Package ‘absorber’

February 23, 2024

Type Package
Title Variable Selection in Nonparametric Models using B-Splines
Version 1.0
Date 2024-02-22
Author Mary E. Savino [aut, cre],
Celine Levy-Leduc [ctb]
Maintainer Mary E. Savino <mary.savino@outlook.fr>
Description A variable selection method using B-Splines in multivariate nonparametric regression models. Based on partial derivatives regularization (ABSORBER) implements a novel variable selection method in a nonlinear multivariate model using B-splines. For further details we refer the reader to the paper Savino, M. E. and Lévy-Leduc, C. (2024), <https://hal.science/hal-04434820>.
License GPL-2
Encoding UTF-8
LazyData true
Depends R (>= 3.5.0), Matrix, sparsegl, fda, parallel
Imports ggplot2, MASS, irlba
Suggests knitr, markdown
VignetteBuilder knitr
NeedsCompilation no
Repository CRAN
Date/Publication 2024-02-23 18:50:08 UTC

R topics documented:

absorber-package ......................................................... 2
absorber ................................................................. 2
plot_selection ........................................................... 3
x_obs ................................................................. 4
y_obs ................................................................. 4

Index 5
absorber-package

Variable Selection in Nonparametric Models using B-Splines

Description

absorber consists of two functions: "absorber.R" and "plot_selection.R". For further information on how to use these functions, we refer the reader to the vignette of the package.

Details

Two datasets are also provided within this package and used as examples of this manual and in the vignette.

Author(s)

Mary E. Savino

Maintainer: Mary E. Savino <mary.savino@outlook.fr>

References


absorber

Variable selection in nonparametric models

Description

This function implements the method described in Savino, M. E. and Lévy-Leduc, C (2024) for variable selection in nonlinear multivariate settings where observations are assumed to satisfy a nonparametric regression model. Each observation point should belong to $[0, 1]^p$.

Usage

absorber(x, y, M = 3, K = 1, all.variables = NULL, parallel = FALSE, nbCore = 1)

Arguments

x
matrix of $p$ columns containing the input values of the observations, each observation belonging to $[0, 1]^p$.

y
vector containing the corresponding response variable associated to the input values x.

M
order of the B-spline basis used in the regression model. Default is 3 (quadratic B-splines).

K
number of evenly spaced knots to use in the B-spline basis. Default value is 1.
plot_selection

all.variables list of characters or integers, labels of the variables. Default is NULL.
parallel logical, if TRUE then a parallelized version of the code is used. Default is FALSE.
nbCore numerical, number of cores used for parallelization, if parallel is set to TRUE.

Value

selec.var list of vectors of the selected variables, one vector for each penalization parameter.
aic.var vector of variables selected using AIC.

Examples

# --- Loading values of x --- #
data('x_obs')
# --- Loading values of the corresponding y --- #
data('y_obs')
x_trunc = x_obs[1:70,,drop=FALSE]
y_trunc = y_obs[1:70]

# --- Variable selection of f1 --- #
absorber(x=x_trunc, y=y_trunc, M = 3)

# --- Parallel computing --- #
absorber(x=x_trunc, y=y_trunc, M = 3, parallel = TRUE, nbCore = 2)

plot_selection Visualization of the selected variables

Description

This function produces a histogram of the variable selection percentage for each variable on which \( f \) depends. It also displays the results obtained with the AIC.

Usage

plot_selection(object)

Arguments

object output obtained with absorber().

Value

This function produces a ggplot2::ggplot() plot to visualize the variables selected with absorber().
Examples

```r
# --- Loading values of x --- #
data('x_obs')
# --- Loading values of the corresponding y ---#
data('y_obs')
x_trunc = x_obs[1:70,,drop=FALSE]
y_trunc = y_obs[1:70]

# --- Variable selection of f1 --- #
res = absorber(x = x_trunc, y = y_trunc, M = 3)
plot_selection(res)
```

---

x_obs

*Observation matrix x of five variables*

Description

An example of 700 observations for the variable selection of function $f_1$ (see Savino and Lévy-Leduc (2024) for more details) with five input variables.

Usage

data("x_obs")

Format

Numeric matrix of 700 rows and 5 columns.

---

y_obs

*Values of the response variable of the noisy observation set of five input variables*

Description

An example of noisy observations obtained by adding a Gaussian noise to $f_1(x_i)$ associated to the input values contained in x_obs.rda. See Savino and Lévy-Leduc (2024) for the expression of $f_1$.

Usage

data("y_obs")

Format

Numeric vector of 700 values.
Index

* package
  absorber-package, 2

absorber, 2
absorber(), 3
absorber-package, 2

ggplot2::ggplot(), 3
plot_selection, 3

x_obs, 4
y_obs, 4