Package ‘bootStepAIC’

February 19, 2015

Title Bootstrap stepAIC
Version 1.2-0
Date 2009-06-04
Author Dimitris Rizopoulos <d.rizopoulos@erasmusmc.nl>
Maintainer Dimitris Rizopoulos <d.rizopoulos@erasmusmc.nl>
Description Model selection by bootstrapping the stepAIC() procedure.
Depends R(>= 2.3.0), MASS
LazyLoad yes
LazyData yes
License GPL (>= 2)
Repository CRAN
Date/Publication 2009-06-04 16:11:54
NeedsCompilation no

R topics documented:

boot.stepAIC ......................................................... 1

Index 5

boot.stepAIC    Bootstraps the Stepwise Algorithm of stepAIC() for Choosing a Model by AIC

Description

Implements a Bootstrap procedure to investigate the variability of model selection under the stepAIC() stepwise algorithm of package MASS.
Usage

\texttt{boot.stepAIC(object, data, B = 100, alpha = 0.05, direction = "backward", k = 2, verbose = FALSE, ...)}

Arguments

- **object**: an object representing a model of an appropriate class; currently, \texttt{"lm"}, \texttt{"aov"}, \texttt{"glm"}, \texttt{"negbin"}, \texttt{"polr"}, \texttt{"survreg"}, and \texttt{"coxph"} objects are supported.
- **data**: a data.frame or a matrix that contains the response variable and covariates.
- **B**: the number of Bootstrap samples.
- **alpha**: the significance level.
- **direction**: the direction argument of \texttt{stepAIC()}.
- **k**: the \texttt{k} argument of \texttt{stepAIC()}.
- **verbose**: logical; if TRUE information about the evolution of the procedure is printed in the screen.
- **...**: extra arguments to \texttt{stepAIC()}, e.g., \texttt{scope}.

Details

The following procedure is replicated \(B\) times:

**Step 1**: Simulate a new data-set taking a sample with replacement from the rows of \texttt{data}.

**Step 2**: Refit the model using the data-set from Step 1.

**Step 3**: For the refitted model of Step 2 run the \texttt{stepAIC()} algorithm.

Summarize the results by counting how many times (out of the \(B\) data-sets) each variable was selected, how many times the estimate of the regression coefficient of each variable (out of the times it was selected) it was statistically significant in significance level \(alpha\), and how many times the estimate of the regression coefficient of each variable (out of the times it was selected) changed signs (see also Austin and Tu, 2004).

Value

An object of class \texttt{BootStep} with components

- **Covariates**: a numeric matrix containing the percentage of times each variable was selected.
- **Sign**: a numeric matrix containing the percentage of times the regression coefficient of each variable had sign + and −.
- **Significance**: a numeric matrix containing the percentage of times the regression coefficient of each variable was significant under the \(alpha\) significance level.
- **OrigModel**: a copy of \texttt{object}.
- **OrigStepAIC**: the result of applying \texttt{stepAIC()} in \texttt{object}.
- **direction**: a copy of the \texttt{direction} argument.
- **k**: a copy of the \texttt{k} argument.
- **BootStepAIC**: a list of length \(B\) containing the results of \texttt{stepAIC()} for each Bootstrap data-set.
**Author(s)**

Dimitris Rizopoulos <d.rizopoulos@erasmusmc.nl>

**References**


**See Also**

`stepAIC` in package MASS

**Examples**

```r
## lm() Example ##
N <- 350
x1 <- runif(N, -4, 4)
x2 <- runif(N, -4, 4)
x3 <- runif(N, -4, 4)
x4 <- runif(N, -4, 4)
x5 <- runif(N, -4, 4)
x6 <- runif(N, -4, 4)
x7 <- factor(sample(sample(letters[1:3], N, rep = TRUE))
Y <- 5 + 3 * x1 + 2 * x2 - 1.5 * x3 - 0.8 * x4 + rnorm(N, sd = 2.5)
data <- data.frame(y, x1, x2, x3, x4, x5, x6, x7)
rm(N, x1, x2, x3, x4, x5, x6, x7)

lmFit <- lm(y ~ (x1 + x2 + x3 + x4 + x5 + x6 + x7), data = data)
boot.stepAIC(lmFit, data)

## glm() Example ##
N <- 200
x1 <- runif(N, -3, 3)
x2 <- runif(N, -3, 3)
x3 <- runif(N, -3, 3)
x4 <- runif(N, -3, 3)
x5 <- factor(sample(sample(letters[1:2], N, rep = TRUE))
eta <- 0.1 + 1.6 * x1 - 2.5 * as.numeric(as.character(x5) == levels(x5)[1])
y1 <- rbinom(N, 1, plogis(eta))
y2 <- rbinom(N, 1, 0.6)
data <- data.frame(y1, y2, x1, x2, x3, x4, x5)
rm(N, x1, x2, x3, x4, x5, eta, y1, y2)

glmFit1 <- glm(y1 ~ x1 + x2 + x3 + x4 + x5, family = binomial, data = data)
glmFit2 <- glm(y2 ~ x1 + x2 + x3 + x4 + x5, family = binomial, data = data)
```

boot.stepAIC(glmFit1, data, B = 50)
boot.stepAIC(glmFit2, data, B = 50)
Index

*Topic regression
  boot.stepAIC, 1

boot.stepAIC, 1