Package 'psav'

June 20, 2021

Type Package

Title Pliable regression spline model using B-splines and auxiliary variables

Version 1.0.0

Author Jae-Kwon Oh and Jae-Hwan Jhong

Maintainer Jae-Kwon Oh <cort0048@gmail.com>

Description

We conducted a study on a regression spline estimator with a few prespecified auxiliary variables. For the implementation of the proposed estimators,we adapted a coordinate descent algorithm. This was implemented by considering a structure of the sum of the residuals squared objective function determined by the B-spline and the auxiliary coefficients. We also considered an efficient stepwise knot selection algorithm based on the Bayesian information criterion.

License 4.0.2

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

NeedsCompilation no

R topics documented:

`Backward` *Knot elimination via Backward*

Description

`Backward` is used as a function to remove knots in `stepwise`. Among the selected knots, the oldest knot is removed and a new knot is selected again.

Usage

```
Backward(x, z, y, inter_knots, left_knots, right_knots, add_knot,
         existing_knot, order, criter, selection_number)
```

Arguments

x	$n \times 1$ predictor vector. n is a sample size.
z	$n \times K$ auxiliary binary variable matrix. K is the number of levels for auxiliary variable.
y	$n \times 1$ response vector.
inter_knots	interior knots
left_knots	left boundary knots
right_knots	right boundary knots
add_knot	knots selected by forward
existing_knot	existing knots excluding newly selected knot (add_knot)
order	order of B-splines. 1 fits constant, 2 fits linear, 3 fits quadratic, 4 fits cubic splines.
criter	criter has a numeric of 8 or 9. 8 indicates "BIC" and 9 indicates "AIC"
selection_number	indicates the numberth of knots selected.

---

fit_model *Pilable Regression Spline model using B-splines and auxiliary variables*

Description

Fit a pilable regression spline model

Usage

```
fit_model(x, z, y, knots, order, max_iter = 1000, epsilon = 1e-5)
```

Arguments

x	$n \times 1$ predictor vector. n is a sample size.
z	$n \times K$ auxiliary binary variable matrix. K is the number of levels for auxiliary variable.
y	$n \times 1$ response vector.
knots	selected knots among the quantiles of x
order	order of B-splines. 1 fits constant, 2 fits linear, 3 fits quadratic, 4 fits cubic splines.
max_iter	maximum iteration of Coodinate Descent Algorithm
epsilon	stopping criteria of Coodinate Descent Algorithm. If the difference in the residual sum of squares is less than epsilon, the algorithm is stopped.

Details

We conduct a study on a statistical learning for modeling and analysis of data with another auxiliary variable in addition to the response and predictor variables. The estimation is expressed as a linear combination of B-splines. Coefficients are estimated by minimizing the residual sum of squares via Coordinate Descent Algorithm. For more details,see Oh and Jhong (0000).

Value

The results are output as a list. The list consists of the following:

`beta`	coefficient vectors for predictor terms
`gamma`	coefficient matrix for auxiliary variable terms
`residuals`	the residuals, that is response minus fitted values.
`fitted_values`	the fitted values
`iter`	number of iteration
`inter_knots`	interior knots
`n`	sample size
`BIC`	Bayesian Information Criterion
`AIC`	Akaike Information Criterion

Author(s)

Jae-Kwon Oh and Jae-Hwan Jhong

`Forward` *Knot selection via Forward*

Description

Choose a knot from the knots other than the knot in `pre_fit`

Usage

```
Forward(x, z, y, inter_knots, left_knots, right_knots,
        pre_fit, order, criter, selection_number)
```

Arguments

`x`	$n \times 1$ predictor vector. n is a sample size.
`z`	$n \times K$ auxiliary binary variable matrix. K is the number of levels for auxiliary variable.
`y`	$n \times 1$ response vector.
`inter_knots`	interior knots
`left_knots`	left boundary knots
`right_knots`	right boundary knots
`pre_fit`	model before choosing a new knot (`fit_model` results)

order	order of B-splines. 1 fits constant, 2 fits linear, 3 fits quadratic, 4 fits cubic splines.
criter	criter has a numeric of 8 or 9. 8 indicates "BIC" and 9 indicates "AIC"
selection_number	indicates the numberth of knots selected.

StepWise *Knot Selection via Stepwise Method*

Description

StepWise is used for knot selection. At every step, the model fits with the fit_model function.

Usage

```
StepWise(x, z, y, max_knots = 10, order = 4,
         criterion = "BIC", max_iter = 1000, epsilon = 1e-5)
```

Arguments

x	$n \times 1$ predictor vector. n is a sample size.
z	$n \times K$ auxiliary binary variable matrix. K is the number of levels for auxiliary variable.
y	$n \times 1$ response vector.
max_knots	the number of initial knots. should specify as the number of that can be divided into small enough segments.
order	order of B-splines. 1 fits constant, 2 fits linear, 3 fits quadratic, 4 fits cubic splines.
criterion	scale used as a criterion for knot selection. Criteria can be selected as "BIC" or "AIC".
max_iter	maximum iteration of Coodinate Descent Algorithm
epsilon	stopping criteria of Coodinate Descent Algorithm. If the difference in the residual sum of squares is less than epsilon, the algorithm is stopped.

Details

Starting from the NULL model, the optimal model is calculated by repeating the selection and removal of knot. Backward is performed when there are more than two knots.

Author(s)

Jae-Kwon Oh and Jae-Hwan Jhong

Examples

```
# Example for Bone Mineral Density Data

library(psav)
library(loon.data)
data("bone")

# When the auxiliary variable is sex
par(mfrow = c(1,2))
x = bone$age
y = bone$rspnbmd
z = bone$sex
tz = transform(z,
               class1 = ifelse(z == "male", 1, 0),
               class2 = ifelse(z == "female", 1, 0))[,2:3]
x_m = x[z == "male"]
x_f = x[z == "female"]

fit = StepWise(x, tz, y, max_knots = 50, order = 4,
               criterion = "BIC", max_iter = 1000, epsilon = 1e-5)
fit_m = fit$fitted_values[z == "male"]
fit_f = fit$fitted_values[z == "female"]

plot(x, y, bty = "n", type = "n", xlab = "age",
     ylab = "Relative spinal bone mineral density",
     cex.lab = 1.3, cex.axis = 1.2)
points(x, y, col = "grey")
lines(sort(x_m), fit_m[order(x_m)], lty = 1, lwd = 2)
lines(sort(x_f), fit_f[order(x_f)], lty = 3, lwd = 2)
abline(v = fit$knots, col = "grey", lty = 2)
legend("topright", c("male", "female"),
       col = "black",
       lty = c(1, 3), lwd = 2,
       bty = "n", cex = 1.1)


# When the auxiliary variable is ethnic
bone_ethnic <- na.omit(bone)
x <- bone_ethnic$age
y <- bone_ethnic$rspnbmd
z <- bone_ethnic$ethnic

tz = transform(z,
               class1 = ifelse(z == "White", 1, 0),
               class2 = ifelse(z == "Hispanic", 1, 0),
               class3 = ifelse(z == "Asian", 1, 0),
               class4 = ifelse(z == "Black", 1, 0))[, 2:5]
x_w = x[z == "White"]
x_h = x[z == "Hispanic"]
x_a = x[z == "Asian"]
x_b = x[z == "Black"]

fit = StepWise(x, tz, y, max_knots = 50, order = 4,
               criterion = "BIC", max_iter = 1000, epsilon = 1e-5)

fit_w = fit$fitted_values[z == "White"]
```

```
fit_h = fit$fitted_values[z == "Hispanic"]
fit_a = fit$fitted_values[z == "Asian"]
fit_b = fit$fitted_values[z == "Black"]

plot(x, y, bty = "n", type = "n", xlab = "age",
     ylab = "Relative spinal bone mineral density",
     cex.lab = 1.3, cex.axis = 1.2)
points(x, y, col = "grey")
lines(sort(x_w), fit_w[order(x_w)], lty = 1, lwd = 2)
lines(sort(x_h), fit_h[order(x_h)], lty = 3, lwd = 2)
lines(sort(x_a), fit_a[order(x_a)], lty = 4, lwd = 2)
lines(sort(x_b), fit_b[order(x_b)], lty = 5, lwd = 2)
abline(v = fit$knots, col = "grey", lty = 2)
legend("topright", c("White", "Hispanic", "Asian", "Black"),
       col = "black",
       bty = "n", cex = 1.1, lwd = 2,
       lty = c(1,3,4,5))
```

Index