Package 'PLStests'

March 7, 2025

Type Package

Title Model Checking for High-Dimensional GLMs via Random Projections

Version 0.1.1

Maintainer Wen Chen <tlqdcw@gmail.com>

Description Provides methods for testing the goodness-of-fit of generalized linear models (GLMs) using random projections. It is specifically designed for high-dimensional scenarios where the number of predictors substantially exceeds the sample size. The statistical methodologies implemented in this package are detailed in the paper by Wen Chen and Falong Tan (2024, <doi:10.48550/arXiv.2412.10721>).

License GPL-3

Encoding UTF-8

LazyData true

Imports glmnet, harmonicmeanp, MASS, psych,stats

RoxygenNote 7.3.2

NeedsCompilation no

Author Wen Chen [aut, cre],
Jie Liu [aut],
Heng Peng [aut],
FaLong Tan [aut],
Lixing Zhu [aut]

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2025-03-07 13:30:02 UTC

Contents

PLStests *Model checking for high dimensional generalized linear models based on random projections*

Description

The function can test goodness-of-fit of a low- or high-dimensional generalized linear model (GLM) by detecting the presence of nonlinearity in the conditional mean function of y given x using the statistics proposed by paper xx. The outputs are p-value of statisitics.

Usage

```
PLStests(y, x, family, b0 = 2, np = 10)
```

Arguments

y	: y Input matrix with n rows, 1-dimensional response vector
x	: x Input matrix with n rows, each a p-dimensional observation vector.
family	: Must be "gaussian" or "binomial" for linear or logistic regression model.
b0	: a paramter to set bindwith, the default value may better for real data analysing.
np	: the number of random projections.

Value

a list with five parameters returned. h stand for b_0. T_alpha: the p value of our statistics by random projection. T_beta: the p value of our statistic by estimated projection. T_cauchy and T_hmp are p value of two combinational method proposed by Liu and Xie (2020) and Wilson (2019) respectively. each method combines p values of np random projections. when the estimated projection is zero, the value set be NA.

References

Chen, W., Liu, J., Peng, H., Tan, F., & Zhu, L. (2024). Model checking for high dimensional generalized linear models based on random projections. arXiv [Stat.ME]. Retrieved from http://arxiv.org/abs/2412.10721

Examples

```
set.seed(100)
data("sonar_mines")
x = sonar_mines[,-1]
y = sonar_mines$y

## make y as 0 or 1 for logistic regression
class1 = "R"
class2 ="M"
y = as.character(y)
y[y==class1]=1
```

```
y[y==class2]=0
y = as.numeric(y)
y = matrix(y,ncol = 1)

## scale x  and make data to be matrix
data_test_x = x
data_test_x = as.matrix(data_test_x)
data_test_y = as.matrix(y)
data_test_x = scale(data_test_x)
PLStests(data_test_y,data_test_x,family="binomial")
```

`sonar_mines` *Example Dataset: sonar_mines*

Description

we evaluate the proposed tests through an analysis of a classification task aimed at distinguishing between sonar signals bounced off a metal cylinder and those bounced off a roughly cylindrical rock. The dataset is available at https://archive.ics.uci.edu/dataset/151/connectionist+bench+sonar+mines+vs+rocks.

Usage

```
sonar_mines
```

Format

A data frame with 208 rows and 61 variables:

V1,V2,V3,V4,V5,V6,V7,V8,V9,V10,V11,V12,V13,V14,V15,V16,V17,V18,V19,V20,V21,V22,V23, V24,V25,V26,V27,V
: Numeric sonar signal attributes (frequencies).

y Class label, a factor with levels 'Mine' and 'Rock'.

Source

from https://archive.ics.uci.edu/dataset/151/connectionist+bench+sonar+mines+vs+rocks

Index