Package ‘midastouch’

February 7, 2016

Type Package
Version 1.3
Title Multiple Imputation by Distance Aided Donor Selection
Date 2016-02-06
Maintainer Philipp Gaffert <philipp.gaffert@web.de>
Depends R (>= 3.2.0)
Imports utils
Suggests mice
Description Contains the function mice.impute.midastouch(). Technically this function is to be run from within the 'mice' package (van Buuren et al. 2011), type ??mice. It substitutes the method 'pmm' within mice by 'midastouch'. The authors have shown that 'midastouch' is superior to default 'pmm'. Many ideas are based on Siddique / Belin 2008's MIDAS.
License GPL-2 | GPL-3
LazyLoad yes
LazyData yes
URL https://www.uni-bamberg.de/fileadmin/uni/fakultaeten/sowi_lehrstuehle/statistik/Personen/Dateien_Florian/properPMM.pdf
NeedsCompilation no
Author Philipp Gaffert [aut, cre],
Florian Meinfelder [aut],
Volker Bosch [aut]
Repository CRAN
Date/Publication 2016-02-07 09:35:46

R topics documented:
mice.impute.midastouch .............................................. 2

Index
mice.impute.midastouch

Predictive Mean Matching with distance aided selection of donors

Description

Imputes univariate missing data using predictive mean matching

Usage

mice.impute.midastouch(y, ry, x, ridge = 1e-05,
midas.kappa = NULL, outout = TRUE, neff = NULL, debug = NULL, ...)

Arguments

y Numeric vector with incomplete data
ry Response pattern of y (TRUE=observed, FALSE=missing)
x Design matrix with length(y) rows and p columns containing complete covariates.
ridge The ridge penalty applied to prevent problems with multicollinearity. The default is ridge = 1e-05, which means that 0.001 percent of the diagonal is added to the cross-product. Larger ridges may result in more biased estimates. For highly noisy data (e.g. many junk variables), set ridge = 1e-06 or even lower to reduce bias. For highly collinear data, set ridge = 1e-04 or higher.
midas.kappa Scalar. If NULL (default) then the optimal kappa gets selected automatically. Alternatively, the user may specify a scalar. Siddique and Belin 2008 find midas.kappa = 3 to be sensible.
outout Logical. If TRUE (default) one model is estimated for each donor (leave-one-out principle). For speedup choose outout = FALSE, which estimates one model for all observations leading to in-sample predictions for the donors and out-of-sample predictions for the recipients. Mind the inappropriateness, though.
neff FOR EXPERTS. Null or character string. The name of an existing environment in which the effective sample size of the donors for each loop (CE iterations times multiple imputations) is supposed to be written. The effective sample size is necessary to compute the correction for the total variance as originally suggested by Parzen, Lipsitz and Fitzmaurice 2005. The objectname is midastouch.neff.
debug FOR EXPERTS. Null or character string. The name of an existing environment in which the input is supposed to be written. The objectname is midastouch.inputlist.
... Other named arguments.
Details

Imputation of y by predictive mean matching, based on Rubin (1987, p. 168, formulas a and b) and Siddique and Belin 2008. The procedure is as follows:

1. Draw a bootstrap sample from the donor pool.
2. Estimate a beta matrix on the bootstrap sample by the leave one out principle.
3. Compute type II predicted values for yobs (nobs x 1) and ymis (nmis x nobsv).
4. Calculate the distance between all yobs and the corresponding ymis.
5. Convert the distances in drawing probabilities.
6. For each recipient draw a donor from the entire pool while considering the probabilities from the model.
7. Take its observed value in y as the imputation.

Value

Numeric vector of length sum(!ry) with imputations

Author(s)

Philipp Gaffert, Florian Meinfelder, Volker Bosch 2015

References


Examples

## from R:: miceL slightly adapted ##

# do default multiple imputation on a numeric matrix
library(midastouch)
library(mice)
imp <- mice(nhanes, method = 'midastouch')
imp

# list the actual imputations for BMI
imp$imp$bmi

# first completed data matrix
complete(imp)

# imputation on mixed data with a different method per column
mice(nhanes2, method = c('sample','midastouch','logreg','norm'))
Index

*Topic mice
  mice.impute.midastouch, 2
  mice.impute, 2
  midastouch (mice.impute.midastouch), 2