R documentation

of 'VA-package.Rd' etc.

July 9, 2016

VA-package	*Variational Approximations and its evaluations in geoadditive (quantile) regression*

Description

This package gives the possibility to estimate Bayesian additive (quantile) regression estimated via variational approximations.

Details

Package:	VA
Type:	Package
Version:	1.0
Date:	2011-10-17
License: 2.14.0 LazyLoad:	yes

The main function of this package is va, a function which uses the theory of variational approximations to calculate the parameters for Bayesian geoadditive regression models for mean regression as well as for quantile regression. The other functions are or for data preperation for the different effects (lin, bsplines, spatial)and (risk)for model evaluation.

Author(s)

Elisabeth Waldmann
University of Goettingen
<ewaldma@uni-goettingen.de>

with contributions from

Thomas Kneib
Universtity of Goettingen
http://www.uni-goettingen.de/de/264255.html

Udo Schroeder
University of Oldenburg
http://www.uni-oldenburg.de/index/personen/?username=USchroeder1

Maintainer: Elisabeth Waldmann

References

Yue, R.Y., Rue, H. *Bayesian inference for additive mixed quantile regression models* Computational Statistics & Data Analysis, (2011), 55, 84-96.

Ormerod, J.T. and Wand, M.P. *Explaining Variational Approximations*. The American Statistician, (2010), 64, 140-153.

Wand, M. P.; Ormerod, J. T.; Padoan, S. A.; and Fruhwirth, R., *Variational Bayes for Elaborate Distributions*, http://ro.uow.edu.au/cssmwp/56 (working paper).

See Also

BayesX

`bsplines` *B-Splines Basis for Bayesian Inference*

Description

Generates a b-spline basis matrix with attributes for Bayesian inference (such as the penalisation matrix and prior parameters). This function is based on the function spline.des from the package **splines**.

Usage

```
bsplines(x, M = 20, degree = 3, order = 2, a_1=1, b_1=.00001, knots=NULL, mx=NULL)
```

Arguments

`x`	data
`M`	number of knots
`degree`	the degree of splines
`order`	the order of the random walk prior
`a_1`	shape parameter for the gamma-prior for the smoothing parameter
`b_1`	scale parameter for the gamma-prior for the smoothing parameter
`knots`	knots at which the spline is to be evaluated
`mx`	value the spline is centered around, default is mean(x)

Value

The matrix returned is the basis matrix. The attributes are K (the penalisation matrix with differences of order "order") and the class of the effect ("splines") as well as the prior parameters a_1 and b_1.

Author(s)

Elisabeth Waldmann

<ewaldma@uni-goettingen.de>

References

Fahrmeir L., Kneib T. and Lang S. (2009) *Regression* Springer, New York

See Also

lin, spatial, random

Examples

```
x <- runif(100,-3,3)
Z <- bsplines(x)
K <- attr(Z, "K")
attr(Z, "class")
```

`lin` *Linear effects for Bayesian inference*

Description

Simplifies the access to prior parameters given in a formula type object.

Usage

```
lin(x, mu0=0, sig0=10000)
```

Arguments

`x`	data
`mu0`	mean for the gaussian prior for the linear effect
`sig0`	variance for the gaussian prior for the linear effect

Value

The vector returned only comprises the data. The attributes are the class of the effect("linear"), and the prior parameters.

Author(s)

Elisabeth Waldmann

See Also

bsplines, spatial

Examples

```
x <- runif(100,-3,3)
Z <- lin(x)
attr(Z, "class")
```

`plot.va` *Plots for VA*

Description

Plots the different effects of a va object.

Usage

```
## S3 method for class 'va'
plot(x,map=NULL,ask=TRUE,linear=TRUE,bsplines=TRUE,spatial=TRUE,random=TRUE,...)
```

Arguments

`x`	object of va-type
`map`	if there is a spatial effect, the map has to be given as boundary file (i.e. .bnd)
`ask`	if `ask=TRUE`, the user will be asked before the page changes to the following plot
`linear`	if `linear = FALSE` there will be no plot of the linear effects, even though there are linear effects in the model
`bsplines`	if `bsplines = FALSE`, there will be no plot of the nonlinear effects, even though there are nonlinear effects in the model
`spatial`	if `spatial = FALSE`, there will be no plot of the spatial effects, even though there are spatial effects in the model
`random`	if `random = FALSE`, there will be no plot of the random effects, even though there are random effects in the model
`...`	Additional arguments

Details

The function returns a plot of densities of the linear effects, the fitted curve of the centralized b-splines (plotted against the covariates), a map with colors refering to the spatial effects and barplots of the random effects. The map is plotted with the function drawmap of the package **BayesX**.

Author(s)

Elisabeth Waldmann
<ewaldma@uni-goettingen.de>

predict.va *Predict method for Variational Approximations*

Description

Predicts values based on a variational approximation object

Usage

```
## S3 method for class 'va'
predict(object, newdata=NULL, ...)
```

Arguments

object	object of va-type
newdata	data for which the prediction is to be done. The data has to be passed in form of matrix or dataframe. The colnames have to be the ones, to which the formula referes. If newdata=NULL the prediction of the va object is returned (usually this is the prediction based on the estimation data set, if the pred-Argument was NULL in the estimation. If there was a prediction-dataset, predict returns the prediction for this dataset)
...	Additional arguments

Value

The prediction is returned as a vector

Note

Random effects are set to zero in the prediction.

Author(s)

Elisabeth Waldmann

random *Random effects*

Description

Generates a basis matrix for a random effect facilitating the access to parameters in a formula type object for Bayesian inference (such as the prior parameters)

Usage

```
random(ind, a_1=1, b_1=1e-5)
```

Arguments

ind	information about the individuals or group
a_1	shape parameter for the gamma-prior for the smoothing parameter
b_1	scale parameter for the gamma-prior for the smoothing parameter

Value

The matrix returned is the basis matrix. The attributes are K (the neighbouring matrix) and the class of the effect ("random") as well as the prior parameters a_1 and b_1.

Author(s)

Elisabeth Waldmann

<ewaldma@uni-goettingen.de>

References

Fahrmeir L., Kneib T. and Lang S. (2009) *Regression* Springer, New York

See Also

lin, bsplines, spatial

---

residuals.va *Residuals for Variational Approximations*

Description

residuals values based on a variational approximation object

Usage

```
## S3 method for class 'va'
residuals(object, ...)
```

Arguments

object	object of va-type
...	Additional arguments

Value

The residuals are returned as a vector

Author(s)

Elisabeth Waldmann
<ewaldma@uni-goettingen.de>

risk *Risk for mean and quantile estimations*

Description

This functions can be used to evaluate the estimation of mean or quantile regression models. It or returns the MSE or the risk function calculated by the checkfunction.

Usage

```
risk(y, pred, quant = FALSE, tau = 0.5)
```

Arguments

y	numeric vector of data
pred	numeric value or vector of the prediction for y
quant	logical. If TRUE the checkfunction is calculated (the default is FALSE)
tau	numerical. The quantile for which the risk is to be calculated. This value has to be between 0 and 1 (the default value is 0.5).

Examples

```
y <- rnorm(1000,3,5)
x <- mean(y)
## MSE of the mean
risk(y,x)

x <- quantile(y, .4)
###
risk(y,x,quant=TRUE, tau=.4)
```

spatial *Gaussian random fields*

Description

Generates a basis matrix for a Gaussian random field facilitating the access to parameters in a formula type object for Bayesian inference (such as the prior parameters)

Usage

```
spatial(geo, map,a_1=1, b_1=.0001)
```

Arguments

geo	spatial information
map	the corresponding map (given as .gra or .bnd)
a_1	shape parameter for the gamma-prior for the smoothing parameter
b_1	scale parameter for the gamma-prior for the smoothing parameter

Value

The matrix returned is the basis matrix. The attributes are K (the neighbouring matrix) and the class of the effect ("spatial") as well as the prior parameters a_1 and b_1.

Author(s)

Elisabeth Waldmann

<ewaldma@uni-goettingen.de>

References

Fahrmeir L., Kneib T. and Lang S. (2009) *Regression* Springer, New York

See Also

lin, bsplines

Examples

```
germany <- read.bnd(system.file("examples/germany.bnd", package="BayesX"))
gramap<-bnd2gra(germany)
centroids <- get.centroids(germany)
geo_ind <- c(sample(1:310,size=1000,replace=TRUE))
geo <- row.names(centroids)[geo_ind]
Z <- spatial(geo, gramap)
K <- attr(Z, "K")
attr(Z, "class")
```

`summary.va` *Summary for VA-Regression*

Description

Summarizes the results of the VA-Regerssion

Usage

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

Arguments

`object`	object of va-type
`...`	Additional arguments

Details

This function returns the formula-object, the quantile the regression was done for (if `quant=TRUE`), a table of the linear coefficients, containing the standard deviation as well as the 0.025 and the 0.975-quantile of the corresponding Gaussian distribution. The nonlinear, spatial and random effects are mentioned with their meaning.

Author(s)

Elisabeth Waldmann
<ewaldma@uni-goettingen.de>

---

va *Estimation of Parameters of Bayesian Additive (Quantile) Regression*

Description

function to estimate parameters of bayesian additive (quantile) regression

Usage

```
va(formula, data = NULL, intercept = TRUE, quant = FALSE, ta = 0.5,
k = 100, sto = 1e-04,
a_0 = 1, b_0 = 1e-05,
plotmse = FALSE, norm = TRUE, pred = NULL, ...)
```

Arguments

formula	a formula object, with the response on the left of a ~ operator, and the terms, separated by + operators, on the right, by default calculated as linear effects. The effects can be changed by using different functions, bsplines, spatial, lin or random.
data	a matrix containing the data
intercept	logical value if an intercept should be included (the default value is TRUE)
quant	logical value indicating if quantile regression is to be calculated (the default value is FALSE)
ta	numerical. The quantile to be estimated. This value has to be between 0 and 1 (the default value is 0.5).
k	maximal number of iterations used
sto	stoping criterion, using the relative change of the risk to stop the program before reaching k iterations
a_0	shape-parameter for gamma-prior for modelprecision parameter
b_0	rate-parameter for gamma-prior for modelprecision parameter
plotmse	logical value, indicating if the risk should be plotted in each iteration (default value is FALSE). Note that the plotted risk refers to the data used for estimation (i.e. defaultwise the normalised data) and therefore might differ highly from the risk calculated afterwards.
norm	logical value, indicating if the data are to be normalized before calculation (default value is TRUE)
pred	matrix of data used for prediction. If there are no data for prediction, the prediction will be done on the data which were used for estimation
...	Extra parameters needed

Value

beta	the estimated linear effects
sig_bet	the covariance matrix of the linear effects
Xlinear	a matrix containing all the linear information
Z	the designmatrix of the nonlinear effects
gamma	list of vectors of ALL estimated nonlinear effects (i.e. bsplines, spatial and random effects)
sig_gam	the covariance matrix of all nonlinear effects
Xnonlinear	a data matrix containing all the nonlinear information
tau_inv	a vector of the inverse of the smoothing parameters
delta	the model precision
i	the number of iterations until the stopping criterion or the maximal number of iterations was reached (note that there is a warning if k is reached before the stopping criterion is fulfilled)
call	formula given to the function
prediction	a vector with the predicted values for either the data the estimation was based on (if pred=NULL) or for the data given in pred
tau	the quantile the estimation was done for (of quant = TRUE)
residuals	a vector of the difference between y and the prediction
replace	a vector which stores the numbers of the iterations in which the parameters in the distribution of the modelprecision were replaced in order to make the numerical integration possible. See section note for further information.

Note

If you estimate the coeffecients for a quantile regression, the approximated precision of the model is calculated via numerical integration using a subroutine in C. If values in this subroutine get too big, the step is repeated with the compensatory values for the parameters of the precision. This only happens for very sparse regions in the datasets (i.e. usually in the margins) and even in these cases the integral will normally be calculated for the real values before convergence. If the real values were not used throughout the whole iterating process, there will be a warning. If the results of your estimations are somehow akward even though there was no warning, you should check va$replace, which returns a vector, telling you in which iteration the parameters were replaced. If you set plotmse=TRUE, the points in which the parameters are replaced are marked in red. If they are replaced in an iteration close to the last iteration, you might want to think about the sparsity of the data in the quantile you tried to estimate (estimating the .01-quantile of a dataset containing 50 observations might not be very sensible anyway). If you want to use the estimations anyway, you should know that even if the point estimators (such as the beta-vector) still seem trustworthy, the covariances get to big (you might consider taking the last estimation BEFORE the replacement to be the closest to the "truth").

Author(s)

Elisabeth Waldmann
University of Goettingen
<ewaldma@uni-goettingen.de>

References

Ormerod, J.T. and Wand, M.P. *Explaining Variational Approximations*. The American Statistician, (2010), 64, 140-153.

Wand, M. P.; Ormerod, J. T.; Padoan, S. A.; and Fruhwirth, R., *Variational Bayes for Elaborate Distributions*, http://ro.uow.edu.au/cssmwp/56 (working paper).

See Also

Graphical presentation can be done by using plot.va.

Examples

```
#data(cars)
#va(dist~speed, data=cars)

## Example for using linear and nonlinear effects
set.seed(123)
x <- runif(100,-3,3)
x2 <- runif(100,-3,3)
x3 <- runif(100,-3,3)
x4 <- runif(100,0,3)
y <- 3 + 3*sin(x)+3*x2+4*x3+log(x4)+rnorm(100,0,1)

L<-va(y~bsplines(x)+x2 +x3 +bsplines(x4), quant=TRUE)
summary(L)
plot(L)

## Example using different quantiles in a heteroscedastic scenario

set.seed(123)
x <- runif(500,0,3)
y <- 3 + 1.5*x+rnorm(500,0,.3*x)

L1<-va(y~x, quant=TRUE, ta=.1)
L3<-va(y~x, quant=TRUE, ta=.3)
L5<-va(y~x, quant=TRUE, ta=.5)
L7<-va(y~x, quant=TRUE, ta=.7)
L9<-va(y~x, quant=TRUE, ta=.9)
plot(x,y, col="lightgrey")
lines(sort(x), L1$pred[order(x)])
lines(sort(x), L3$pred[order(x)])
lines(sort(x), L5$pred[order(x)])
lines(sort(x), L7$pred[order(x)])
lines(sort(x), L9$pred[order(x)])
```