

Package ‘smoothic’

October 27, 2021

Type Package

Title Variable Selection Using a Smooth Information Criterion

Version 0.1.0

Maintainer Meadhbh O'Neill <meadhbh.oneill@ul.ie>

Description Implementation of the SIC epsilon-telescope method, either using single or multi-parameter regression. This package contains the data analyses from O'Neill and Burke (2021). ``Variable Selection Using a Smooth Information Criterion for Multi-Parameter Regression Models". <[arXiv:2110.02643](https://arxiv.org/abs/2110.02643)>.

License GPL-3

URL <https://github.com/meadhbh-oneill/smoothic>

Encoding UTF-8

LazyData true

RoxygenNote 7.1.2

NeedsCompilation no

Author Meadhbh O'Neill [aut, cre],
Kevin Burke [aut]

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2021-10-27 14:40:02 UTC

R topics documented:

bostonhouseprice	2
pcancer	3
smoothic	4
sniffer	5
summary.smoothic	6

Index	8
--------------	----------

bostonhouseprice *Boston House Price Data*

Description

Data, which come from a study by Harrison Jr and Rubinfeld (1978), examining the association between median house prices in a particular community with various community characteristics.

Usage

bostonhouseprice

Format

A data frame with 506 rows and 9 variables:

crime crimes committed per capita

rooms average number of rooms per house

radial index of accessibility to radial highways

stratio average student-teacher ratio of schools in the community

lowstat percentge of the population that are "lower status"

lnox log(annual average nitrogen oxide concentration (pphm))

lproptax log(property tax per \$1000)

ldist log(weighted distances to five employment centres in the Boston region)

lprice log(median house price (\$))

Source

<https://CRAN.R-project.org/package=wooldridge>

References

Harrison Jr, D. and Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of environmental economics and management*, 5(1):81-102.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.

pcancer

Prostate Cancer Data

Description

Data, which come from a study by Stamey et al. (1989), examining the correlation between the level of prostate-specific antigen (PSA) and various clinical measures in men who were about to receive a radical prostatectomy.

Usage

pcancer

Format

A data frame with 97 rows and 9 variables:

lcavol log(cancer volume (cm³))

lweight log(prostate weight (g))

age age of the patient

lbph log(amount of benign prostatic hyperplasia (cm²))

svi presence of seminal vesicle invasion (1=yes, 0=no)

lcp log(capsular penetration (cm))

gleason Gleason score

pgg45 percentage of Gleason scores four of five

lpsa log(PSA (ng/mL))

Source

<https://web.stanford.edu/~hastie/ElemStatLearn/datasets/prostate.data>

References

Stamey, T. A., Kabalin, J. N., McNeal, J. E., Johnstone, I. M., Freiha, F., Redwine, E. A., and Yang, N. (1989). Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate. ii. radical prostatectomy treated patients. *The Journal of urology*, 141(5):1076-1083.

smoothic

*Variable Selection Using a Smooth Information Criterion (SIC)***Description**

Implements the SIC ϵ -telescope method, either using single or multi-parameter regression. Returns estimated coefficients, estimated standard errors (SEE) and the value of the penalized likelihood function. Note that the function will scale the predictors to have unit variance, however, the final estimates are converted back to their original scale.

Usage

```
smoothic(
  formula,
  data,
  model = "mpr",
  lambda = "log(n)",
  epsilon_1 = 10,
  epsilon_T = 1e-05,
  steps_T = 100,
  zero_tol = 1e-08,
  tol = 1e-08,
  max_it = 10000,
  initial_step = 10,
  max_step_it = 1000
)
```

Arguments

formula	An object of class " <code>formula</code> ": a two-sided object with response on the left hand side and the model variables on the right hand side.
data	A data frame containing the variables in the model; the data frame should be unstandardized.
model	The type of regression to be implemented, either <code>model = "mpr"</code> for multi-parameter regression, or <code>model = "spr"</code> for single parameter regression (i.e., classical normal linear regression). Defaults to <code>model="mpr"</code> .
lambda	Value of penalty tuning parameter. Suggested values are " <code>log(n)</code> " and " <code>2</code> " for the BIC and AIC respectively. Defaults to <code>lambda="log(n)"</code> for the BIC case.
epsilon_1	Starting value for ϵ -telescope. Defaults to 10.
epsilon_T	Final value for ϵ -telescope. Defaults to <code>1e-05</code> .
steps_T	Number of steps in ϵ -telescope. Defaults to 100.
zero_tol	Coefficients below this value are treated as being zero. Defaults to <code>1e-08</code> .
tol	Convergence tolerance for the optimization. Defaults to <code>1e-08</code> .
max_it	Maximum number of iterations to performed before the optimization is terminated. Defaults to <code>1e+04</code> .

<code>initial_step</code>	Initial step length for step halving in Newton-Raphson algorithm. Defaults to 10.
<code>max_step_it</code>	Maximum allowable number of steps to take for step halving in Newton-Raphson algorithm. Defaults to 1e+03.

Value

A list with estimates and estimated standard errors.

- `coefficients` - vector of coefficients.
- `see` - vector of estimated standard errors.
- `model` - the matched type of model which is called.
- `plike` - value of the penalized likelihood function.

Author(s)

Meadhbh O'Neill

References

O'Neill, M. and Burke, K. (2021) Variable Selection Using a Smooth Information Criterion for Multi-Parameter Regression Models. <arXiv:2110.02643>

Examples

```
# Sniffer Data -----
# MPR Model ----
results <- smoothic(
  formula = y ~ .,
  data = sniffer,
  model = "mpr"
)
summary(results)
```

sniffer

Sniffer Data

Description

Data examining the factors that impact the amount of hydrocarbon vapour released when gasoline is pumped into a tank.

Usage

```
sniffer
```

Format

A data frame with 125 rows and 5 variables:

tanktemp initial tank temperature (degrees F)
gastemp temperature of the dispensed gasoline (degrees F)
tankpres initial vapour pressure in the tank (psi)
gaspres vapour pressure of the dispensed gasoline (psi)
y hydrocarbons emitted (g)

Source

<https://CRAN.R-project.org/package=alr4>

References

Weisberg, S. (2014). Applied Linear Regression, 4th edition. Hoboken NJ: Wiley.

summary.smoothic *Summarising Smooth Information Criterion (SIC) Fits*

Description

summary method class “smoothic”

Usage

```
## S3 method for class 'smoothic'
summary(object, ...)
```

Arguments

object an object of class “smoothic” which is the result of a call to `smoothic`.
... further arguments passed to or from other methods.

Value

A list containing the following components:

- **model** - the matched model from the smoothic object.
- **coefmat** - a typical coefficient matrix whose columns are the estimated regression coefficients, estimated standard errors (SEE) and p-values.
- **plike** - value of the penalized likelihood function.

Author(s)

Meadhbh O’Neill

Examples

```
# Sniffer Data -----  
# MPR Model ----  
results <- smoothic(  
  formula = y ~ .,  
  data = sniffer,  
  model = "mpr"  
)  
summary(results)
```

Index

* datasets

bostonhouseprice, 2

pcancer, 3

sniffer, 5

bostonhouseprice, 2

formula, 4

pcancer, 3

print.summary.smoothic

(summary.smoothic), 6

smoothic, 4, 6

sniffer, 5

summary.smoothic, 6