Package 'gagam'

April 10, 2020

Type Package

Title Genetic Algorithm for Generalized Additive Models

Version 0.1

Author Mark Cus <mark-cus@outlook.com>

Maintainer Mark Cus <mark-cus@outlook.com>

Description This package implements a genetic algorithm which performs simulatenous variable selection and structure discovery in generalized additive models.

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.0

Depends utils,
> stats,
> mgcv,
> parallel,
> foreach,
> doParallel,
> doMC

URL https://github.com/markcus1/gagam

BugReports https://github.com/markcus1/gagam/issues

Suggests knitr,
> rmarkdown,
> mlbench

VignetteBuilder knitr

References Cus, Mark. 2020. ``Simultaneous Variable Selection And Structure Discovery In Generalized Additive Models''. https://github.com/markcus1/gagam/blob/master/GAGAMpaper.pdf.

R topics documented:

1

---

gagam   *Genetic Algorithm for Generalized Additive Models*

Description

Implements the genetic algorithm for simultaneous variable selection and structure discovery in generalized additive models. For a given dependent variable and a set of explanatory variables, the genetic algorithm determines which regressors should be included linearly, which nonparametrically, and which should be excluded from the regression equation. The aim is to minimize the Bayesian Information Criterion value of the model.

Usage

```
gagam(
  y,
  x,
  pop_size = 500,
  Kvar = 15,
  Kint = 0,
  no_gen = 100,
  p_m = 0.05,
  p_int = 0.1,
  p_nonpar = 0.1,
  p_int_nonpar = 0.1,
  multicore = TRUE,
  cores = NULL,
  k = 10,
  bs = "cr",
  family = gaussian(),
  method = "REML",
  optimizer = c("outer", "newton"),
  reduc = NULL,
  always_par = NULL
)
```

Arguments

y	Vector, matrix, data frame, or factor containing observations of the dependent variable.
x	Matrix or data frame containing all considered explanatory variables. If the columns have names those will be used for variable names in the final output.
pop_size	Size of the population (needs to be a multiple of 500). Default is 500.
Kvar	Maximum number of variables allowed in the final model. Default is 15.
Kint	Maximum number of interactions allowed in the final model. Default is 0.
no_gen	Number of generations until convergence. Default is 100.
p_m	Mutation rate for variables. Default is 0.05.
p_int	Mutation rate for interactions of variables. Default is 0.1.
p_nonpar	Mutation rate for the linear/nonparametric indicators for variables. Default is 0.1.

p_int_nonpar	Mutation rate for the linear/nonparametric indicators for interactions. Default is 0.1.
multicore	Whether to use multiple cores in computation. Strongly recommended but may not work on Windows. Default is TRUE.
cores	Number of cores to use with multicore. Default (NULL) uses all cores.
k	Basis dimension for nonparametric terms estimated using splines. Default is 10.
bs	Spline basis for nonparametric terms. Specified as a two letter character string. Default is the natural cubic spline, bs="cr". See smooth.terms for an overview of what is available.
family	Specifies the family for the gam (see family and family.mgcv). Default is gaussian().
method	Specifies the metric for smoothing parameter selection (see gam). Default is "REML".
optimizer	Specifies the numerical optimization algorithm for the gam (see gam). Default is c("outer","newton").
reduc	Implements additional variable elimination methods at the end of the run of the genetic algorithm. User can choose between methods 1, 2, and 3. Multiple methods can be chosen. E.g. reduc=c(1) or reduc=c(1,3). See GAGAM paper for an explanation of the methods. Default is NULL.
always_par	Vector of the column numbers (in x) of the variables always estimated parametrically (for noncontinuous predictors).

Value

A list containing: gam object (fitted best model), vector of indexes or names of variables included linearly, vector of indexes or names of variables included nonparametrically (and the same lists for interactions if Kint is greater than 0).

References

Cus, Mark. 2020. "Simultaneous Variable Selection And Structure Discovery In Generalized Additive Models". http://www.markcus.com/paper.pdf.

Examples

```
N <- 500
set.seed(123)
xdat <- matrix(rnorm(N*10,0,1),nrow=N,ncol=10)
ydat <- 4*xdat[,1]+5*xdat[,2]+6*xdat[,3]+(xdat[,4])^2 + 4 + rnorm(N,0,0.25)

## Not run:
example_gagam <- gagam(ydat,xdat,Kvar = 6,no_gen = 50)

## End(Not run)
```

---

plot.gagam   *Plot GAGAM*

Description

Plot method for gagam objects.

Usage

```
## S3 method for class 'gagam'
plot(x, ..., reduc = NULL)
```

Arguments

x	Fitted gagam object
...	Any other arguments to pass to `plot.gam`.
reduc	If you wish to summarize one of the reduced models, specify the index of reduction. Specify only one index at a time. Default is NULL.

---

predict.gagam   *Prediction from GAGAM*

Description

Predict method for gagam objects.

Usage

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

Arguments

object	gagam object (result of the gagam() function)
...	Any other arguments to pass to `predict.gam`.
newdata	Matrix or data frame with explanatory variables. The order of columns and their names must be the same as that used to construct the gagam object. If missing, the function extracts in-sample predictions.
reduc	If prediction should be from one of the reduced models, specify the index of reduction. Specify only one index at a time. Default is NULL.

Value

A vector of predictions.

Examples

```
N <- 1000
set.seed(123)
xdat2 <- matrix(rnorm(N*10,0,1),nrow=N,ncol=10)
ydat2 <- 4*xdat2[,1]+5*xdat2[,2]+6*xdat2[,3]+(xdat2[,4])^2 + 4 + rnorm(N,0,0.25)

xdat <- xdat2[1:500,]
ydat <- ydat2[1:500]

## Not run:
example_gagam <- gagam(ydat,xdat,Kvar = 6,no_gen = 50)

## End(Not run)

xdat <- xdat2[501:1000,]
ydat <- ydat2[501:1000]

newdata <- as.data.frame(cbind(as.matrix(ydat),xdat))

## Not run:
predict_gagam(example_gagam,newdata)

## End(Not run)
```

---

summary.gagam   *Summarize GAGAM*

Description

Summary method for gagam objects.

Usage

```
## S3 method for class 'gagam'
summary(object, ..., reduc = NULL)
```

Arguments

object	Fitted gagam object
...	Any other arguments to pass to summary.gam.
reduc	If you wish to summarize one of the reduced models, specify the index of reduction. Specify only one index at a time. Default is NULL.

Value

Summary of the gam object.

Index