Package 'moose'

September 9, 2022

Title Mean Squared Out-of-Sample Error Projection

Version 0.0.1

Description Projects mean squared out-of-sample error for a linear regression based upon the methodology developed in Rohlfs (2022) <doi:10.48550/arXiv.2209.01493>. It consumes as inputs the lm object from an estimated OLS regression (based on the ``training sample'') and a data.frame of out-of-sample cases (the ``test sample'') that have non-missing values for the same predictors. The test sample may or may not include data on the outcome variable; if it does, that variable is not used. The aim of the exercise is to project what what mean squared out-of-sample error can be expected given the predictor values supplied in the test sample. Output consists of a list of three elements: the projected mean squared out-of-sample error, the projected out-of-sample R-squared, and a vector of out-of-sample ``hat'' or ``leverage'' values, as defined in the paper.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.2.1

NeedsCompilation no

Author Chris Rohlfs [aut, cre] (<https://orcid.org/0000-0001-7714-9231>)

Maintainer Chris Rohlfs <car2228@columbia.edu>

Repository CRAN

Date/Publication 2022-09-09 08:20:02 UTC

R topics documented:

moose *moose: mean squared out-of-sample error projection*

Description

This function projects the mean squared out-of-sample error for a linear regression

Usage

```
moose(reg, dataset)
```

Arguments

reg	an lm object containing the regression to project out-of-sample
dataset	a data.frame containing new cases for out-of-sample projection

Value

mse	Projected mean squared out-of-sample error
R2o	Projected out-of-sample R-squared
hat	Leverage for each out-of-sample observation. For each i, this is the sum of the squared elements of xi [X'X]^-1 X', where X is the predictor matrix from the training sample.

Examples

```
# set the seed for reproducibility of the example
set.seed(04251978)
# randomly generate 100 observations of data
mydata <- data.frame(x1=rnorm(100),x2=rnorm(100),x3=rnorm(100))
# true outcome variable is y = x1 + x2 + x3 + e
y <- mydata$x1 + mydata$x2 + mydata$x3 + rnorm(100)
# regression with the first 25 observations from the dataset
reg <- lm(y ~ x1 + x2 + x3,data=cbind(y,mydata)[1:25,])
# using the predictor values from the first 25 observations,
# project the out-of-sample error we can expect in the case of
# "non-stochastic" predictors whose values are the same in the
# test sample as in the training sample.
# note that mydata does not include the outcome variable.
same.predictor.values.error <- moose(reg,mydata[1:25,])
# by comparison, the in-sample R-squared value observed
# in training is:
summary(reg)$r.squared
# using the predictor values from the next 75 obsevervations,
# project the out-of-sample error we can expect in the case
# of stochastic predictors whose values potentially differ
# from those used in training.
new.predictor.values.error <- moose(reg,mydata[26:100,])
```

```
# by comparison, the actual mse and out-of-sample R-squared value
# obtained from observations 26-100 of this random sample are:
mse <- mean((y[26:100]-predict(reg,mydata[26:100,]))^2)
mse
m.total.sqs <- mean((y[26:100]-mean(y[26:100]))^2)
r2o <- 1-mse/m.total.sqs
r2o
```

Index