

Package ‘BISSpkg’

May 17, 2016

Type Package

Title Bootstrap Imputation with Variable Selection

Version 1.0

Date 2016-03-16

Author Zhifan Sang, Yi Deng, and Qi Long

Maintainer Zhifan Sang <zsang@emory.edu>

Description Bootstrap imputation through direct use of stability selection, as well as Bayesian Lasso Regression. Built-in imputation models are provided for continuous data (normal linear regression), binary data (logistic regression) and poisson data (log-linear regression).

Depends R (>= 3.2.0),
MIHD,
MASS,
mice,
parcor,
blasso,
glmnet

License GPL-2|GPL-3

RoxygenNote 5.0.1

LazyData true

R topics documented:

| | |
|--------------------------|---|
| BI_SS | 1 |
| detect.missing | 3 |

| | |
|--------------|----------|
| Index | 4 |
|--------------|----------|

| | |
|-------|--|
| BI_SS | <i>Bootstrap Imputation and Variable Selection</i> |
|-------|--|

Description

This function deploys bootstrap imputation and variable selection on (X,y) data with missing values

Usage

```
BI_SS(X, Y, family, link = NULL, missing_col, MI.method, BI.size, pi, nsteps)
```

Arguments

| | |
|-------------|---|
| X | a n x p matrix of n observation and p predictors, where NAs present as missing values |
| Y | a n x 1 vector represent the outcome parameter, esponse variable. Quantitative for family="gaussian", or family="poisson" (non-negative counts). For family="binomial" should be either a factor with two levels, or a two-column matrix of counts or proportions (the second column is treated as the target class; for a factor, the last level in alphabetical order is the target class). For "binomial" case, if y is presented as a vector, it will be coerced into a factor. |
| family | of regression, could be one of the following values: "gaussian", "binomial", "poisson") |
| missing_col | index of the missing column in X (only one column is missing in this case) |
| MI.method | the type of imputation method. Can be "mice", or "balasso". |
| BI.size | number of bootstrap imputed datasets |
| pi | threshold value for adaptive lasso variable selection |
| nsteps | number of steps used in stability selection |
| reg.type | the type of regression used in the second step, could be lasso, lasso.adp |

Details

This function first performs the bootstrap imputation based on mice method(mice package) and balasso method(MIHD) method on data consists of outcome parameter and predictors with missing values to obtain bootstrap imputed dataset. Then this function use lasso and stability selection with randomized lasso to conduct variable selection on the bootstrap imputed datasets, and obtain selection indicators. Lastly, this function calculate selected predictor estimates.

Value

a list of output values including Bootstrap imputed dataset(Bimp), Final variable selected indicator(S.fin), final parameter estimates for the regression(beta.fin), final selection probability for lasso (S1.prob), final selection for variable selection indicator for stability selection with randomized lasso(S2.prob)

Author(s)

Zhifan Sang

See Also

MIdurr

Examples

```
output_lr_1=BI_SS(lr_1[, -1], lr_1$Y, family="gaussian", link=NULL, missing_col=1, MI.method="mice", BI.size=100,
```

`detect.missing`*Finding the indicator matrix for missing values*

Description

This function develops a simple way to generate the indicator matrix for matrix with missing values

Usage

```
detect.missing(x)
```

Arguments

`x` a $n \times p$ matrix of n observation and p predictors, where NAs present as missing values

Details

This function runs NA search and replace on the input matrix, return a binary matrix of the same size. If the (i,j) element in the input matrix is NA, then the (i,j) element of output matrix is 1; Otherwise, the (i,j) element of the output matrix is 0

Value

a $n \times p$ binary indicator matrix where 1 indicates value is missing, 0 indicates value is not missing.

Author(s)

Zhifan Sang

See Also

`is.na`

Index

BI_SS, 1

detect.missing, 3