Package 'globalboosttest'

February 19, 2015

Version 1.1-0

Date 2010-08-02

Title Testing the additional predictive value of high-dimensional data

Author Anne-Laure Boulesteix <boulesteix@ibe.med.uni-muenchen.de>,
Torsten Hothorn <torsten.hothorn@stat.uni-muenchen.de>.

Maintainer Anne-Laure Boulesteix <boulesteix@ibe.med.uni-muenchen.de>

Depends R (>= 2.8), mboost (>= 2.0-0), survival

Suggests

Description 'globalboosttest' implements a permutation-based testing
procedure to globally test the (additional) predictive value of
a large set of predictors given that a small set of predictors
is already available. Currently, 'globalboosttest' supports
binary outcomes (via logistic regression) and survival outcomes
(via Cox regression). It is based on boosting regression as
implemented in the package 'mboost'.

License GPL (>= 2)

URL

Repository CRAN

Repository/R-Forge/Project globalboosttest

Repository/R-Forge/Revision 4

Date/Publication 2012-10-29 08:58:54

NeedsCompilation no

R topics documented:

---

globalboosttest   *Testing the additional predictive value of high-dimensional data*

Description

The function globalboosttest implements a permutation-based testing procedure to globally test the (additional) predictive value of a large set of predictors given that a small set of predictors is already available.

Usage

```
globalboosttest(X,Y,Z=NULL,nperm=1000,mstop=1000,mstopAIC=FALSE,pvalueonly=TRUE,plot=FALSE,...)
```

Arguments

X	A n x p matrix or data frame with observations in rows and variables in columns, whose additional predictive value has to be tested.
Y	Either a n-vector of type factor (if the prediction outcome is binary), or a numeric vector of length n (if the prediction outcome is numeric and uncensored), or a Surv object (if the prediction outcome is a survival time).
Z	A n x q matrix or data frame with observations in rows and variables in columns, on which we want to condition. Note that q should be smaller than n. If Z=NULL, the function globalboosttest simply assesses the predictive value of X without conditioning.
nperm	The number of permutations used to derived the p-value.
mstop	A numeric vector giving the number(s) of boosting steps at which the p-value has to be calculated.
mstopAIC	If TRUE, the best number of boosting steps is determined based on AIC using the non-permuted data from the range 1:max(mstop).
pvalueonly	Should the function return only the permutation p-value or also the risk for all numbers of boosting steps and all permutations?
plot	If TRUE, a plot representing the minimized criterion for real data (in black) and permuted data (in grey).
...	Further arguments to be passed to the plot function if plot=TRUE.

Details

See Boulesteix and Hothorn (2009) for details on the methodology. If mstopAIC=TRUE, the number of boosting steps is chosen from 1 to max(mstop) independently of the specific values included in the vector mstop.

Value

A list with the following arguments

riskreal	A numeric vector of length max(mstop) giving the risk computed from the original data set with mstop from 1 to max(mstop) (if pvalueonly=FALSE).
riskperm	A npermxmax(mstop) matrix giving the risk computed from the nperm permuted data sets with mstop from 1 to max(mstop) (if pvalueonly=FALSE).
mstopAIC	The number of boosting steps selected using the AIC-based procedure (if mstopAIC=TRUE).
pvalue	A numeric vector of length length(mstop) (if mstopAIC=FALSE) or length(mstop)+1 (if mstopAIC=TRUE) giving the permutation-pvalues obtained for each considered value of mstop

Author(s)

Anne-Laure Boulesteix (http://www.ibe.med.uni-muenchen.de/organisation/mitarbeiter/020_professuren/boulesteix/eng.html),

Torsten Hothorn (http://www.statistik.lmu.de/~hothorn/)

References

A. L. Boulesteix and Torsten Hothorn (2010). Testing the additional predictive value of high-dimensional data. BMC Bioinformatics 10:78.

Examples

```
# load globalboosttest library
library(globalboosttest)

# load the simulated data with binary outcome
data(simdatabin)
attach(simdatabin)
# Test with 25 permutations
test<-globalboosttest(X=X,Y=Y,Z=Z,nperm=25,mstop=c(100,500,1000))


# load the simulated data with survival outcome
data(simdatasurv)
attach(simdatasurv)
# Test with 25 permutations
test<-globalboosttest(X=X,Y=Surv(time,status),Z=NULL,nperm=25,mstop=c(100,500,1000),mstopAIC=FALSE)
```

| simdatabin | *Simulated data with binary outcome* |

Description

Simulated data with binary outcome Y. The data X and Z are non-informative in the sense that they were randomly generated independently of Y.

Usage

```
data(simdatabin)
```

Details

Simulated data for demonstration purpose only.

Value

A list with the following arguments:

X	A 50 x 200 matrix (with observations in rows and variables in columns) whose additional predictive value has to be tested.
Y	A vector of length 50 and type factor
Z	A 50 x 5 matrix (with observations in rows and variables in columns) on which we want to condition.

Author(s)

Anne-Laure Boulesteix (http://www.ibe.med.uni-muenchen.de/organisation/mitarbeiter/020_professuren/boulesteix/eng.html)

See Also

simdatasurv

Examples

```
# load globalboosttest library
library(globalboosttest)

# load the simulated data with binary outcome
data(simdatabin)
# Visualize the dimensions
dim(simdatabin$X)
dim(simdatabin$Z)
simdatabin$Y
```

---

simdatasurv   *Simulated data with survival outcome*

Description

Simulated data with binary outcome Y. The data X and Z are non-informative in the sense that they were randomly generated independently of Y.

Usage

```
data(simdatasurv)
```

Details

Simulated data for demonstration purpose only.

Value

A list with the following arguments:

X	A 50 x 200 matrix (with observations in rows and variables in columns) whose additional predictive value has to be tested.
status	A numeric vector of length 50 giving the survival status with values 0 (alive) and 1 (dead).
time	A numeric vector of length 50 giving the observation time (which equals the survival time for subjects with status=1)
Z	A 50 x 5 matrix (with observations in rows and variables in columns) on which we want to condition.

Author(s)

Anne-Laure Boulesteix (http://www.ibe.med.uni-muenchen.de/organisation/mitarbeiter/020_professuren/boulesteix/eng.html),

See Also

simdatabin

Examples

```
# load globalboosttest library
library(globalboosttest)

# load the simulated data with survival outcome
data(simdatasurv)
# Visualize the dimensions
dim(simdatasurv$X)
dim(simdatasurv$Z)
```

```
simdatasurv$status
simdatasurv$time
```

Index