Package 'olr'

October 14, 2022

Title Optimal Linear Regression

Version 1.1

Date 2020-01-07

Author Mathew Fok

Maintainer Mathew Fok <mfok@stevens.edu>

Description The optimal linear regression olr(), runs all the possible combinations of linear regression equations. The olr() returns the equation which has the greatest adjusted R-squared term or the greatest R-squared term based on the user's discretion. Essentially, the olr() returns the best fit equation out of all the possible equations. R-squared increases with the addition of an explanatory variable whether it is 'significant' or not, thus this was developed to eliminate that conundrum. Adjusted R-squared is preferred to overcome this phenomenon, but each combination will still produce different results and this will return the best one. Complimentary functions are included which list all of the equations, all of the equations in ascending order, a function to give the user a specific model's summary, and the list of adjusted R-squared terms & R-squared terms. A 'Python' version is available at: <https://pypi.org/project/olr/>.

URL https://github.com/MatHatter, https://pypi.org/project/olr/

Encoding UTF-8

Depends R (>= 2.10)

LazyData true

Imports plyr, utils, stats

RoxygenNote 6.1.1.9000

NeedsCompilation no

Repository CRAN

Date/Publication 2020-01-08 18:10:02 UTC

R topics documented:

---

olr *olr: Optimal Linear Regression*

Description

The main olr() runs all of the possible linear regression equation combinations, which are all of the combinations of dependent variables respect to the independent variable. In essence, the olr() returns the best fit linear regression model. The user can prompt the olr() to return either the best fit statistical summary of either the greatest adjusted R-squared, or the greatest R-squared term. R-squared increases with the addition of an explanatory variable whether it is 'significant' or not, thus this was developed to eliminate that conundrum. Adjusted R-squared is preferred to overcome this phenomenon, but each combination will still produce different results and this will return the best one.

Usage

```
olr(dataset, responseName = NULL, predictorNames = NULL,
  adjr2 = TRUE)

olrmodels(dataset, responseName = NULL, predictorNames = NULL)

olrformulas(dataset, responseName = NULL, predictorNames = NULL)

olrformulaorder(dataset, responseName = NULL, predictorNames = NULL)

adjr2list(dataset, responseName = NULL, predictorNames = NULL)

r2list(dataset, responseName = NULL, predictorNames = NULL)
```

Arguments

`dataset`	is defined by the user and points to the name of the dataset that is being used.
`responseName`	the response variable name defined as a string. For example, it represents a header in the data table.
`predictorNames`	the predictor variable or variables that are the terms that are to be regressed against the `responseName`. Place desired headers from the `dataset` in here as a character vector.
`adjr2`	`adjr2` = `TRUE` returns the regression summary for the maximum adjusted R-squared term. `adjr2` = `FALSE` returns the regression summary for the maximum R-squared term.

Details

Complimentary functions below follow the format: function(dataset, responseName = NULL, predictorNames = NULL)

olrmodels: returns the list of models accompanied by the coefficients. After typing in `olrmodels(dataset, responseName, predictorNames)` type the desired summary number to the right of the comma in the brackets: `[,x]` where x equals the desired summary number. For example, `olrmodels(dataset, responseName, predictorNames)[,8]`

olrformulas: returns the list of olr() formulas

olrformulasorder: returns the formulas with the predictors (dependent variables) in ascending order

adjr2list: list of the adjusted R-squared terms

r2list: list of the R-squared terms

When `responseName` and `predictorNames` are `NULL`, then the first column in the `dataset` is set as the `responseName` and the remaining columns are the `predictorNames`.

A 'Python' version is available at <https://pypi.org/project/olr>.

Value

The regression summary for the adjusted R-squared or the R-squared, specified with `TRUE` or `FALSE` in the olr().

Examples

```
file <- system.file("extdata", "oildata.csv", package = "olr", mustWork = TRUE)
oildata <- read.csv(file, header = TRUE)

dataset <- oildata
responseName <- 'OilPrices'
predictorNames <- c('SP500', 'RigCount', 'API', 'Field_Production', 'OperableCapacity', 'Imports')

olr(dataset, responseName, predictorNames, adjr2 = TRUE)
```

Index