Package 'VIF'

February 19, 2015

Version 1.0

Date 2011-10-06

Title VIF Regression: A Fast Regression Algorithm For Large Data

Author Dongyu Lin <dongyu.lin@gmail.com>

Maintainer Dongyu Lin <dongyu.lin@gmail.com>

Description This package implements a fast regression algorithm for building linear model for large data as defined in the paper ``VIF-Regression: A Fast Regression Algorithm for Large Data (2011), Journal of the American Statistical Association, Vol. 106, No. 493: 232-247" by Dongyu Lin, Dean P. Foster, and Lyle H. Ungar.

License GPL (>= 2)

URL http://gosset.wharton.upenn.edu/~foster/auction/auction.html

Repository CRAN

Date/Publication 2012-10-29 08:58:07

NeedsCompilation no

R topics documented:

housingexp *Boston Housing Data with 3-Way Interactions*

Description

This data set is developed based on the original Boston Housing Data. We use `MEDV` as the repsonse `y`. In the matrix of variables `x`, we include all the other 13 variables, their second and third orders, their two-way and three-way interactions.

Usage

```
data(housingexp)
```

Source

http://stat.wharton.upenn.edu/~buja/STAT-541/boston.dat

syn *A Synthetic Data set For Testing VIF-Regression*

Description

This data set contains a data set that can be used to test the VIF-regression.

Usage

```
data(syn)
```

Format

A list contains:

`y` a numeric vector giving the response, 1000 x 1

`x` a numeric matrix of variables, containing 200 variables, 1000 x 200

`true` a vector of true variables that generate `y`

vif *Fitting Linear Models using VIF-Regression*

Description

vif selects variables for a linear model. It returns a subset of variables for building a linear model.

Usage

```
vif(y, x, w0 = 0.05, dw = 0.05, subsize = 200, trace = TRUE, mode = c("dense", "sparse"))
```

Arguments

y	the response.
x	an optional data frame or matrix containing the variables in the model.
w0	the initial wealth.
dw	the incremental wealth attained if a variable is included in the model.
subsize	the size of the subsample to approximate the variance inflation factor.
trace	logical. if TRUE a list containing current wealth, current test level, absolute t value and p-value for the current variable will be printed out.
mode	"dense" or "sparse", specifying one of the two alpha-investings that should be used. Default is "sparse".

Value

A list containing:

select	the chosen subset of variable.
modelmatrix	the model matrix that is ready for fitting a linear model.

Author(s)

Dongyu Lin <dongyu.lin@gmail.com>

References

Dongyu Lin, Dean P. Foster, and Lyle H. Ungar. (2011). VIF-Regression: A Fast Regression Algorithm for Large Data. *Journal of the American Statistical Association*, Vol. 106, No. 493: 232–247. http://gosset.wharton.upenn.edu/~foster/research/vif_jasa_final.pdf

The data sets used in the paper can be downloaded via following links:

Boston Housing Data: http://gosset.wharton.upenn.edu/~foster/auction/boston.csv

Bankruptcy Data: http://gosset.wharton.upenn.edu/~foster/auction/bankruptcy.csv

Call Center Data: http://gosset.wharton.upenn.edu/~foster/auction/calldata.tar.gz

Many others: http://gosset.wharton.upenn.edu/~foster/auction/auction.html.

Examples

```
data(syn);
vif.sel <- vif(syn$y, syn$x, trace = FALSE);
vif.sel$select;
syn$true;

data(housingexp);
colnames(housingexp$x);
vif.sel <- vif(housingexp$y, housingexp$x, w0 = 0.0005, dw = 0.005, subsize = 300, trace = FALSE);
```

Index