Package 'abe'

October 30, 2017

Type Package

Title Augmented Backward Elimination

Version 3.0.1

Date 2017-10-25

Author Rok Blagus [aut, cre], Sladana Babic [ctb]

Maintainer Rok Blagus <rok.blagus@mf.uni-lj.si>

Description Performs augmented backward elimination and checks the stability of the obtained model. Augmented backward elimination combines significance or information based criteria with the change in estimate to either select the optimal model for prediction purposes or to serve as a tool to obtain a practically sound, highly interpretable model. More details can be found in Dunkler et al. (2014) <doi:10.1371/journal.pone.0113677>.

License GPL (>= 2)

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2017-10-30 09:14:07 UTC

R topics documented:

1

abe	*Augmented Backward Elimination*

Description

Function ″abe″ performs Augmented backward elimination where variable selection is based on the change-in-estimate and significance or information criteria. It can also make a backward-selection based on significance or information criteria only by turning off the change-in-estimate criterion.

Usage

```
abe(fit, data = NULL, include = NULL, active = NULL, tau = 0.05,
  exp.beta = TRUE, exact = FALSE, criterion = "alpha", alpha = 0.2,
  type.test = "Chisq", type.factor = NULL, verbose = T)
```

Arguments

fit	An object of a class ″lm″, ″glm″ or ″coxph″ representing the fit. Note, the functions should be fitted with argument x=TRUE and y=TRUE.
data	data frame used when fitting the object `fit`.
include	a vector containing the names of variables that will be included in the final model. These variables are used as only passive variables during modeling. These variables might be exposure variables of interest or known confounders. They will never be dropped from the working model in the selection process, but they will be used passively in evaluating change-in-estimate criteria of other variables. Note, variables which are not specified as include or active in the model fit are assumed to be active and passive variables.
active	a vector containing the names of active variables. These less important explanatory variables will only be used as active, but not as passive variables when evaluating the change-in-estimate criterion.
tau	Value that specifies the threshold of the relative change-in-estimate criterion. Default is set to 0.05.
exp.beta	Logical specifying if exponent is used in formula to standardize the criterion. Default is set to TRUE.
exact	Logical, specifies if the method will use exact change-in-estimate or its approximation. Default is set to FALSE, which means that the method will use approximation proposed by Dunkler et al. Note, setting to TRUE can severely slow down the algorithm, but setting to FALSE can in some cases lead to a poor approximation of the change-in-estimate criterion.
criterion	String that specifies the strategy to select variables for the black list. Currently supported options are significance level `'alpha'`, Akaike information criterion `'AIC'` and Bayesian information criterion `'BIC'`. If you are using significance level, in that case you have to specify the value of 'alpha' (see parameter `alpha`) and type of the test statistic (see parameter `type.test`). Default is set to ″alpha″.

alpha	Value that specifies the level of significance as explained above. Default is set to 0.2.
type.test	String that specifies which test should be performed in case the criterion = "alpha". Possible values are "F" and "Chisq" (default) for class "lm", "Rao", "LRT", "Chisq" (default), "F" for class "glm" and "Chisq" for class "coxph". See also [drop1](#).
type.factor	String that specifies how to treat factors, see details, possible values are "factor" and "individual".
verbose	Logical that specifies if the variable selection process should be printed. Note: this can severely slow down the algorithm.

Details

Using the default settings ABE will perform augmented backward elimination based on significance. The level of significance will be set to 0.2. All variables will be treated as "passive or active". Approximated change-in-estimate will be used. Threshold of the relative change-in-estimate criterion will be 0.05. Setting tau to a very large number (e.g. Inf) turns off the change-in-estimate criterion, and ABE will only perform backward elimination. Specifying "alpha" = 0 will include variables only because of the change-in-estimate criterion, as then variables are not safe from exclusion because of their p-values. Specifying "alpha" = 1 will always include all variables.

When using type.factor="individual" each dummy variable of a factor is treated as an individual explanatory variable, hence only this dummy variable can be removed from the model (warning: use sensible coding for the reference group). Using type.factor="factor" will look at the significance of removing all dummy variables of the factor and can drop the entire variable from the model.

Value

An object of class "lm", "glm" or "coxph" representing the model chosen by abe method.

Author(s)

Rok Blagus, <rok.blagus@mf.uni-lj.si>

Sladana Babic

References

Daniela Dunkler, Max Plischke, Karen Lefondre, and Georg Heinze. Augmented backward elimination: a pragmatic and purposeful way to develop statistical models. PloS one, 9(11):e113677, 2014.

See Also

[abe.boot](#), [lm](#), [glm](#) and [coxph](#)

Examples

```
# simulate some data:

set.seed(1)
n=100
x1<-runif(n)
x2<-runif(n)
x3<-runif(n)
y<--5+5*x1+5*x2+ rnorm(n,sd=5)
dd<-data.frame(y,x1,x2,x3)

# fit a simple model containing only numeric covariates
fit<-lm(y~x1+x2+x3,x=TRUE,y=TRUE,data=dd)

# perform ABE with "x1" as only passive and "x2" as only active
# using the exact change in the estimate of 5% and significance
# using 0.2 as a threshold
abe.fit<-abe(fit,data=dd,include="x1",active="x2",
tau=0.05,exp.beta=FALSE,exact=TRUE,criterion="alpha",alpha=0.2,
type.test="Chisq",verbose=TRUE)

summary(abe.fit)

# similar example, but turn off the change-in-estimate and perform
# only backward elimination

abe.fit<-abe(fit,data=dd,include="x1",active="x2",
tau=Inf,exp.beta=FALSE,exact=TRUE,criterion="alpha",alpha=0.2,
type.test="Chisq",verbose=TRUE)

summary(abe.fit)

# an example with the model containing categorical covariates:
dd$x3<-rbinom(n,size=3,prob=1/3)
dd$y1<--5+5*x1+5*x2+ rnorm(n,sd=5)
fit<-lm(y1~x1+x2+factor(x3),x=TRUE,y=TRUE,data=dd)

# treat "x3" as a single covariate:

abe.fit.fact<-abe(fit,data=dd,include="x1",active="x2",
tau=0.05,exp.beta=FALSE,exact=TRUE,criterion="alpha",alpha=0.2,
type.test="Chisq",verbose=TRUE,type.factor="factor")

summary(abe.fit.fact)

# treat each dummy of "x3" as a separate covariate:

abe.fit.ind<-abe(fit,data=dd,include="x1",active="x2",
tau=0.05,exp.beta=FALSE,exact=TRUE,criterion="alpha",alpha=0.2,
type.test="Chisq",verbose=TRUE,type.factor="individual")

summary(abe.fit.ind)
```

---

abe.boot *Bootstrapped Augmented Backward Elimination*

Description

Performs Augmented backward elimination on re-sampled datasets using different bootstrap and re-sampling techniques.

Usage

```
abe.boot(fit, data = NULL, include = NULL, active = NULL, tau = 0.05,
  exp.beta = TRUE, exact = FALSE, criterion = "alpha", alpha = 0.2,
  type.test = "Chisq", type.factor = NULL, num.boot = 100,
  type.boot = c("bootstrap", "mn.bootstrap", "subsampling"),
  prop.sampling = 0.632)
```

Arguments

fit	An object of a class "lm", "glm" or "coxph" representing the fit. Note, the functions should be fitted with argument x=TRUE and y=TRUE.
data	data frame used when fitting the object fit.
include	a vector containing the names of variables that will be included in the final model. These variables are used as passive variables during modeling. These variables might be exposure variables of interest or known confounders. They will never be dropped from the working model in the selection process, but they will be used passively in evaluating change-in-estimate criteria of other variables. Note, variables which are not specified as include or active in the model fit are assumed to be active and passive variables.
active	a vector containing the names of active variables. These less important explanatory variables will only be used as active, but not as passive variables when evaluating the change-in-estimate criterion.
tau	Value that specifies the threshold of the relative change-in-estimate criterion. Default is set to 0.05.
exp.beta	Logical specifying if exponent is used in formula to standardize the criterion. Default is set to TRUE.
exact	Logical, specifies if the method will use exact change-in-estimate or approximated. Default is set to FALSE, which means that the method will use approximation proposed by Dunkler et al. Note, setting to TRUE can severely slow down the algorithm, but setting to FALSE can in some cases lead to a poor approximation of the change-in-estimate criterion.
criterion	String that specifies the strategy to select variables for the blacklist. Currently supported options are significance level 'alpha', Akaike information criterion 'AIC' and Bayesian information criterion 'BIC'. If you are using significance level, in that case you have to specify the value of 'alpha' (see parameter alpha). Default is set to "alpha".

alpha	Value that specifies the level of significance as explained above. Default is set to 0.2.
type.test	String that specifies which test should be performed in case the `criterion = "alpha"`. Possible values are `"F"` and `"Chisq"` (default) for class `"lm"`, `"Rao"`, `"LRT"`, `"Chisq"` (default), `"F"` for class `"glm"` and `"Chisq"` for class `"coxph"`. See also [drop1](#).
type.factor	String that specifies how to treat factors, see details, possible values are `"factor"` and `"individual"`.
num.boot	number of bootstrap re-samples
type.boot	String that specifies the type of bootstrap. Possible values are `"bootstrap"`, `"mn.bootstrap"`, `"subsampling"`, see details
prop.sampling	Sampling proportion. Only applicable for `type.boot="mn.bootstrap"` and `type.boot="subsampling"`, defaults to 0.632. See details.

Details

type.boot can be bootstrap (n observations drawn from the original data with replacement), mn.bootstrap (m out of n observations drawn from the original data with replacement), subsampling (m out of n observations drawn from the original data without replacement), where m is [prop.sampling*n].

Value

an object of class abe for which summary and plot functions are available. A list with the following elements:

models the final models obtained after performing ABE on re-sampled datasets, each object in the list is of the same class as fit

alpha the vector of significance levels used

tau the vector of threshold values for the change-in-estimate

num.boot number of re-sampled datasets

criterion criterion used when constructing the black-list

all.vars a list of variables used when estimating fit

fit.or the initial model

Author(s)

Rok Blagus, <rok.blagus@mf.uni-lj.si>

Sladana Babic

References

Daniela Dunkler, Max Plischke, Karen Lefondre, and Georg Heinze. Augmented backward elimination: a pragmatic and purposeful way to develop statistical models. PloS one, 9(11):e113677, 2014.

Riccardo De Bin, Silke Janitza, Willi Sauerbrei and Anne-Laure Boulesteix. Subsampling versus Bootstrapping in Resampling-Based Model Selection for Multivariable Regression. Biometrics 72, 272-280, 2016.

See Also

abe, summary.abe, plot.abe

Examples

```
# simulate some data and fit a model

set.seed(1)
n=100
x1<-runif(n)
x2<-runif(n)
x3<-runif(n)
y<--5+5*x1+5*x2+ rnorm(n,sd=5)
dd<-data.frame(y=y,x1=x1,x2=x2,x3=x3)
fit<-lm(y~x1+x2+x3,x=TRUE,y=TRUE,data=dd)

# use ABE on 50 bootstrap re-samples considering different
# change-in-estimate thresholds and significance levels

fit.boot<-abe.boot(fit,data=dd,include="x1",active="x2",
tau=c(0.05,0.1),exp.beta=FALSE,exact=TRUE,
criterion="alpha",alpha=c(0.2,0.05),type.test="Chisq",
num.boot=50,type.boot="bootstrap")

summary(fit.boot)

# use ABE on 50 subsamples randomly selecting 50% of subjects
# considering different change-in-estimate thresholds and
# significance levels

fit.boot<-abe.boot(fit,data=dd,include="x1",active="x2",
tau=c(0.05,0.1),exp.beta=FALSE,exact=TRUE,
criterion="alpha",alpha=c(0.2,0.05),type.test="Chisq",
num.boot=50,type.boot="subsampling",prop.sampling=0.5)

summary(fit.boot)
```

---

plot.abe   *Plot Function*

Description

Plot function for the bootstrapped version of ABE.

Usage

```
## S3 method for class 'abe'
plot(x, type.plot = c("coefficients", "models", "variables"),
  alpha = NULL, tau = NULL, variable = NULL, ...)
```

Arguments

x	an object of class "abe", an object returned by a call to abe.boot
type.plot	string which specifies the type of the plot. See details.
alpha	values of alpha for which the plot is to be made (can be a vector of length >1)
tau	values of tau for which the plot is to be made (can be a vector of length >1)
variable	variables for which the plot is to be made (can be a vector of length >1)
...	Arguments to be passed to methods, such as graphical parameters (see barplot, hist).

Details

when using type.plot="coefficients" the function plots a histogram of the estimated regression coefficients for the specified variables, alpha(s) and tau(s) obtained from different re-sampled datasets. When the variable is not included in the final model, its regression coefficient is set to zero. When using type.plot="variables" the function plots a barplot of the relative inclusion frequencies of the specified variables, for the specified values of alpha and tau. When using type.plot="models" the function plots a barplot of the relative frequencies of the final models for specified alpha(s) and tau(s).

Author(s)

Rok Blagus, <rok.blagus@mf.uni-lj.si>

Sladana Babic

See Also

abe.boot, summary.abe

Examples

```
set.seed(1)
n=100
x1<-runif(n)
x2<-runif(n)
x3<-runif(n)
y<--5+5*x1+5*x2+ rnorm(n,sd=5)
dd<-data.frame(y=y,x1=x1,x2=x2,x3=x3)
fit<-lm(y~x1+x2+x3,x=TRUE,y=TRUE,data=dd)

fit.boot<-abe.boot(fit,data=dd,include="x1",active="x2",
tau=c(0.05,0.1),exp.beta=FALSE,exact=TRUE,
criterion="alpha",alpha=c(0.2,0.05),type.test="Chisq",
num.boot=50,type.boot="bootstrap")

plot(fit.boot,type.plot="coefficients",
alpha=0.2,tau=0.1,variable=c("x1","x3"),
col="light blue")

plot(fit.boot,type.plot="variables",
```

```
alpha=0.2,tau=0.1,variable=c("x1","x2","x3"),
col="light blue",horiz=TRUE,las=1)

par(mar=c(4,6,4,2))
plot(fit.boot,type.plot="models",
alpha=0.2,tau=0.1,col="light blue",horiz=TRUE,las=1)
```

---

summary.abe   *Summary Function*

Description

makes a summary of a bootstrapped version of ABE

Usage

```
## S3 method for class 'abe'
summary(object, conf.level = 0.95, ...)
```

Arguments

object	an object of class "abe", an object returned by a call to abe.boot
conf.level	the confidence level, defaults to 0.95
...	additional arguments affecting the summary produced.

Value

a list with the following elements:

var.rel.frequencies: inclusion relative frequencies for all variables from the initial model

model.rel.frequencies: relative frequencies of the final models

var.coefs: bootstrap medians and percentiles for the estimates of the regression coefficients for each variable from the initial model

Author(s)

Rok Blagus, <rok.blagus@mf.uni-lj.si>

Sladana Babic

See Also

abe.boot, plot.abe

Examples

```
set.seed(1)
n=100
x1<-runif(n)
x2<-runif(n)
x3<-runif(n)
y<--5+5*x1+5*x2+ rnorm(n,sd=5)
dd<-data.frame(y=y,x1=x1,x2=x2,x3=x3)
fit<-lm(y~x1+x2+x3,x=TRUE,y=TRUE,data=dd)

fit.boot<-abe.boot(fit,data=dd,include="x1",active="x2",
tau=c(0.05,0.1),exp.beta=FALSE,exact=TRUE,
criterion="alpha",alpha=c(0.2,0.05),type.test="Chisq",
num.boot=50,type.boot="bootstrap")

summary(fit.boot)$var.rel.frequencies
```

Index