

Package ‘spselect’

August 29, 2016

Type Package

Title Selecting Spatial Scale of Covariates in Regression Models

Version 0.0.1

Date 2016-08-29

Author Lauren Grant, David Wheeler

Maintainer Lauren Grant <pacele@vcu.edu>

Depends tester, magic, pracma

Description Fits spatial scale (SS) forward stepwise regression, SS incremental forward stagewise regression, SS least angle regression (LARS), and SS lasso models. All area-level covariates are considered at all available scales to enter a model, but the SS algorithms are constrained to select each area-level covariate at a single spatial scale.

License GPL (>= 2)

LazyLoad yes

NeedsCompilation no

Repository CRAN

Date/Publication 2016-08-29 18:53:24

R topics documented:

spselect-package	2
lars.ss	3
lasso.ss	4
stagewise.ss	6
stepwise.ss	7
X	8
X.3D	9
y	10

Index	11
--------------	-----------

`spselect-package`*Selecting spatial scale of area-level covariates in regression models*

Description

Fits spatial scale (SS) forward stepwise regression, SS incremental forward stagewise regression, SS least angle regression (LARS), and SS lasso models. All area-level covariates are considered at all available scales to enter a model, but the SS algorithms are constrained to select each area-level covariate at a single spatial scale.

Details

Package: spselect
Type: Package
Version: 0.0.1
Date: 2016-08-29
License: GPL (>=2)
LazyLoad: yes

Author(s)

Lauren Grant, David Wheeler

Maintainer: Lauren Grant <pacele@vcu.edu>

References

Grant LP, Gennings C, Wheeler, DC. (2015). Selecting spatial scale of covariates in regression models of environmental exposures. *Cancer Informatics*, 14(S2), 81-96. doi: 10.4137/CIN.S17302

Examples

```
data(y)
data(X.3D)
y.name <- "y"
ss <- c("ind", "ss1", "ss2")
mod_forward.step.ss_1 <- stepwise.ss(y, X.3D, y.name, ss, 1)
```

lars.ss	<i>Spatial scale least angle regression (LARS)</i>
---------	--

Description

This function fits a spatial scale (SS) LARS model.

Usage

```
lars.ss(y, X, ss, a.lst, S.v, C.v, col.plot, verbose=TRUE, plot=TRUE)
```

Arguments

<code>y</code>	A numeric response vector
<code>X</code>	A data frame of numeric variables
<code>ss</code>	A vector of names to identify the different levels of covariates available as potential candidates for model input
<code>a.lst</code>	A list of identity matrices, where each column indicates a particular level or spatial scale for a specified covariate (e.g., <code>ss1_x2</code>)
<code>S.v</code>	A vector of positive integers, where each number denotes the number of spatial scales associated with a particular covariate
<code>C.v</code>	A vector, where all values are initialized to 0
<code>col.plot</code>	A vector of colors (corresponding to each SS) used in the coefficient path plot
<code>verbose</code>	If TRUE, details are printed as the algorithm progresses
<code>plot</code>	If TRUE, a coefficient path plot is generated

Details

This function estimates coefficients using the SS LARS modeling approach. The function also provides summary details and plots a coefficient path plot.

Value

A list with the following items:

<code>beta</code>	Regression coefficient estimates from all set of model solutions
<code>beta.aic</code>	Regression coefficient estimates from final model
<code>ind.v</code>	Vector of indices to denote the corresponding columns of X associated with each active predictor
<code>aic.v</code>	Vector of Akaike information criterion (AIC) values
<code>stack.ss</code>	Vector of indices to indicate the level at which each covariate enters the model

Author(s)

Lauren Grant, David Wheeler

References

Grant LP, Gennings C, Wheeler, DC. (2015). Selecting spatial scale of covariates in regression models of environmental exposures. *Cancer Informatics*, 14(S2), 81-96. doi: 10.4137/CIN.S17302

Examples

```

data(y)
data(X)

names.X <- colnames(X)

ss <- c("ind", "ss1", "ss2")

a.lst <- list(NULL)
a.lst[[1]] <- 1
dim(a.lst[[1]]) <- c(1,1)
dimnames(a.lst[[1]]) <- list(NULL, names.X[1])

a.lst[[2]] <- diag(2)
dimnames(a.lst[[2]]) <- list(NULL, names.X[c(2,3)])

a.lst[[3]] <- diag(2)
dimnames(a.lst[[3]]) <- list(NULL, names.X[c(4,5)])

S.v <- c(1,2,2)
C.v <- rep(0,length(a.lst))

mod_LARS.ss <- lars.ss(y, X, ss, a.lst, S.v, C.v, c("black", "red", "green"))

```

lasso.ss

Spatial scale lasso

Description

This function fits a spatial scale (SS) lasso model.

Usage

```
lasso.ss(y, X, ss, a.lst, S.v, C.v, col.plot, verbose=TRUE, plot=TRUE)
```

Arguments

<code>y</code>	A numeric response vector
<code>X</code>	A data frame of numeric variables
<code>ss</code>	A vector of names to identify the different levels of covariates available as potential candidates for model input
<code>a.lst</code>	A list of identity matrices, where each column indicates a particular level or spatial scale for a specified covariate (e.g., <code>ss1_x2</code>)

<code>S.v</code>	A vector of positive integers, where each number denotes the number of spatial scales associated with a particular covariate
<code>C.v</code>	A vector, where all values are initialized to 0
<code>col.plot</code>	A vector of colors (corresponding to each SS) used in the coefficient path plot
<code>verbose</code>	If TRUE, details are printed as the algorithm progresses
<code>plot</code>	If TRUE, a coefficient path plot is generated

Details

This function estimates coefficients using the SS lasso modeling approach. The function also provides summary details and plots a coefficient path plot.

Value

A list with the following items:

<code>beta</code>	Regression coefficient estimates from all set of model solutions
<code>beta.aic</code>	Regression coefficient estimates from final model
<code>ind.v</code>	Vector of indices to denote the corresponding columns of X associated with each active predictor
<code>aic.v</code>	Vector of Akaike information criterion (AIC) values
<code>stack.ss</code>	Vector of indices to indicate the level at which each covariate enters the model

Author(s)

Lauren Grant, David Wheeler

References

Grant LP, Gennings C, Wheeler, DC. (2015). Selecting spatial scale of covariates in regression models of environmental exposures. *Cancer Informatics*, 14(S2), 81-96. doi: 10.4137/CIN.S17302

Examples

```
data(y)
data(X)

names.X <- colnames(X)

ss <- c("ind", "ss1", "ss2")

a.lst <- list(NULL)
a.lst[[1]] <- 1
dim(a.lst[[1]]) <- c(1,1)
dimnames(a.lst[[1]]) <- list(NULL, names.X[1])

a.lst[[2]] <- diag(2)
dimnames(a.lst[[2]]) <- list(NULL, names.X[c(2,3)])
```

```

a.lst[[3]] <- diag(2)
dimnames(a.lst[[3]]) <- list(NULL, names.X[c(4,5)])

S.v <- c(1,2,2)
C.v <- rep(0,length(a.lst))

mod_lasso.ss <- lasso.ss(y, X, ss, a.lst, S.v, C.v, c("black", "red", "green"))

```

stagewise.ss

Spatial scale incremental forward stagewise regression

Description

This function fits a spatial scale (SS) incremental forward stagewise regression model.

Usage

```
stagewise.ss(y, X, X.3D, ss, increment, tolerance, col.plot, verbose=TRUE, plot=TRUE)
```

Arguments

y	A numeric response vector
X	A data frame of numeric variables
X.3D	A 3-D or stacked array of numeric variables, where each stack represents a particular level of covariates (i.e., individual- and area-level variables at more than one spatial scale). In cases where values are only present for a covariate at certain levels, that covariate is assigned missing values at all other levels.
ss	A vector of names to identify the different levels of covariates available as potential candidates for model input
increment	A positive step size
tolerance	A small, positive value used as a stopping criterion when none of the predictors are correlated with the residuals. The algorithm stops if the overall maximum correlation is less than a specified tolerance.
col.plot	A vector of colors (corresponding to each SS) used in the coefficient path plot
verbose	If TRUE, details are printed as the algorithm progresses
plot	If TRUE, a coefficient path plot is generated

Details

This function estimates coefficients using the SS forward stagewise regression approach. The function also provides summary details and plots a coefficient path plot.

Value

A list with the following items:

beta.final	Regression coefficient estimates from final model
stack.ss	Vector of indices to indicate the level at which each covariate enters the model

Author(s)

Lauren Grant, David Wheeler

References

Grant LP, Gennings C, Wheeler, DC. (2015). Selecting spatial scale of covariates in regression models of environmental exposures. *Cancer Informatics*, 14(S2), 81-96. doi: 10.4137/CIN.S17302

Examples

```
data(y)
data(X)
data(X.3D)
ss <- c("ind", "ss1", "ss2")
mod_forward.stage.ss_0.1 <- stagewise.ss(y, X ,X.3D, ss, 0.1, 0.1, c("black", "red", "green"))
```

stepwise.ss

Spatial scale forward stepwise regression

Description

This function fits a spatial scale (SS) forward stepwise regression model.

Usage

```
stepwise.ss(y, X.3D, y.name, ss, epsilon, verbose=TRUE)
```

Arguments

y	A numeric response vector
X.3D	A 3-D or stacked array of numeric variables, where each stack represents a particular level of covariates (i.e., individual- and area-level variables at more than one spatial scale). In cases where values are only present for a covariate at certain levels, that covariate is assigned missing values at all other levels.
y.name	A name for y
ss	A vector of names to identify the different levels of covariates available as potential candidates for model input
epsilon	A positive value used as a stopping criterion when there is inadequate improvement in the model's performance. The algorithm stops if the difference in the Akaike information criterion (AIC) between the current model and the proposed model is less than epsilon.
verbose	If TRUE, details are printed as the algorithm progresses

Details

This function estimates coefficients using the SS forward stepwise regression approach. The function also estimates the model fit and provides summary details.

Value

A list with the following items:

<code>beta.final</code>	Regression coefficient estimates from final model
<code>aic.final</code>	AIC for final model
<code>summary.final</code>	Summary output of final model
<code>stack.ss</code>	Vector of indices to indicate the level at which each covariate enters the model

Author(s)

Lauren Grant, David Wheeler

References

Grant LP, Gennings C, Wheeler, DC. (2015). Selecting spatial scale of covariates in regression models of environmental exposures. *Cancer Informatics*, 14(S2), 81-96. doi: 10.4137/CIN.S17302

Examples

```
data(y)
data(X.3D)
y.name <- "y"
ss <- c("ind", "ss1", "ss2")
mod_forward.step.ss_1 <- stepwise.ss(y, X.3D, y.name, ss, 1)
```

X

Input data X

Description

Simulated input data

Usage

```
data(X)
```

Format

A data frame with 20 observations on the following 5 variables:

<code>x1</code>	a numeric vector
<code>ss1_x2</code>	a numeric vector
<code>ss2_x2</code>	a numeric vector
<code>ss1_x3</code>	a numeric vector
<code>ss2_x3</code>	a numeric vector

Details

The data consist of simulated variables, including an individual-level covariate (x1) and two area-level covariates (x2, x3) available at two different spatial scales (ss1, ss2).

Examples

```
data(X)
```

X.3D

Input data X.3D

Description

Simulated input data (in stacked array format)

Usage

```
data(X.3D)
```

Format

A 20x3x3 stacked array with the following stacks:

ind a numeric array containing an individual-level variable (x1)

ss1 a numeric array containing area-level variables (x2, x3) available at ss1

ss2 a numeric array containing area-level variables (x2, x3) available at ss2

Details

The data consist of simulated variables, including an individual-level covariate (x1) and two area-level covariates (x2, x3) available at two different spatial scales (ss1, ss2). The data are in the form of a 3-D or stacked array, where each stack represents a particular level of covariates, including spatial scale. The first stack contains the individual-level variable; the second and third stacks contain the area-level variables at the ss1 and ss2 levels, respectively. Note that in cases where values are only present for a covariate at certain levels, that covariate is assigned missing values at all other levels.

References

Grant LP, Gennings C, Wheeler, DC. (2015). Selecting spatial scale of covariates in regression models of environmental exposures. *Cancer Informatics*, 14(S2), 81-96. doi: 10.4137/CIN.S17302

Examples

```
data(X.3D)
```

y	<i>Response data y</i>
---	------------------------

Description

Simulated response data

Usage

`data(y)`

Format

A numeric response vector with 20 observations

Examples

`data(y)`

Index

*Topic **datasets**

X, 8

X.3D, 9

y, 10

*Topic **models**

lars.ss, 3

lasso.ss, 4

stagewise.ss, 6

stepwise.ss, 7

*Topic **package**

spselect-package, 2

lars.ss, 3

lasso.ss, 4

spselect (spselect-package), 2

spselect-package, 2

stagewise.ss, 6

stepwise.ss, 7

X, 8

X.3D, 9

y, 10