

Package ‘mcb’

June 5, 2020

Type Package

Title Model Confidence Bounds

Version 0.1.15

Author Yang Li, Yichen Qin, Heming Deng

Maintainer Heming Deng<dheming@ruc.edu.cn>

Description When choosing proper variable selection methods, it is important to consider the uncertainty of a certain method. The model confidence bound for variable selection identifies two nested models (upper and lower confidence bound models) containing the true model at a given confidence level. A good variable selection method is the one of which the model confidence bound under a certain confidence level has the shortest width. When visualizing the variability of model selection and comparing different model selection procedures, model uncertainty curve is a good graphical tool. A good variable selection method is the one of whose model uncertainty curve will tend to arch towards the upper left corner. This function aims to obtain the model confidence bound and draw the model uncertainty curve of certain single model selection method under a coverage rate equal or little higher than user-given confidential level. About what model confidence bound is and how it work please see Li,Y., Luo,Y., Ferrari,D., Hu,X. and Qin,Y. (2019) Model Confidence Bounds for Variable Selection. *Biometrics*, 75:392-403. <DOI:10.1111/biom.13024>. Besides, 'flare' is needed only you apply the SQRT or LAD method ('mcb' totally has 8 methods). Although 'flare' has been archived by CRAN, you can still get it in <<https://CRAN.R-project.org/package=flare>> and the latest version is useful for 'mcb'.

License GPL (>= 2)

Encoding UTF-8

LazyData true

Imports parallel,methods,leaps,lars,MASS,glmnet,ncvreg,smoothmest,ggplot2,reshape2

Depends R (>= 3.6.0)

Suggests flare, testthat

NeedsCompilation no

Repository CRAN

Date/Publication 2020-06-05 13:00:02 UTC

R topics documented:

Diabetes	2
mcb	2
mcb.compare	4

Index	6
--------------	----------

Diabetes	<i>Diabetes</i>
----------	-----------------

Description

This diabetes data set has $n = 352$ samples and there are $p = 6$ predictors: lamotrigine (ltg), total serum cholesterol (tc), total cholesterol (tch), low- and high-density lipoprotein (ldl and hdl) and glucose (glu). The response variable is the measurement of the disease progression one year after baseline.

Usage

Diabetes

Format

A dataframe containing 352 records

References

B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Annals of statistics*, 32(2):407–499, 2004.

mcb	<i>Model Confidence Bound</i>
-----	-------------------------------

Description

When choosing proper variable selection methods, it is important to consider the uncertainty of a certain method. The MCB for variable selection identifies two nested models (upper and lower confidence bound models) containing the true model at a given confidence level. A good variable selection method is the one of which the MCB under a certain confidence level has the shortest width. When visualizing the variability of model selection and comparing different model selection procedures, Model uncertainty curve is a good graphical tool. A good variable selection method is the one of whose MUC will tend to arch towards the upper left corner. This function aims to obtain the MCB and draw the MUC of certain single model selection method under a coverage rate equal or little higher than user-given confidential level.

Usage

```
mcb(x, y, B=200, lambda=NA, method='Lasso', level=0.95, seed=122)
```

Arguments

x	input matrix; each column is an observation vector of certain independent variable, and will be given a name automatically in the order of x1, x2, x3...
y	y is a matrix of one column which presents the response vector B number of bootstrap replicates to perform, default value is 200.
B	number of bootstrap replicates to perform; Default value is 200.
lambda	A user supplied lambda value. It is the penalty tuning parameter for the variable selection method tested. The default value is the optimization outcome automatically computed in consideration of the specific case.
method	Default value is 'Lasso'; user can choose from 'aLasso', 'Lasso', 'SCAD', 'MCP', 'stepwise', 'LAD', 'SQRT'
level	a positive value between 0 and 1, like the concept of confidence level for point estimation; Default value is 0.95
seed	seed for bootstrap procedures; Default value is 122;

Value

The mcb method returns an object of class "mcb" The generic accessor functions mcb, mucplot and mcbframe extract various useful features of the value returned by mcb. An object of class "mcb" is a list containing at least the following components:

mcb	a list containing the bootstrap coverage rate (which is the closest to the user-given confidence level) and the corresponding model confidence bound of the user-chosen variable selection method in the form of lower confidence bound and upper confidence bound.
mucplot	plot of the model uncertainty curve for this specific user-chosen variable selectionmethod.
mcbframe	a dataframe containing all the information about MCBs for the specific variable selectionmethod under all bootstrap coverage rates including width(w), lower confidence bound(lcb) and upper confidence bound(ucb) for each bootstrap coverage rate(bcr)

References

Li,Y., Luo,Y., Ferrari,D., Hu,X. and Qin,Y. (2019) Model Confidence Bounds for Variable Selection. *Biometrics*, 75:392-403.

Examples

```
data(Diabetes) # load data
x <- Diabetes[,c('S1', 'S2', 'S3', 'S4', 'S5')]
y <- Diabetes[,c('Y')]
x <- data.matrix(x)
```

```

y <- data.matrix(y)
result <- mcb(x=x, y=y)
# plot of the model uncertainty curve
result$mucplot
# a list containing the bootstrap coverage rate and mcb
result$mcb
# a dataframe containing all the information about MCBs
result$mcbframe

```

mcb.compare	<i>Comparisons of Model Confidence Bounds for Different Variable selection Methods</i>
-------------	--

Description

This function is a supplement of the function `mcb`. It is used to compare different variable selection methods and would return all the MUCs on same canvas. A good variable selection method's MUC will tend to arch towards the upper left corner.

Usage

```
mcb.compare(x, y, B=200, lambdas=NA, methods=NA, level=0.95, seed=122)
```

Arguments

<code>x</code>	input matrix presenting independent variables as in <code>mcb</code> .
<code>y</code>	response vector as in <code>mcb</code> .
<code>B</code>	number of bootstrap replicates to perform; Default value is 200.
<code>lambdas</code>	A vector of penalty tuning parameters for each variable selection method. The default values are the optimal choices for each selection method computed automatically.
<code>methods</code>	a vector including all variable selection methods the user wants to test and compare. The default value is <code>c('aLasso', 'Lasso', 'SCAD', 'MCP', 'stepwise', 'LAD', 'SQRT')</code>
<code>level</code>	user-defined confidence level as in <code>mcb</code> ; Default value is 0.95.
<code>seed</code>	Default value is 122.

Value

The `mcb.compare` method returns an object of class "mcb.compare". An object of class "mcb.compare" is a list containing at least the following components:

<code>mcb</code>	a list containing the bootstrap coverage rate and the corresponding model confidence bound for all user-given variable selection methods.
<code>mucplot</code>	plot of the model uncertainty curves for all variable selection methods and could be used to choose the best method.
<code>mcbframe</code>	a list containing all the information about MCBs for all variable selection methods under all available bootstrap coverage rates.

References

Li,Y., Luo,Y., Ferrari,D., Hu,X. and Qin,Y. (2019) Model Confidence Bounds for Variable Selection. *Biometrics*, 75:392-403.

Examples

```
data(Diabetes) # load data
x <- Diabetes[,c('S1', 'S2', 'S3', 'S4', 'S5')]
y <- Diabetes[,c('Y')]
x <- data.matrix(x)
y <- data.matrix(y)
result <- mcb.compare(x=x, y=y)
# plot of the model uncertainty curves for all variable selection methods
result$mucplot
# a list containing the bootstrap coverage rate and mcb which based on Lasso
result$mcb$Lasso
# a dataframe containing all the information about MCBs which based on Lasso
result$mcbframe$Lasso
```

Index

Diabetes, [2](#)

mcb, [2](#)

mcb.compare, [4](#)