Package 'GMDHreg'

August 2, 2020

Type Package

Title Regression using GMDH Algorithms

Version 0.2.1

Author Manuel Villacorta Tilve

Maintainer Manuel Villacorta Tilve <mvt.oviedo@gmail.com>

Description Regression using GMDH algorithms from Prof. Alexey G. Ivakhnenko.
Group Method of Data Handling (GMDH), or polynomial neural networks, is a family of inductive algorithms
that performs gradually complicated polynomial models and selecting the best solution by an external criterion.
In other words, inductive GMDH algorithms give possibility finding automatically interrelations in data, and
selecting an optimal structure of model or network.
The package includes GMDH Combinatorial, GMDH MIA (Multilayered Iterative Algorithm), GMDH GIA (Generalized Iterative Algorithm) and GMDH Combinatorial with Active Neurons.
An introduction of GMDH algorithms:
Farlow, S.J. (1981): ``The GMDH algorithm of Ivakhnenko", The American Statistician, 35(4), pp. 210-215. <doi:10.2307/2683292>
Ivakhnenko A.G. (1968): ``The Group Method of Data Handling -
A Rival of the Method of Stochastic Approximation", Soviet Automatic Control, 13(3), pp. 43-55.

License GPL-3

Encoding UTF-8

LazyData true

Depends R (>= 2.15)

Imports stats, utils

Suggests knitr, rmarkdown

VignetteBuilder knitr

RoxygenNote 7.0.2

NeedsCompilation no

Repository CRAN

Date/Publication 2020-08-02 01:32:11 UTC

R topics documented:

`gmdh.combi` *GMDH Combinatorial*

Description

Build a regression model performing GMDH Combinatorial.
This is the basic GMDH algorithm. For more information, please read the package's vignette.

Usage

```
gmdh.combi(X, y, G = 2, criteria = c("PRESS", "test", "ICOMP"),
  x.test = NULL, y.test = NULL)
```

Arguments

`X`	matrix with N>1 columns and M rows, containing independent variables in the model. Be careful, N>4 and G=2, could be computationally very expensive and time consuming. The data must not contain NAs
`y`	vector or matrix containing dependent variable in the model. The data must not contain NAs
`G`	polynomial degree. 0: linear regression without quadratic and interactrion terms. 1: linear regression with interaction terms. 2: original Ivakhnenko quadratic polynomial.
`criteria`	GMDH external criteria. Values: • PRESS: Predicted Residual Error Sum of Squares. It take into account all information in data sample and it is computed without recalculating of system for each test point. • test: use x.test and y.test to estimate RMSE (Root Mean Squeare Errors).

- ICOMP: Index of Informational Complexity. Like PRESS, it is computed without recalculating of system.

`x.test` matrix with a sample randomly drawn from the initial data. This sample should not be included in X.
It is used when criteria = test.

`y.test` vector or matrix with y values correspond with x.test values.

Value

An object of class 'combi'. This is a list with two elements: results and G.
Results is a list with two elements:

- coef: coeficients of final selected GMDH Combinatorial model.
- CV: external criteria value for selected model.

G the grade of polynomial used in GMDH Combinatorial model.

References

Bozdogan, H. and Haughton, D.M.A. (1998): "Information complexity criteria for regression models", Computational Statistics & Data Analysis, 28, pp. 51-76 <doi: 10.1016/S0167-9473(98)00025-5>

Hild, Ch. R. and Bozdogan, H. (1995): "The use of information-based model selection criteria in the GMDH algorithm", Systems Analysis Modelling Simulation, 20(1-2), pp. 29-50

Ivakhnenko A.G. (1968): "The Group Method of Data Handling - A Rival of the Method of Stochastic Approximation", Soviet Automatic Control, 13(3), pp. 43-55

Examples

```
set.seed(123)
x <- matrix(data = c(rnorm(1050)), ncol = 3, nrow = 350)
colnames(x) <- c("a", "b", "c")
y <- matrix(data = c(10 + x[, "a"] + x[, "b"]^2 + x[, "c"]^3), ncol = 1)
colnames(y) <- "y"
x.test <- x[1:10, ]
y.test <- y[1:10]
x <- x[-c(1:10), ]
y <- y[-c(1:10)]

mod <- gmdh.combi(X = x, y = y, criteria = "PRESS")
pred <- predict(mod, x.test)
summary(sqrt((pred - y.test)^2))
```

`gmdh.combi.twice` *GMDH Twice-Multilayered Combinatorial*

Description

Build a regression model performing GMDH Twice-Multilayered Combinatorial (TMC).
For more information, please read the package's vignette.

Usage

```
gmdh.combi.twice(X, y, criteria = c("PRESS", "test", "ICOMP"), G = 2,
  x.test = NULL, y.test = NULL)
```

Arguments

`X`	matrix with N>1 columns and M rows, containing independent variables in the model. Be careful, N>4 and G=2, could be computationally very expensive and time consuming. The data must not contain NAs
`y`	vector or matrix containing dependent variable in the model. The data must not contain NAs
`criteria`	GMDH external criteria. Values: • PRESS: Predicted Residual Error Sum of Squares. It take into account all information in data sample and it is computed without recalculating of system for each test point. • test: use x.test and y.test to estimate RMSE (Root Mean Squeare Errors). • ICOMP: Index of Informational Complexity. Like PRESS, it is computed without recalculating of system.
`G`	polynomial degree. 0: linear regression without quadratic and interactrion terms. 1: linear regression with interaction terms. 2: original Ivakhnenko quadratic polynomial.
`x.test`	matrix with a sample randomly drawn from the initial data. This sample should not be included in X. It is used when criteria = test.
`y.test`	vector or matrix with y values correspond with x.test values.

Value

An object of class 'combitwice'. This is a list with two elements: results and G
Results is a list with two elements:

- coef: coeficients of final selected GMDH Combinatorial model.
- CV: external criteria value for selected model.

G the grade of polynomial used in GMDH Combinatorial model.

References

Bozdogan, H. and Haughton, D.M.A. (1998): "Information complexity criteria for regression models", Computational Statistics & Data Analysis, 28, pp. 51-76 <doi: 10.1016/S0167-9473(98)00025-5>

Hild, Ch. R. and Bozdogan, H. (1995): "The use of information-based model selection criteria in the GMDH algorithm", Systems Analysis Modelling Simulation, 20(1-2), pp. 29-50

Ivakhnenko A.G. (1968): "The Group Method of Data Handling - A Rival of the Method of Stochastic Approximation", Soviet Automatic Control, 13(3), pp. 43-55

Examples

```
set.seed(123)
x <- matrix(data = c(rnorm(1050)), ncol = 3, nrow = 350)
colnames(x) <- c("a", "b", "c")
y <- matrix(data = c(10 + x[, "a"] + x[, "b"]^2 + x[, "c"]^3), ncol = 1)
colnames(y) <- "y"
x.test <- x[1:10, ]
y.test <- y[1:10]
x <- x[-c(1:10), ]
y <- y[-c(1:10)]

mod <- gmdh.combi.twice(X = x, y = y, criteria = "PRESS")
pred <- predict(mod, x.test)
summary(sqrt((pred - y.test)^2))
```

---

gmdh.gia *GMDH GIA*

Description

Build a regression model performing GMDH GIA (Generalized Iterative Algorithm) with Active Neurons (Combinatorial algorithm).
For more information, please read the package's vignette.

Usage

```
gmdh.gia(X, y, prune = 40, criteria = c("PRESS", "test", "ICOMP"),
  x.test = NULL, y.test = NULL)
```

Arguments

X	matrix with N>3 columns and M rows, containing independent variables in the model. The data must not contain NAs
y	vector or matrix containing dependent variable in the model. The data must not contain NAs
prune	an integer whose recommended minimum value is the number of initial regressors. The maximum value will depend on the available RAM. It is recommended to work with the maximum value, but it can be computationally very expensive. Prune is the selected number of neurons from layer i to layer i+1. The resulting layer i+1 has prune(prune-1)/2 neurons; for example with prune=150, the resulting nerurons will be 11.175
criteria	GMDH external criteria. Values: • PRESS: predicted residual error sum of squares. • test: use x.test and y.test to estimate RMSE (root mean squeare errors). • ICOMP: Index of Informational Complexity. Like PRESS, it is computed without recalculating of system.
x.test	matrix with a sample randomly drawn from the initial data. It is used when criteria = test. This sample should not be included in X.
y.test	vector or matrix with y values correspond with x.test values.

Value

An object of class gia.

References

Bozdogan, H. and Haughton, D.M.A. (1998): "Information complexity criteria for regression models", Computational Statistics & Data Analysis, 28, pp. 51-76 <doi: 10.1016/S0167-9473(98)00025-5>

Farlow, S.J. (1981): "The GMDH algorithm of Ivakhnenko", The American Statistician, 35(4), pp. 210-215. <doi:10.2307/2683292>

Hild, Ch. R. and Bozdogan, H. (1995): "The use of information-based model selection criteria in the GMDH algorithm", Systems Analysis Modelling Simulation, 20(1-2), pp. 29-50

Ivakhnenko A.G. (1968): "The Group Method of Data Handling - A Rival of the Method of Stochastic Approximation", Soviet Automatic Control, 13(3), pp. 43-55

Stepashko, V. Bulgakova, O. and Zosimov V. (2018): "Construction and Research of the Generalized Iterative GMDH Algorithm with Active Neurons", Advances in Intelligent Systems and Computing II, pp. 492-510 <doi:10.1007/978-3-319-70581-1_35>

Examples

```
set.seed(123)
x <- matrix(data = c(rnorm(1000)), ncol = 5, nrow = 200)
colnames(x) <- c("a", "b", "c", "d", "e")
y <- matrix(data = c(10 + x[, "a"] * x[, "e"]^3), ncol = 1)
colnames(y) <- "y"
x.test <- x[1:10, ]
y.test <- y[1:10]
x <- x[-c(1:10), ]
y <- y[-c(1:10)]

mod <- gmdh.gia(X = x, y = y, prune = 5, criteria = "PRESS")
pred <- predict(mod, x.test)
summary(sqrt((pred - y.test)^2))
```

---

gmdh.mia *GMDH MIA*

Description

Build a regression model performing GMDH MIA (Multilayered Iterative Algorithm).
For more information, please read the package's vignette.

Usage

```
gmdh.mia(X, y, prune = 150, criteria = c("PRESS", "test", "ICOMP"),
  x.test = NULL, y.test = NULL)
```

Arguments

X	matrix with N>3 columns and M rows, containing independent variables in the model. The data must not contain NAs
y	vector or matrix containing dependent variable in the model. The data must not contain NAs
prune	an integer whose recommended minimum value is the number of initial regressors. The maximum value will depend on the available RAM. It is recommended to work with the maximum value, but it can be computationally very expensive.

Prune is the selected number of neurons from layer i to layer i+1. The resulting layer i+1 has prune(prune-1)/2 neurons; for example with prune=150, the resulting nerurons will be 11.175

`criteria` GMDH external criteria. Values:

- PRESS: Predicted Residual Error Sum of Squares. It take into account all information in data sample and it is computed without recalculating of system for each test point.
- test: use x.test and y.test to estimate RMSE (Root Mean Squeare Errors).
- ICOMP: Index of Informational Complexity. Like PRESS, it is computed without recalculating of system.

`x.test` matrix with a sample randomly drawn from the initial data.
It is used when criteria = test.
This sample should not be included in X.

`y.test` vector or matrix with y values correspond with x.test values.

Value

An object of class mia.

References

Bozdogan, H. and Haughton, D.M.A. (1998): "Information complexity criteria for regression models", Computational Statistics & Data Analysis, 28, pp. 51-76 <doi: 10.1016/S0167-9473(98)00025-5>

Farlow, S.J. (1981): "The GMDH algorithm of Ivakhnenko", The American Statistician, 35(4), pp. 210-215. <doi:10.2307/2683292>

Hild, Ch. R. and Bozdogan, H. (1995): "The use of information-based model selection criteria in the GMDH algorithm", Systems Analysis Modelling Simulation, 20(1-2), pp. 29-50

Ivakhnenko A.G. (1968): "The Group Method of Data Handling - A Rival of the Method of Stochastic Approximation", Soviet Automatic Control, 13(3), pp. 43-55

Examples

```
set.seed(123)
x <- matrix(data = c(rnorm(1000)), ncol = 5, nrow = 200)
colnames(x) <- c("a", "b", "c", "d", "e")
y <- matrix(data = c(10 + x[, "a"] * x[, "e"]^3), ncol = 1)
colnames(y) <- "y"
x.test <- x[1:10, ]
y.test <- y[1:10]
x <- x[-c(1:10), ]
```

```
y <- y[-c(1:10)]

mod <- gmdh.mia(X = x, y = y, prune = 5, criteria = "PRESS")
pred <- predict(mod, x.test)
summary(sqrt((pred - y.test)^2))
```

`predict.combi` *Predict GMDH Combinatorial*

Description

Calculates GMDH Combinatorial model predictions for new data.

Usage

```
## S3 method for class 'combi'
predict(object, newdata, ...)
```

Arguments

`object`	an object of class 'combi'
`newdata`	matrix containing dependent variables in the model, wich the predictions are calculated.
`...`	other undocumented arguments

Value

A matrix with predictions.

Examples

```
set.seed(123)
x <- matrix(data = c(rnorm(1050)), ncol = 3, nrow = 350)
colnames(x) <- c("a", "b", "c")
y <- matrix(data = c(10 + x[, "a"] + x[, "b"]^2 + x[, "c"]^3), ncol = 1)
colnames(y) <- "y"
x.test <- x[1:10, ]
y.test <- y[1:10]
x <- x[-c(1:10), ]
y <- y[-c(1:10)]

mod <- gmdh.combi(X = x, y = y, criteria = "PRESS")
pred <- predict(mod, x.test)
summary(sqrt((pred - y.test)^2))
```

predict.combitwice *Predict GMDH Twice-Multilayered Combinatorial*

Description

Calculates GMDH Twice-Multilayered Combinatorial model predictions for new data.

Usage

```
## S3 method for class 'combitwice'
predict(object, newdata, ...)
```

Arguments

object	an object of class 'combitwice'
newdata	matrix containing dependent variables in the model, wich the predictions are calculated.
...	other undocumented arguments

Value

A matrix with predictions.

Examples

```
set.seed(123)
x <- matrix(data = c(rnorm(1050)), ncol = 3, nrow = 350)
colnames(x) <- c("a", "b", "c")
y <- matrix(data = c(10 + x[, "a"] + x[, "b"]^2 + x[, "c"]^3), ncol = 1)
colnames(y) <- "y"
x.test <- x[1:10, ]
y.test <- y[1:10]
x <- x[-c(1:10), ]
y <- y[-c(1:10)]

mod <- gmdh.combi.twice(X = x, y = y, criteria = "PRESS")
pred <- predict(mod, x.test)
summary(sqrt((pred - y.test)^2))
```

predict.gia *Predict GMDH GIA object*

Description

Calculates GMDH GIA Twice model predictions for new data.

Usage

```
## S3 method for class 'gia'
predict(object, newdata, ...)
```

Arguments

object	an object of class 'giatwice'
newdata	matrix containing dependent variables in the model, wich the predictions are calculated.
...	other undocumented arguments

Value

A matrix with predictions.

Examples

```
set.seed(123)
x <- matrix(data = c(rnorm(1000)), ncol = 5, nrow = 200)
colnames(x) <- c("a", "b", "c", "d", "e")
y <- matrix(data = c(10 + x[, "a"] * x[, "e"]^3), ncol = 1)
colnames(y) <- "y"
x.test <- x[1:10, ]
y.test <- y[1:10]
x <- x[-c(1:10), ]
y <- y[-c(1:10)]

mod <- gmdh.gia(X = x, y = y, prune = 5, criteria = "PRESS")
pred <- predict(mod, x.test)
summary(sqrt((pred - y.test)^2))
```

predict.mia *Predict GMDH MIA object*

Description

Calculates GMDH MIA model predictions for new data.

Usage

```
## S3 method for class 'mia'
predict(object, newdata, ...)
```

Arguments

object	an object of class 'mia'
newdata	matrix containing dependent variables in the model, wich the predictions are calculated.
...	other undocumented arguments

Value

A matrix with predictions.

Examples

```
set.seed(123)
x <- matrix(data = c(rnorm(1000)), ncol = 5, nrow = 200)
colnames(x) <- c("a", "b", "c", "d", "e")
y <- matrix(data = c(10 + x[, "a"] * x[, "e"]^3), ncol = 1)
colnames(y) <- "y"
x.test <- x[1:10, ]
y.test <- y[1:10]
x <- x[-c(1:10), ]
y <- y[-c(1:10)]

mod <- gmdh.mia(X = x, y = y, prune = 5, criteria = "PRESS")
pred <- predict(mod, x.test)
summary(sqrt((pred - y.test)^2))
```

Index