Package ‘mispr’

April 10, 2018

Type Package
Title Multiple Imputation with Sequential Penalized Regression
Version 1.0.0
Author Faisal Maqbool Zahid
Maintainer Faisal Maqbool Zahid <faisalmz99@yahoo.com>
Depends R (>= 3.3.0)
Imports e1071, MASS, penalized, stats
License GPL-2
Encoding UTF-8
LazyData true
RoxygenNote 6.0.1
NeedsCompilation no
Repository CRAN
Date/Publication 2018-04-10 13:24:44 UTC

### R topics documented:

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>data1</td>
<td>2</td>
</tr>
<tr>
<td>data2</td>
<td>2</td>
</tr>
<tr>
<td>mispr</td>
<td>3</td>
</tr>
</tbody>
</table>

Index 5
Simulated data with 50 covariates

**Description**

Data1 artificially generated dataframe with n=100 and p=50. Missing values using MAR (missing at random) mechanism are artificially generated in 10 covariates.

**Usage**

```r
data(data1)
```

**Format**

An object of class `data.frame` with 100 rows and 51 columns.

**Examples**

```r
data(data1)
```

Simulated data with 200 covariates

**Description**

Data2 artificially generated dataframe with n=100 and p=200. Missing values using MAR (missing at random) mechanism are artificially generated in 10 covariates.

**Usage**

```r
data(data2)
```

**Format**

An object of class `data.frame` with 100 rows and 201 columns.

**Examples**

```r
data(data2)
```
mispr

Multiple Imputation with Sequential Penalized Regression

Description
Generates Multivariate Imputations using sequential regression with L2 penalization.

Usage
mispr(x, x.select = FALSE, pen = FALSE, maxit = 5, m = 5,
      track = FALSE, init.method = "random", L2.fix = NULL, cv = TRUE,
      maxL2 = 2^10)

Arguments
x
A data frame or a matrix containing the incomplete data. Missing values are coded as NA.

x.select
A Boolean flag. If TRUE, linearly dependent columns will be removed before fitting of each imputation model. If FALSE, the linearly dependent columns will be removed only when number of predictors is greater than the sample size for fitting an imputation model. The default is FALSE.

pen
A Boolean flag. If TRUE, each imputation model will be fitted with L2 penalty. If FALSE, maximum likelihood estimation (MLE) will be used. However, if MLE fails, L2 penalty is used for fitting the imputation model. The default is FALSE.

maxit
A scalar giving the number of iterations. The default is 5.

m
Number of multiple imputations. The default is m=5.

track
A Boolean flag. If TRUE, mispr will print additional information about iterations on console. The default is FALSE for silent computation.

init.method
Method for initialization of missing values. random for filling NA in each column with a random sample from the observed values of that column. median for mean imputation.

L2.fix
Fixed value of ridge penalty (optional) to use for each imputation model. For default i.e., NULL, L2 penalty will be decided with k-fold cross-validation.

cv
A Boolean flag. If TRUE that is default, optimal value of L2 penalty will be decided independently for each imputation model using 5-fold cross-validation.

maxL2
The maximum value of the tuning parameter for L2 penalization to be used for optimizing the cross-validated likelihood. Default value is $2^{10}$.

Details
Generates multiple imputations for incomplete multivariate data by fitting a sequence of regression models using L2 penalty iteratively. Missing data can occur in one or more variables of the data. In each step of the iteration, ridge regression is fitted according to the distributional form of the missing variable taken as a response. All other variables are taken as predictors. If some predictors are incomplete, the most # recently generated imputations are used to complete the predictors before using them as a predictor.
Value

a list containing the number of imputed datasets, number of iterations used to obtain imputed data, list of multiply imputed datasets, and summary of missing values.

Author(s)

Faisal Maqbool Zahid <faisalmz99@yahoo.com>.

References


Examples

data(data1)
# Select a subset of data1
x=data1[, 1:10]
res1 = mispr(x)
# to get 3 multiply imputed datasets
res2 = mispr(x, m=3)
Index

*Topic datasets
  data1, 2
  data2, 2

data1, 2
data2, 2
mispr. 3