Package ‘parcor’
February 20, 2015

Type Package
Title Regularized estimation of partial correlation matrices
Version 0.2-6
Date 2014-09-04
Depends MASS, glmnet, ppls, Epi, GeneNet
Author Nicole Kraemer, Juliane Schaefer
Maintainer Nicole Kraemer <kraemer_r_packages@yahoo.de>
Description The package estimates the matrix of partial correlations based on different regularized regression methods: lasso, adaptive lasso, PLS, and Ridge Regression. In addition, the package provides model selection for lasso, adaptive lasso and Ridge regression based on cross-validation.
License GPL (>= 2)
LazyLoad yes
Repository CRAN
NeedsCompilation no
Date/Publication 2014-09-04 13:43:28

R topics documented:

parcor-package ...................................................... 2
adalasso ............................................................. 2
adalasso.net .......................................................... 4
Beta2parcor .......................................................... 5
lm.ridge.univariate ............................................... 7
mylars ............................................................... 8
performance.pcor .................................................... 9
pls.net .............................................................. 10
ridge.cv ............................................................ 12
ridge.net ........................................................... 13
sym2vec ............................................................. 14

Index 15
Description

This package contains different methods to estimate the matrix of partial correlations based on a \((n \times p)\) matrix \(X\) of observations. For low-dimensional settings \((p > n)\), the matrix of partial correlations can be estimated based on \(p\) least-squares regression fits. However, in high-dimensional scenarios \((p < n)\), these least-squares problems are ill-posed and need to be regularized. This package contains four different regularized regression techniques for the estimation of the partial correlations: lasso, adaptive lasso, ridge regression, and Partial Least Squares. In addition, the package provides model selection for lasso, adaptive lasso and Ridge regression based on cross-validation.

Details

<table>
<thead>
<tr>
<th>Package</th>
<th>parcor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Package</td>
</tr>
<tr>
<td>Version</td>
<td>0.2-6</td>
</tr>
<tr>
<td>Date</td>
<td>2014-09-04</td>
</tr>
<tr>
<td>License</td>
<td>GPL2 or newer</td>
</tr>
<tr>
<td>LazyLoad</td>
<td>yes</td>
</tr>
</tbody>
</table>

Author(s)

Nicole Kraemer, Juliane Schaefer
Maintainer: Nicole Kraemer <kraemer_r_packages@yahoo.de>

References

http://www.biomedcentral.com/1471-2105/10/384/

adalasso

Adaptive Lasso

Description

This function computes the lasso and adaptive lasso solution based on k-fold cross-validation. The initial weights for adaptive lasso are computed from a lasso fit.
Usage
adalasso(x, y, k = 10, use.Gram = TRUE, both = TRUE, intercept = TRUE)

Arguments

x matrix of input observations. The rows of x contain the samples, the columns of
x contain the observed variables
y vector of responses. The length of y must equal the number of rows of X
k the number of splits in k-fold cross-validation. The same k is used for the esti-
mation of the weights and the estimation of the penalty term for adaptive lasso.
Default is k = 10.
use.Gram When the number of variables is very large, you may not want LARS to pre-
compute the Gram matrix. Default is use.Gram = TRUE.
both Logical. If both = FALSE, only the lasso coefficients are computed. Default is
both = TRUE.
intercept Should an intercept be included? Default is intercept = TRUE.

Details
In each of the k-fold cross-validation steps, the weights for adaptive lasso are computed in terms of
a lasso fit. (The optimal value of the penalty term is selected via k-fold cross-validation). Note that
this implies that a lasso solution is computed k*k times!

Value

intercept.lasso intercept for lasso. If intercept = FALSE was specified, the intercept is set to 0.
intercept.adalasso intercept for adaptive lasso. If intercept = FALSE was specified, the intercept is
set to 0.
coefficients.adalasso regression coefficients for adaptive lasso.
coefficients.lasso regression coefficients for lasso.
cv.lasso cv error for the optimal lasso model.
cv.adalasso cv error for the optimal adaptive lasso model.
lambda.lasso optimal lambda value for lasso-
lambda.adalasso optimal lambda value for adaptive lasso.

Author(s)
Nicole Kraemer, Juliane Schaefer
References


http://www.biomedcentral.com/1471-2105/10/384/

See Also

Beta2parcor, adalasso.net

Examples

```r
n<-100  # number of observations
p<-60   # number of variables
X<-matrix(rnorm(n*p),ncol=p)
y<-rnorm(n)
ada.object<-adalasso(X,y,k=10)
```

adalasso.net Partial Correlations with (Adaptive) Lasso

Description

This function computes the matrix of partial correlations based on an estimation of the corresponding regression models via lasso and adaptive lasso respectively.

Usage

```r
adalasso.net(X, k = 10, use.Gram = FALSE, both = TRUE, verbose = FALSE, intercept = TRUE)
```

Arguments

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>matrix of observations. The rows of X contain the samples, the columns of X contain the observed variables.</td>
</tr>
<tr>
<td>k</td>
<td>the number of splits in k-fold cross-validation. The same k is used for the estimation of the weights and the estimation of the penalty term for adaptive lasso. Default value is k=10.</td>
</tr>
<tr>
<td>use.Gram</td>
<td>When the number of variables is very large, you may not want LARS to pre-compute the Gram matrix. Default is use.Gram=FALSE.</td>
</tr>
<tr>
<td>both</td>
<td>Logical. If both=FALSE, only the lasso solution is computed. Default is both=TRUE.</td>
</tr>
<tr>
<td>verbose</td>
<td>Print information on conflicting signs etc. Default is verbose=FALSE</td>
</tr>
<tr>
<td>intercept</td>
<td>Should an intercept be included in the regression models? Default is intercept=TRUE.</td>
</tr>
</tbody>
</table>
Details

For each of the columns of \( x \), a regression model based on (adaptive) lasso is computed. In each of the \( k \)-fold cross-validation steps, the weights for adaptive lasso are computed in terms of a lasso fit. (The optimal value of the penalty term is selected via \( k \)-fold cross-validation). Note that this implies that a lasso solution is computed \( k \times k \) times! Finally, the results of the regression models are transformed via the function \( \text{Beta2parcor} \).

Value

- \( \text{pcor.adalasso} \): estimated matrix of partial correlation coefficients for adaptive lasso.
- \( \text{pcor.lasso} \): estimated matrix of partial correlation coefficients for lasso.

Author(s)

Nicole Kraemer

References

  [http://www.biomedcentral.com/1471-2105/10/384/](http://www.biomedcentral.com/1471-2105/10/384/)

See Also

- \( \text{Beta2parcor} \), \( \text{adalasso} \)

Examples

```r
n<-20
p<-10
X<-matrix(rnorm(n*p),ncol=p)
pc<-adalasso.net(X,k=5)
```

Beta2parcor  Computation of partial correlation coefficients

Description

This function computes the matrix of partial correlation coefficients based on the results of the corresponding regression models.
Usage

Beta2parcor(Beta, verbose=FALSE)

Arguments

Beta: matrix of regression coefficients
verbose: print information on conflicting signs etc. Default is verbose=FALSE.

Details

A well-known result (Whittaker, 1990) shows that the matrix of partial correlation coefficients can be estimated by computing a least squares regression model for each variable. If there are more variables than observations, the least squares problem is ill-posed and needs regularization. The matrix Beta stores the regression coefficients of any user-defined regression method. The function Beta2parcor computes the corresponding matrix of partial correlations.

Value

matrix of partial correlation coefficients

Note

This is an internal function.

Author(s)

Nicole Kraemer

References

http://www.biomedcentral.com/1471-2105/10/384/

See Also

ridge.net, adalasso.net, pls.net

Examples

# this is an internal function and should not be called by the user
lm.ridge.univariate

Ridge Regression for a single predictor variable

Description

This function computes the ridge regression coefficients for a model with a single predictor variable.

Usage

\texttt{lm.ridge.univariate(x,y,lambda=0, scale=TRUE)}

Arguments

\begin{itemize}
  \item \texttt{x} \hspace{1cm} \text{vector of predictor observations.}
  \item \texttt{y} \hspace{1cm} \text{vector of responses. The length of y must equal the length of x}
  \item \texttt{lambda} \hspace{1cm} \text{vector of penalty terms. Default is lambda=0}
  \item \texttt{scale} \hspace{1cm} \text{Scale x and y? Default is scale=TRUE.}
\end{itemize}

Details

This function replaces the R function \texttt{lm.ridge} if only one predictor variable is used, as the latter function produces an error message in this case.

Value

A matrix with two columns and \text{length(lambda)} rows. In each row, the intercept (first column) and the regression coefficient (second column) is stored.

Author(s)

Nicole Kraemer

See Also

\texttt{ridge.net}, \texttt{ridge.cv}

Examples

\begin{verbatim}
  n<-100  # number of observations
  x<-rnorm(100)
  y<-rnorm(n)
  ridge.object<-lm.ridge.univariate(x,y,lambda=1:10)
\end{verbatim}
**mylars**  
*Cross-validation for Lasso*

**Description**

This function computes the cross-validation-optimal regression coefficients for lasso.

**Usage**

mylars(X, y, k = 10, use.Gram = TRUE, normalize = TRUE, intercept = TRUE)

**Arguments**

- **X**: matrix of observations. The rows of X contain the samples, the columns of X contain the observed variables.
- **y**: vector of responses. The length of y must equal the number of rows of X.
- **k**: the number of splits in k-fold cross-validation. Default is k=10.
- **use.Gram**: When the number of variables is very large, you may not want LARS to pre-compute the Gram matrix. Default is use.Gram = TRUE.
- **normalize**: Should the columns of X be scaled? Default is normalize = TRUE.
- **intercept**: Should an intercept be included? Default is intercept = TRUE.

**Details**

We use the glmnet() function from the glmnet package to compute the fit. Note that in Kraemer et al. (2009), we used the lars() function from the lars package, which is much slower than glmnet().

**Value**

- **lambda**: vector of parameter values from which the optimal parameter is selected.
- **cv**: cross-validated error for all lambda values.
- **lambda.opt**: cross-validation optimal parameter.
- **cv.lasso**: cv error for the optimal model.
- **intercept**: cross-validation optimal intercept. If intercept = FALSE was specified, the intercept is set to 0.
- **coefficients**: cross-validation optimal regression coefficients, without intercept.

**Author(s)**

Nicole Kraemer
References


http://www.biomedcentral.com/1471-2105/10/384/

See Also

Beta2parcor, adalasso

Examples

```r
n<-20
p<-50
X<-matrix(rnorm(n*p),ncol=p)
y<-rnorm(n)
dummy<-mylars(X,y)
```

```
| performance.pcor | Quality of estimated partial correlations |
```

Description

This function computed various performance measures of the estimated matrix of partial correlations.

Usage

```r
performance.pcor(inferred.pcor, true.pcor=NULL, fdr=TRUE, cutoff.ggm=0.8, verbose=FALSE, plot.it=FALSE)
```

Arguments

- `inferred.pcor`: matrix of estimated partial correlations
- `true.pcor`: true matrix of partial correlations. Default is `true.pcor=NULL`
- `fdr`: logical. If `fdr=TRUE`, the entries of `inferred.pcor` are tested for significance. Default is `fdr=TRUE`
- `cutoff.ggm`: default cutoff for significant partial correlations. Default is `cutoff.ggm=0.8`
- `verbose`: Print information on test results etc.. Default is `verbose=FALSE`
- `plot.it`: Plot test results and ROC-curves. Default is `plot.it=FALSE`
Details

This function computes a range of performance measures: The function always returns the number of selected edges, the binary matrix that encodes the edges, the connectivity and the percentage of positive correlations. If \texttt{true.pcor} is provided, the function also returns the power (= true positive rate), the false positive rate and the positive predictive value. For non-sparse estimates that involve testing (i.e. \texttt{fdr=TRUE}) the function also returns the area under the curve, and a pair of vectors of false and true positive rates. The latter can e.g. be used to plot a ROC-curve.

Value

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>num.selected</td>
<td>number of selected edges</td>
</tr>
<tr>
<td>adj</td>
<td>binary matrix that encodes the existence of an edge between two nodes.</td>
</tr>
<tr>
<td>connectivity</td>
<td>vector of length \texttt{ncol(inferred.pcor)}. Its ith entry indicated the number of nodes that are connected to the ith node.</td>
</tr>
<tr>
<td>positive.cor</td>
<td>percentage of positive partial correlations out of all selected edges.</td>
</tr>
<tr>
<td>power</td>
<td>power (if \texttt{true.pcor} is provided)</td>
</tr>
<tr>
<td>ppv</td>
<td>positive predictive value (if \texttt{true.pcor} is provided)</td>
</tr>
<tr>
<td>tpr</td>
<td>true positive rate (=power) (if \texttt{true.pcor} is provided)</td>
</tr>
<tr>
<td>fpr</td>
<td>true positive rate (=power) (if \texttt{true.pcor} is provided)</td>
</tr>
<tr>
<td>auc</td>
<td>area under the curve (if \texttt{true.pcor} is provided and \texttt{fdr=TRUE})</td>
</tr>
<tr>
<td>TPR</td>
<td>vector of true positive rates corresponding to varying cut-offs (if \texttt{true.pcor} is provided and \texttt{fdr=TRUE})</td>
</tr>
<tr>
<td>FPR</td>
<td>vector of false positive rates corresponding to varying cut-offs (if \texttt{true.pcor} is provided and \texttt{fdr=TRUE})</td>
</tr>
</tbody>
</table>

Author(s)

Juliane Schaefer, Nicole Kraemer

References


http://www.biomedcentral.com/1471-2105/10/384/

\textbf{Description}

This function computes the matrix of partial correlations via an estimation of the corresponding regression models via Partial Least Squares.
**Usage**

```r
pls.net(xL scale = TRUE, k = 10L, ncomp = 15L, verbose = FALSE)
```

**Arguments**

- `x`: matrix of observations. The rows of `x` contain the samples, the columns of `x` contain the observed variables.
- `scale`: Scale the columns of `x`? Default is `scale = TRUE`.
- `k`: Number of splits in k-fold cross-validation. Default value is `k = 10`.
- `ncomp`: Maximal number of components. Default is `15`.
- `verbose`: Print information on conflicting signs etc. Default is `verbose = FALSE`.

**Details**

For each of the columns of `x`, a regression model based on Partial Least Squares is computed. The optimal model is determined via cross-validation. The results of the regression models are transformed via the function `beta2parcor`.

**Value**

- `pcor`: estimated matrix of partial correlation coefficients.
- `m`: optimal number of components for each of the `ncol(x)` regression models.

**Author(s)**

Nicole Kraemer

**References**


[http://www.biomedcentral.com/1471-2105/10/384/](http://www.biomedcentral.com/1471-2105/10/384/)

**Examples**

```r
n <- 20
p <- 40
X <- matrix(rnorm(n*p), ncol = p)
pc <- pls.net(X, ncomp = 10L, k = 5)
```
Description

This function computes the optimal ridge regression model based on cross-validation.

Usage

`ridge.cv(x, y, lambda, scale = TRUE, k = 10, plot.it = FALSE)`

Arguments

- `x`: matrix of input observations. The rows of `x` contain the samples, the columns of `x` contain the observed variables.
- `y`: vector of responses. The length of `y` must equal the number of rows of `x`.
- `lambda`: Vector of penalty terms.
- `scale`: Scale the columns of `x`? Default is `scale=TRUE`.
- `k`: Number of splits in k-fold cross-validation. Default value is `k=10`.
- `plot.it`: Plot the cross-validation error as a function of `lambda`? Default is `FALSE`.

Value

- `intercept`: cross-validation optimal intercept.
- `coefficients`: cross-validation optimal regression coefficients.
- `lambda.opt`: optimal value of `lambda`.

Author(s)

Nicole Kraemer

See Also

`ridge.net`

Examples

```r
n<-100 # number of observations
p<-60 # number of variables
X<-matrix(rnorm(n*p),ncol=p)
y<-rnorm(n)
ridge.object<-ridge.cv(X,y)
```
Description

This function computes the matrix of partial correlations via an estimation of the corresponding regression models via Ridge Regression.

Usage

```
ridge.net(X, lambda, plot.it = FALSE, scale = TRUE, k = 10, verbose=FALSE)
```

Arguments

- `X`: matrix of observations. The rows of `X` contain the samples, the columns of `X` contain the observed variables.
- `lambda`: Vector of penalty terms.
- `scale`: Scale the columns of `X`? Default is scale=TRUE.
- `k`: Number of splits in k-fold cross-validation. Default value is k=10.
- `plot.it`: Plot the cross-validation error as a function of lambda? Default is FALSE.
- `verbose`: Print information on conflicting signs etc. Default is verbose=FALSE

Value

- `pcor`: estimated matrix of partial correlations.
- `lambda.opt`: optimal value of lambda for each of the ncol regression models.

Author(s)

Nicole Kraemer

References


http://www.biomedcentral.com/1471-2105/10/384/

See Also

ridge.cv

Examples

```
n<-20
p<-40
X<-matrix(rnorm(n*p),ncol=p)
pc<-ridge.net(X,k=5)
```
sym2vec

Transform symmetric matrix to vector

Description
This function transforms the upper triangle (without diagonal) of a matrix A into a vector.

Usage
sym2vec(A)

Arguments
A  quadratic matrix of size p x p

Details
The entries of the matrix A are stacked line-by-line.

Value
vector of length p(p-1)/2

Note
This is an internal function.

Author(s)
Nicole Kraemer

Examples
p<-7
A<-matrix(rnorm(p*p),ncol=p)
v<-sym2vec(A)
Index

*Topic multivariate
  adalasso, 2
  adalasso.net, 4
  Beta2parcor, 5
  mylars, 8
  performance.pcor, 9
  pls.net, 10
  ridge.cv, 12
  ridge.net, 13
  sym2vec, 14

*Topic package
  parcor-package, 2

*Topic univariate
  lm.ridge.univariate, 7

adalasso, 2, 5, 9
adalasso.net, 4, 4, 6
Beta2parcor, 4, 5, 5, 9
lm.ridge.univariate, 7
mylars, 8
parcor (parcor-package), 2
parcor-package, 2
performance.pcor, 9
pls.net, 6, 10
ridge.cv, 7, 12, 13
ridge.net, 6, 7, 12, 13
sym2vec, 14