

Package ‘glmnetSE’

November 5, 2021

Type Package

Title Add Nonparametric Bootstrap SE to 'glmnet' for Selected Coefficients (No Shrinkage)

Version 0.0.1

Author Sebastian Bahr [cre, aut]

Description Builds a LASSO, Ridge, or Elastic Net model with 'glmnet' or 'cv.glmnet' with bootstrap inference statistics (SE, CI, and p-value) for selected coefficients with no shrinkage applied for them. Model performance can be evaluated on test data and an automated alpha selection is implemented for Elastic Net. Parallelized computation is used to speed up the process. The methods are described in Friedman et al. (2010) <doi:10.18637/jss.v033.i01> and Simon et al. (2011) <doi:10.18637/jss.v039.i05>.

License GPL-3

URL <https://github.com/sebastianbahr/glmnetSE>

BugReports <https://github.com/sebastianbahr/glmnetSE/blob/main/issues>

Depends R (>= 4.0.0)

Imports boot, glmnet, graphics, parallel, stats

Encoding UTF-8

Language en-US

RoxygenNote 7.1.1

NeedsCompilation no

Maintainer Sebastian Bahr <sebastian.bahr@unibe.ch>

Repository CRAN

Date/Publication 2021-11-05 15:40:02 UTC

R topics documented:

glmnetSE	2
plot.glmnetSE	4
summary.glmnetSE	5

glmnetSE	<i>Add Nonparametric Bootstrap SE to 'glmnet' for Selected Coefficients (No Shrinkage)</i>
----------	--

Description

Builds a LASSO, Ridge, or Elastic Net model with `glmnet` or `cv.glmnet` with bootstrap inference statistics (SE, CI, and p-value) for selected coefficients with no shrinkage applied for them. Model performance can be evaluated on test data and an automated alpha selection is implemented for Elastic Net. Parallelized computation is used to speed up the process.

Usage

```
glmnetSE(
  data,
  cf.no.shrnkg,
  alpha = 1,
  method = "10CVoneSE",
  test = "none",
  r = 250,
  nlambdas = 100,
  seed = 0,
  family = "gaussian",
  type = "basic",
  conf = 0.95,
  perf.metric = "mse",
  ncore = "mx.core"
)
```

Arguments

<code>data</code>	A data frame, tibble, or matrix object with the outcome variable in the first column and the feature variables in the following columns. Note: all columns beside the first one are used as feature variables. Feature selection has to be done beforehand.
<code>cf.no.shrnkg</code>	A character string of the coefficients whose effect size will be interpreted, the inference statistic is of interest and therefore no shrinkage will be applied.
<code>alpha</code>	Alpha value [0,1]. An alpha of 0 results in a ridge regression, a value of 1 in a LASSO, and a value between 0 and 1 in an Elastic Net. If a sequence of possible alphas is passed to the <code>alpha</code> argument the alpha of the best performing model (based on the selected method and <code>perf.metric</code>) is selected - default is 1.
<code>method</code>	A character string defining if 10-fold cross-validation is used or not. Possible methods are <code>none</code> : no cross-validation is applied and the coefficients for $\lambda = 0.1$ are selected. <code>10CVoneSE</code> : 10-fold cross-validation is applied and the lambda of the least complex model with an MSE within one standard error of

	the smallest MSE is selected. 10CVmin: 10-fold cross-validation is applied and the lambda at which the MSE is the smallest is selected - default is 10CVoneSE.
test	A data frame, tibble, or matrix object with the same outcome and feature variables as supplied to data which includes test-observations not used for the training of the model.
r	Number of nonparametric bootstrap repetitions - default is 250
nlambda	Number of tested lambda values - default is 100.
seed	Seed set for the cross-validation and bootstrap sampling - default 0 which means no seed set.
family	A character string representing the used model family either gaussian or binomial - default is gaussian.
type	A character string indicating the type of calculated bootstrap intervals. It can be norm, basic, perc, or bca. For more information check the boot.ci package - default is basic.
conf	Indicates the confidence interval level - default is 0.95.
perf.metric	A character string indicating the used performance metric to evaluate the performance of different lambdas and the final model. Can be either mse (mean squared error), mae (mean absolute error), class (classification error), or auc (area under the curve). Is not applied when method none is used - default is mse.
ncore	A numerical value indicates the number of build clusters and used cores in the computation. If not defined the maximum available number of cores of the OS -1 is used mx.core. It is not possible to use more than 32 cores, because efficiency decreases rapidly at this point see (Sloan et al. 2014) - default is mx.core.

Value

glmnetSE object which output can be displayed using `summary()` or `summary.glmnetSE()`. If family binomial and performance metric auc is used it is possible to plot the ROC curve with `plot()` or `plot.glmnetSE()`.

Author(s)

Sebastian Bahr, <sebastian.bahr@unibe.ch>

References

- Friedman J., Hastie T. and Tibshirani R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1), 1-22. <https://www.jstatsoft.org/v33/i01/>.
- Simon N., Friedman J., Hastie T. and Tibshirani R. (2011). Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *Journal of Statistical Software*, 39(5), 1-13. <https://www.jstatsoft.org/v39/i05/>.
- Efron, B. and Tibshirani, R. (1993) An Introduction to the Bootstrap. Chapman & Hall. https://cds.cern.ch/record/526679/files/0412042312_TOC.pdf
- Sloan T.M., Piotrowski M., Forster T. and Ghazal P. (2014) Parallel Optimization of Bootstrapping in R. <https://arxiv.org/ftp/arxiv/papers/1401/1401.6389.pdf>

See Also

[summary.glmnetSE](#) and [plot.glmnetSE](#) methods.

Examples

```
# LASSO model with gaussian function, no cross validation, a seed of 123, and
# the coefficient of interest is Education. Two cores are used for the computation

glmnetSE(data=swiss, cf.no.shrnkg = c("Education"), alpha=1, method="none", seed = 123, ncore = 2)

# Ridge model with binomial function, 10-fold cross validation selecting the lambda
# at which the smallest MSE is achieved, 500 bootstrap repetitions, no seed, the
# misclassification error is used as performance metric, and the coefficient of
# interest are Education and Catholic. Two cores are used for the computation.

# Generate dichotom variable
swiss$Fertility <- ifelse(swiss$Fertility >= median(swiss$Fertility), 1, 0)

glmnetSE(data=swiss, cf.no.shrnkg = c("Education", "Catholic"), alpha=0, method="10CVmin", r=500,
          seed = 0, family="binomial", perf.metric = "class", ncore = 2)

# Elastic Net with gaussian function, automated alpha selection, selection the lambda
# within one standard deviation of the best model, test data to obtain the performance
# metric on it, a seed of 123, bias-corrected and accelerated confidence intervals, a
# level of 0.9, the performance metric MAE, and the coefficient of interest is Education.
# Two cores are used for the computation

# Generate a train and test set
set.seed(123)
train_sample <- sample(nrow(swiss), 0.8*nrow(swiss))

swiss.train <- swiss[train_sample, ]
swiss.test  <- swiss[-train_sample, ]

glmnetSE(data=swiss.train, cf.no.shrnkg = c("Education"), alpha=seq(0.1,0.9,0.1),
          method="10CVoneSE", test = swiss.test, seed = 123, family = "gaussian", type = "bca",
          conf = 0.9, perf.metric = "mae", ncore = 2)
```

plot.glmnetSE

Plot ROC Curve of a fitted glmnetSE Model on Test Data

Description

Plot the ROC curve of a fitted model [glmnetSE](#) (family binomial and performance metric auc) on supplied test data.

Usage

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

Arguments

x A model of the class glmnetSE of family binomial and performance metric auc for which the ROC curve should be plotted.

... Additional arguments affecting the plot produced.

Value

The ROC curve of a glmnetSE object.

Examples

```
# Generate dichotom variable  
  
swiss$Fertility <- ifelse(swiss$Fertility >= median(swiss$Fertility), 1, 0)  
  
# Generate a train and test set  
set.seed(1234)  
train_sample <- sample(nrow(swiss), 0.8*nrow(swiss))  
  
swiss.train <- swiss[train_sample, ]  
swiss.test  <- swiss[-train_sample, ]  
  
# Estimate model  
  
glmnetSE.model <- glmnetSE(data=swiss.train, cf.no.shrnkg = c("Education"),  
alpha=seq(0.1,0.9,0.1), method = "10CVoneSE", test = swiss.test, seed = 123,  
family = "binomial", perf.metric = "auc", ncore = 2)  
  
# Plot ROC curve of the fitted model on swiss.test data  
  
plot(glmnetSE.model)
```

Description

Print the coefficients with standard errors, confidence intervals, and p-values of a `glmnetSE` model. The inference statistics are only available for the coefficients without shrinkage applied. They would be biased otherwise. Only if cross-fold validation is used in the `glmnetSE` model, the selected performance metric is displayed. If test data is supplied the performance metric on the train as test data is displayed.

Usage

```
## S3 method for class 'glmnetSE'  
summary(object, ...)
```

Arguments

<code>object</code>	A model of the class <code>glmnetSE</code> .
<code>...</code>	Additional arguments affecting the summary produced.

Value

The output of a `glmnetSE` object and the performance metric if cross-fold validation is used.

Examples

```
# Estimate model  
  
glmnetSE.model <- glmnetSE(data=swiss,cf.no.shrnkg = c("Education"), ncore = 2)  
  
# Display model output with summary  
  
summary(glmnetSE.model)
```

Index

- * **ROC**
 - plot.glmnetSE, 4
 - * **bootstrap**
 - glmnetSE, 2
 - * **errors**
 - glmnetSE, 2
 - * **glmnetSE**
 - plot.glmnetSE, 4
 - summary.glmnetSE, 5
 - * **glmnet**
 - glmnetSE, 2
 - * **output**
 - summary.glmnetSE, 5
 - * **plot**
 - plot.glmnetSE, 4
 - * **results**
 - summary.glmnetSE, 5
 - * **shrinkage**
 - glmnetSE, 2
 - * **standard**
 - glmnetSE, 2
 - * **summary**
 - summary.glmnetSE, 5
- boot.ci, 3
- cv.glmnet, 2
- glmnet, 2
- glmnetSE, 2, 4, 6
- plot.glmnetSE, 4, 4
- summary.glmnetSE, 4, 5