Package ‘PredictiveRegression’

February 19, 2015

Version 0.1-4
Date 2012-10-29
Title Prediction Intervals for Three Basic Statistical Models
Author Vladimir Vovk <vovk@cs.rhul.ac.uk> and Ilia Nouretdinov
<ilia@cs.rhul.ac.uk>
Maintainer Vladimir Vovk <vovk@cs.rhul.ac.uk>
Description Three prediction algorithms described in the paper
License file LICENSE
Repository CRAN
Date/Publication 2012-10-29 19:07:29
NeedsCompilation no
License_restricts_use yes

R topics documented:

gausspred ............................................................... 1
iidpred ................................................................. 3
mvapred ............................................................... 4

Index 6

__________________________

gausspred Gauss predictor

__________________________

Description

Prediction intervals based on the Gauss linear model
Usage

gausspred(train, test, epsilons = c(0.05, 0.01))

Arguments

train
Training set as a matrix of size \( N \times (K + 1) \). Each row describes an observation. Columns 1 to \( K \) are the explanatory variables, and column \( K + 1 \) is the response variables.

test
Test set as a matrix of size \( N_2 \times K \). Each row corresponds to an observation (but without the response variable). Columns 1 to \( K \) are the explanatory variables.

epsilons
Vector of several significance levels. Each significance level \( \text{epsilons}[j] \) is a number between 0 and 1. The default value is (5%, 1%).

Value

The output is a list of three elements.

output[[1]]
The matrix of lower bounds of prediction intervals. Its size is \( N_2 \times N_\epsilon \), where \( N_2 \) is the number of test observations and \( N_\epsilon \) is the number of significance levels. The element \( \text{output}[[1]][i,j] \) of \( \text{output}[[1]] \) is the lower bound \( a \) of the prediction interval \( [a, b] \) for the \( i \)th test observation and for the \( j \)th significance level \( \text{epsilons}[j] \) in the vector \( \text{epsilons} \).

output[[2]]
The matrix of upper bounds \( b \), with the same structure as \( \text{output}[[1]] \). Typically \( a = \text{output}[[1]][i,j] \) and \( b = \text{output}[[2]][i,j] \) are real numbers such that \( a \leq b \). Exceptions: \( a \) is allowed to be \( -\infty \) and \( b \) is allowed to be \( \infty \); the only case where \( a > b \) is \( a = \infty \) and \( b = -\infty \) (the empty prediction \( [a, b] \)).

output[[3]]
The termination code: 0 = normal termination; 1 = illegal parameters (the training and test sets have different numbers of explanatory variables); 2 = too few observations.

References


Examples

```r
train <- matrix(c(1,2,3,4, 2.01,2.99,4.01,4.99), nrow=4, ncol=2);
test <- matrix(c(0,10,20), nrow=3, ncol=1);
output <- gausspred(train, test, c(0.05,0.2));
print(output[[1]]);
print(output[[2]]);
```
 iidpred  

IID predictor

Description

Prediction intervals based on the IID model

Usage

iidpred(train,test,epsilons=c(0.05,0.01),ridge=0)

Arguments

- **train**: Training set as a matrix of size \( N \times K + 1 \). Each row describes an observation. Columns 1 to \( K \) are the explanatory variables, and column \( K + 1 \) is the response variables.

- **test**: Test set as a matrix of size \( N_2 \times K \). Each row corresponds to an observation (but without the response variable). Columns 1 to \( K \) are the explanatory variables.

- **epsilons**: Vector of several significance levels. Each significance level \( \text{epsilons}[j] \) is a number between 0 and 1. The default value is (5%,1%).

- **ridge**: Ridge coefficient, a nonnegative number. The default value is 0; setting it to a small positive constant might lead to more stable results.

Value

The output is a list of three elements.

- **output[[1]]**: The matrix of lower bounds of prediction intervals. Its size is \( N_2 \times N_\epsilon \), where \( N_2 \) is the number of test observations and \( N_\epsilon \) is the number of significance levels. The element \( \text{output[[1]]}[i,j] \) of \( \text{output[[1]]} \) is the lower bound \( a \) of the prediction interval \([a, b]\) for the \( i \)th test observation and for the \( j \)th significance level \( \text{epsilons}[j] \) in the vector \( \text{epsilons} \).

- **output[[2]]**: The matrix of upper bounds \( b \), with the same structure as \( \text{output[[1]]} \). Typically \( a = \text{output[[1]]}[i,j] \) and \( b = \text{output[[2]]}[i,j] \) are real numbers such that \( a \leq b \). Exceptions: \( a \) is allowed to be \(-\infty\) and \( b \) is allowed to be \( \infty \); the only case where \( a > b \) is \( a = \infty \) and \( b = -\infty \) (the empty prediction \([a, b]\)).

- **output[[3]]**: The termination code: 0 = normal termination; 1 = illegal parameters (the training and test sets have different numbers of explanatory variables); 2 = too few observations for all significance levels.
References


Examples

```r
train <- matrix(c(0,10,20,30, 1.01,10.99,21.01,30.99), nrow=4, ncol=2)
test <- matrix(c(5,15,25), nrow=3, ncol=1)
output <- iidpred(train,test,c(0.05,0.2),0.01)
print(output[[1]])
print(output[[2]])
```

mvapred

*MVA predictor*

Description

Prediction intervals based on the MVA model

Usage

```r
mvapred(train,test,epsilons=c(0.05,0.01),ridge=0)
```

Arguments

- **train**: Training set as a matrix of size $N \times K + 1$. Each row describes an observation. Columns 1 to $K$ are the explanatory variables, and column $K + 1$ is the response variables.
- **test**: Test set as a matrix of size $N_2 \times K$. Each row corresponds to an observation (but without the response variable). Columns 1 to $K$ are the explanatory variables.
- **epsilons**: Vector of several significance levels. Each significance level $\text{epsilons}[j]$ is a number between 0 and 1. The default value is (5%,1%).
- **ridge**: Ridge coefficient, a nonnegative number. The default value is 0; setting it to a small positive constant might lead to more stable results.

Value

The output is a list of three elements.

- **output[[1]]**: The matrix of lower bounds of prediction intervals. Its size is $N_2 \times N_\epsilon$, where $N_2$ is the number of test observations and $N_\epsilon$ is the number of significance levels. The element $\text{output[[1]][i,j]}$ of $\text{output[[1]]}$ is the lower bound $a$ of the prediction interval $[a, b]$ for the $i$th test observation and for the $j$th significance level $\text{epsilons}[j]$ in the vector $\text{epsilons}$. 
output[[2]] The matrix of upper bounds $b$, with the same structure as output[[1]]. Typically $a = \text{output[[1]]}[i,j]$ and $b = \text{output[[2]]}[i,j]$ are real numbers such that $a \leq b$. Exceptions: $a$ is allowed to be $-\infty$ and $b$ is allowed to be $\infty$; the only case where $a > b$ is $a = \infty$ and $b = -\infty$ (the empty prediction $[a,b]$).

output[[3]] The termination code: 0 = normal termination; 1 = illegal parameters (the training and test sets have different numbers of explanatory variables); 2 = too few observations.

References


Examples

```r
train <- matrix(c(0,10,20,30, 1.01,10.99,21.01,30.99), nrow=4,ncol=2);
test <- matrix(c(5,15,25), nrow=3, ncol=1);
output <- mvapred(train,test,c(0.05,0.2),0.01);
print(output[[1]]);
print(output[[2]]);
```
Index

*Topic models
  gausspred, 1
  iidpred, 3
  mvapred, 4
*Topic multivariate
  mvapred, 4
*Topic regression
  gausspred, 1
  iidpred, 3
  mvapred, 4

gausspred, 1

iidpred, 3

mvapred, 4