Package 'sparsestep'

January 27, 2017

Version 1.0.0

Date 2017-01-26

Title SparseStep Regression

Description Implements the SparseStep model for solving regression problems with a sparsity constraint on the parameters. The SparseStep regression model was proposed in Van den Burg, Groenen, and Alfons (2017) <https://arxiv.org/abs/1701.06967>. In the model, a regularization term is added to the regression problem which approximates the counting norm of the parameters. By iteratively improving the approximation a sparse solution to the regression problem can be obtained. In this package both the standard SparseStep algorithm is implemented as well as a path algorithm which uses golden section search to determine solutions with different values for the regularization parameter.

License GPL (>= 2)

Imports graphics

Depends R (>= 3.0.0), Matrix (>= 1.0-6)

Classification/MSC 62J05, 62J07

URL https://github.com/GjjvdBurg/SparseStep, https://arxiv.org/abs/1701.06967

BugReports https://github.com/GjjvdBurg/SparseStep

RoxygenNote 5.0.1

NeedsCompilation no

Author Gertjan van den Burg [aut, cre], Patrick Groenen [ctb], Andreas Alfons [ctb]

Maintainer Gertjan van den Burg <gertjanvandenburg@gmail.com>

Repository CRAN

Date/Publication 2017-01-27 10:18:45

R topics documented:

`sparsestep-package` *SparseStep: Approximating the Counting Norm for Sparse Regularization*

Description

In the SparseStep regression model the ordinary least-squares problem is augmented with an approximation of the exact ℓ_0 pseudonorm. This approximation is made increasingly more accurate in the SparseStep algorithm, resulting in a sparse solution to the regression problem. See the references for more information.

SparseStep functions

The main SparseStep functions are:

`sparsestep` Fit a SparseStep model for a given range of λ values

`path.sparsestep` Fit the SparseStep model along a path of λ values which are generated such that a model is created at each possible level of sparsity, or until a given recursion depth is reached.

Other available functions are:

`plot` Plot the coefficient path of the SparseStep model.

`predict` Predict the outcome of the linear model using SparseStep

`coef` Get the coefficients from the SparseStep model

`print` Print a short description of the SparseStep model

Author(s)

Gerrit J.J. van den Burg, Patrick J.F. Groenen, Andreas Alfons
Maintainer: Gerrit J.J. van den Burg <gertjanvandenburg@gmail.com>

References

Van den Burg, G.J.J., Groenen, P.J.F. and Alfons, A. (2017). *SparseStep: Approximating the Counting Norm for Sparse Regularization*, arXiv preprint arXiv:1701.06967 [stat.ME]. URL https://arxiv.org/abs/1701.06967.

Examples

```
x <- matrix(rnorm(100*20), 100, 20)
y <- rnorm(100)
fit <- sparsestep(x, y)
plot(fit)
fits <- path.sparsestep(x, y)
plot(fits)
x2 <- matrix(rnorm(50*20), 50, 20)
y2 <- predict(fits, x2)
```

`coef.sparsestep` *Get the coefficients of a fitted SparseStep model*

Description

Returns the coefficients of the SparseStep model.

Usage

```
## S3 method for class 'sparsestep'
coef(object, ...)
```

Arguments

`object`	a `sparsestep` object
`...`	further argument are ignored

Value

The coefficients of the SparseStep model (i.e. the betas) as a dgCMatrix. If the model was fitted with an intercept this will be the first row in the resulting matrix.

Author(s)

Gerrit J.J. van den Burg, Patrick J.F. Groenen, Andreas Alfons
Maintainer: Gerrit J.J. van den Burg <gertjanvandenburg@gmail.com>

References

Van den Burg, G.J.J., Groenen, P.J.F. and Alfons, A. (2017). *SparseStep: Approximating the Counting Norm for Sparse Regularization*, arXiv preprint arXiv:1701.06967 [stat.ME]. URL https://arxiv.org/abs/1701.06967.

Examples

```
x <- matrix(rnorm(100*20), 100, 20)
y <- rnorm(100)
fit <- sparsestep(x, y)
coef(fit)
```

`path.sparsestep` *Approximate path algorithm for the SparseStep model*

Description

Fits the entire regularization path for SparseStep using a Golden Section search. Note that this algorithm is approximate, there is no guarantee that the solutions _between_ induced values of lambdas do not differ from those calculated. For instance, if solutions are calculated at λ_i and λ_{i+1}, this algorithm ensures that λ_{i+1} has one more zero than the solution at λ_i (provided the recursion depth is large enough). There is however no guarantee that there are no different solutions between λ_i and λ_{i+1}. This is an ongoing research topic.

Note that this path algorithm is not faster than running the `sparsestep` function with the same λ sequence.

Usage

```
path.sparsestep(x, y, max.depth = 10, gamma0 = 1000, gammastop = 1e-04,
  IMsteps = 2, gammastep = 2, normalize = TRUE, intercept = TRUE,
  force.zero = TRUE, threshold = 1e-07, XX = NULL, Xy = NULL,
  use.XX = TRUE, use.Xy = TRUE, quiet = FALSE)
```

Arguments

`x`	matrix of predictors
`y`	response
`max.depth`	maximum recursion depth
`gamma0`	starting value of the gamma parameter
`gammastop`	stopping value of the gamma parameter
`IMsteps`	number of steps of the majorization algorithm to perform for each value of gamma
`gammastep`	factor to decrease gamma with at each step
`normalize`	if TRUE, each variable is standardized to have unit L2 norm, otherwise it is left alone.
`intercept`	if TRUE, an intercept is included in the model (and not penalized), otherwise no intercept is included
`force.zero`	if TRUE, absolute coefficients smaller than the provided threshold value are set to absolute zero as a post-processing step, otherwise no thresholding is performed

threshold	threshold value to use for setting coefficients to absolute zero
XX	The X'X matrix; useful for repeated runs where X'X stays the same
Xy	The X'y matrix; useful for repeated runs where X'y stays the same
use.XX	whether or not to compute X'X and return it
use.Xy	whether or not to compute X'y and return it
quiet	don't print search info while running

Value

A "sparsestep" S3 object is returned, for which print, predict, coef, and plot methods exist. It has the following items:

call	The call that was used to construct the model.
lambda	The value(s) of lambda used to construct the model.
gamma0	The gamma0 value of the model.
gammastop	The gammastop value of the model
IMsteps	The IMsteps value of the model
gammastep	The gammastep value of the model
intercept	Boolean indicating if an intercept was fitted in the model
force.zero	Boolean indicating if a force zero-setting was performed.
threshold	The threshold used for a forced zero-setting
beta	The resulting coefficients stored in a sparse matrix format (dgCMatrix). This matrix has dimensions nvar x nlambda
a0	The intercept vector for each value of gamma of length nlambda
normx	Vector used to normalize the columns of x
meanx	Vector of column means of x
XX	The matrix X'X if use.XX was set to TRUE
Xy	The matrix X'y if use.Xy was set to TRUE

Author(s)

Gerrit J.J. van den Burg, Patrick J.F. Groenen, Andreas Alfons
Maintainer: Gerrit J.J. van den Burg <gertjanvandenburg@gmail.com>

References

Van den Burg, G.J.J., Groenen, P.J.F. and Alfons, A. (2017). *SparseStep: Approximating the Counting Norm for Sparse Regularization*, arXiv preprint arXiv:1701.06967 [stat.ME]. URL https://arxiv.org/abs/1701.06967.

See Also

coef, print, predict, plot, and sparsestep.

Examples

```
x <- matrix(rnorm(100*20), 100, 20)
y <- rnorm(100)
pth <- path.sparsestep(x, y)
```

`plot.sparsestep` *Plot the SparseStep path*

Description

Plot the coefficients of the SparseStep path

Usage

```
## S3 method for class 'sparsestep'
plot(x, ...)
```

Arguments

`x`	a `sparsestep` object
`...`	further argument to matplot

Author(s)

Gerrit J.J. van den Burg, Patrick J.F. Groenen, Andreas Alfons
Maintainer: Gerrit J.J. van den Burg <gertjanvandenburg@gmail.com>

References

Van den Burg, G.J.J., Groenen, P.J.F. and Alfons, A. (2017). *SparseStep: Approximating the Counting Norm for Sparse Regularization*, arXiv preprint arXiv:1701.06967 [stat.ME]. URL https://arxiv.org/abs/1701.06967.

Examples

```
x <- matrix(rnorm(100*20), 100, 20)
y <- rnorm(100)
fit <- sparsestep(x, y)
plot(fit)
pth <- path.sparsestep(x, y)
plot(pth)
```

`predict.sparsestep` *Make predictions from a SparseStep model*

Description

Predicts the outcome variable for the SparseStep model for each value of lambda supplied to the model.

Usage

```
## S3 method for class 'sparsestep'
predict(object, newx, ...)
```

Arguments

`object`	Fitted `sparsestep` object
`newx`	Matrix of new values for `x` at which predictions are to be made.
`...`	further argument are ignored

Value

a matrix of numerical predictions of size nobs x nlambda

Author(s)

Gerrit J.J. van den Burg, Patrick J.F. Groenen, Andreas Alfons
Maintainer: Gerrit J.J. van den Burg <gertjanvandenburg@gmail.com>

References

Van den Burg, G.J.J., Groenen, P.J.F. and Alfons, A. (2017). *SparseStep: Approximating the Counting Norm for Sparse Regularization*, arXiv preprint arXiv:1701.06967 [stat.ME]. URL https://arxiv.org/abs/1701.06967.

Examples

```
x <- matrix(rnorm(100*20), 100, 20)
y <- rnorm(100)
fit <- sparsestep(x, y)
yhat <- predict(fit, x)
```

print.sparsestep *Print the fitted SparseStep model*

Description

Prints a short text of a fitted SparseStep model

Usage

```
## S3 method for class 'sparsestep'
print(x, ...)
```

Arguments

x	a sparsestep object to print
...	further argument are ignored

Author(s)

Gerrit J.J. van den Burg, Patrick J.F. Groenen, Andreas Alfons
Maintainer: Gerrit J.J. van den Burg <gertjanvandenburg@gmail.com>

References

Van den Burg, G.J.J., Groenen, P.J.F. and Alfons, A. (2017). *SparseStep: Approximating the Counting Norm for Sparse Regularization*, arXiv preprint arXiv:1701.06967 [stat.ME]. URL https://arxiv.org/abs/1701.06967.

Examples

```
x <- matrix(rnorm(100*20), 100, 20)
y <- rnorm(100)
fit <- sparsestep(x, y)
print(fit)
```

sparsestep *Fit the SparseStep model*

Description

Fits the SparseStep model for a chosen values of the regularization parameter.

Usage

```
sparsestep(x, y, lambda = c(0.1, 0.5, 1, 5, 10), gamma0 = 1000,
  gammastop = 1e-04, IMsteps = 2, gammastep = 2, normalize = TRUE,
  intercept = TRUE, force.zero = TRUE, threshold = 1e-07, XX = NULL,
  Xy = NULL, use.XX = TRUE, use.Xy = TRUE)
```

Arguments

x	matrix of predictors
y	response
lambda	regularization parameter
gamma0	starting value of the gamma parameter
gammastop	stopping value of the gamma parameter
IMsteps	number of steps of the majorization algorithm to perform for each value of gamma
gammastep	factor to decrease gamma with at each step
normalize	if TRUE, each variable is standardized to have unit L2 norm, otherwise it is left alone.
intercept	if TRUE, an intercept is included in the model (and not penalized), otherwise no intercept is included
force.zero	if TRUE, absolute coefficients smaller than the provided threshold value are set to absolute zero as a post-processing step, otherwise no thresholding is performed
threshold	threshold value to use for setting coefficients to absolute zero
XX	The X'X matrix; useful for repeated runs where X'X stays the same
Xy	The X'y matrix; useful for repeated runs where X'y stays the same
use.XX	whether or not to compute X'X and return it (boolean)
use.Xy	whether or not to compute X'y and return it (boolean)

Value

A "sparsestep" S3 object is returned, for which print, predict, coef, and plot methods exist. It has the following items:

call	The call that was used to construct the model.
lambda	The value(s) of lambda used to construct the model.
gamma0	The gamma0 value of the model.
gammastop	The gammastop value of the model
IMsteps	The IMsteps value of the model
gammastep	The gammastep value of the model
intercept	Boolean indicating if an intercept was fitted in the model
force.zero	Boolean indicating if a force zero-setting was performed.

threshold	The threshold used for a forced zero-setting
beta	The resulting coefficients stored in a sparse matrix format (dgCMatrix). This matrix has dimensions nvar x nlambda
a0	The intercept vector for each value of gamma of length nlambda
normx	Vector used to normalize the columns of x
meanx	Vector of column means of x
XX	The matrix X'X if use.XX was set to TRUE
Xy	The matrix X'y if use.Xy was set to TRUE

Author(s)

Gerrit J.J. van den Burg, Patrick J.F. Groenen, Andreas Alfons
Maintainer: Gerrit J.J. van den Burg <gertjanvandenburg@gmail.com>

References

Van den Burg, G.J.J., Groenen, P.J.F. and Alfons, A. (2017). *SparseStep: Approximating the Counting Norm for Sparse Regularization*, arXiv preprint arXiv:1701.06967 [stat.ME]. URL https://arxiv.org/abs/1701.06967.

See Also

coef, print, predict, plot, and path.sparsestep.

Examples

```
x <- matrix(rnorm(100*20), 100, 20)
y <- rnorm(100)
fit <- sparsestep(x, y)
```

Index