tr>
<td>thresh</td>
<td>6</td>
</tr>
<tr>
<td>tilting</td>
<td>6</td>
</tr>
</tbody>
</table>

Index 8
Description

Implements an algorithm for variable selection in high-dimensional linear regression using the "tilted correlation", a way of measuring the contribution of each variable to the response which takes into account high correlations among the variables in a data-driven way.

Details

Package: tilting
Type: Package
Version: 1.1.1
Date: 2016-12-22
License: GPL (>= 2)
LazyLoad: yes

The main function of the package is `tilting`.

Author(s)

Haeran Cho, Piotr Fryzlewicz

Maintainer: Haeran Cho <haeran.cho@bristol.ac.uk>

References


Examples

```r
X <- matrix(rnorm(100*100), 100, 100) # 100-by-100 design matrix
y <- apply(X[,1:5], 1, sum)+rnorm(100) # first five variables are significant

tilt <- tilting(X, y, op=2)
tilt$active.nhat # returns the finally selected variables
```
**col.norm**

*Compute the L2 norm of each column*

**Description**

The function returns a vector containing the L2 norm of each column for a given matrix.

**Usage**

```r
col.norm(x)
```

**Arguments**

- `x` a matrix for which the column norms are computed.

**Value**

A vector containing the L2 norm of the columns of `X` is returned.

**Author(s)**

Haeran Cho

---

**get.thr**

*Select a threshold for sample correlation matrix*

**Description**

The function selects a threshold for sample correlation matrix.

**Usage**

```r
get.thr(C, n, p, max.num = 1, alpha = NULL, step = NULL)
```

**Arguments**

- `C` sample correlation matrix of a design matrix.
- `n` the number of observations of the design matrix.
- `p` the number of variables of the design matrix.
- `max.num` the number of times for which the threshold selection procedure is repeated. Usually `max.num==1` is used.
- `alpha` The level at which the false discovery rate is controlled. When `alpha==NULL`, it is set to be `1/sqrt(p)`.
- `step` the size of a step taken when screening the `p(p-1)/2` off-diagonal elements of `C`. 
Value

thr selected threshold.
thr.seq when max.num>1, the sequence of thresholds selected at each iteration.

Author(s)
Haeran Cho

References

Description
The function returns an estimate of the coefficient vector for a linear regression problem by setting the coefficients corresponding to a given index set to be the least squares estimate and the rest to be equal to zero.

Usage
lse.beta(X, y, active = NULL)

Arguments
X design matrix.
y response vector.
active the index set on which the least squares estimate is computed.

Value
An estimate of the coefficient vector is returned as above. If active==NULL, a vector of zeros is returned.

Author(s)
Haeran Cho
**projection**

*Compute the projection matrix onto a given set of variables*

**Description**

The function computes the projection matrix onto a set of columns of a given matrix.

**Usage**

```r
projection(x, active = NULL)
```

**Arguments**

- `x`: a matrix containing the columns onto which the projection matrix is computed.
- `active`: an index set of the columns of `X`.

**Value**

Returns the projection matrix onto the columns of "X" whose indices are included in "active". When `active==NULL`, a null matrix is returned.

**Author(s)**

Haeran Cho

---

**select.model**

*Select the final model*

**Description**

The function returns the final model as a subset of the active set chosen by Tilted Correlation Screening algorithm, for which the extended BIC is minimised.

**Usage**

```r
select.model(bic.seq, active)
```

**Arguments**

- `bic.seq`: the sequence of extended BIC at each iteration.
- `active`: the index set of selected variables by Tilted Correlation Screening algorithm.

**Value**

The index set of finally selected variables is returned.
**Author(s)**
Haeran Cho

**thresh**

Hard-threshold a matrix

**Description**
For a given matrix and a threshold, the function performs element-wise hard-thresholding based on the absolute value of each element.

**Usage**
thresh(C, alph, eps = 1e-10)

**Arguments**
- `C`: a matrix on which the hard-thresholding is performed.
- `alph`: threshold.
- `eps`: effective zero.

**Value**
Returns the matrix C after hard-thresholding.

**Author(s)**
Haeran Cho

**tilting**

Variable selection via Tilted Correlation Screening algorithm

**Description**
Given a design matrix and a response vector, the function selects a threshold for the sample correlation matrix, computes an adaptive measure for the contribution of each variable to the response variable based on the thus-thresholded sample correlation matrix, and chooses a variable at each iteration. Once variables are selected in the "active" set, the extended BIC is used for the final model selection.

**Usage**

tilting(X, y, thr.step = NULL, thr.rep = 1, max.size = NULL, max.count = NULL, op = 2, bic.gamma = 1, eps = 1e-10)
Arguments

- **X**: design matrix.
- **y**: response vector.
- **thr.step**: a step size used for threshold selection. When `thr.step==NULL`, it is chosen automatically.
- **thr.rep**: the number of times for which the threshold selection procedure is repeated.
- **max.size**: the maximum number of the variables conditional on which the contribution of each variable to the response is measured (when `max.size==NULL`, it is set to be half the number of observations).
- **max.count**: the maximum number of iterations.
- **op**: when `op==1`, rescaling 1 is used to compute the tilted correlation. If `op==2`, rescaling 2 is used.
- **bic.gamma**: a parameter used to compute the extended BIC.
- **eps**: an effective zero.

Value

- **active**: active set containing the variables selected over the iterations.
- **thr.seq**: a sequence of thresholds selected over the iterations.
- **bic.seq**: extended BIC computed over the iterations.
- **active.hat**: finally chosen variables using the extended BIC.

Author(s)

Haeran Cho

References


Examples

```r
X<-matrix(rnorm(100*100), 100, 100) # 100-by-100 design matrix
y<-apply(X[,1:5], 1, sum)+rnorm(100) # first five variables are significant

tilt<-tilting(X, y, op=2)
tilt$active.hat # returns the finally selected variables
```
Index

col.norm, 3
get.thr, 3
lse.beta, 4
projection, 5
select.model, 5
thresh, 6
tilting, 2, 6
tilting-package, 2