

Package ‘AutoStepwiseGLM’

November 14, 2018

Type Package

Title Builds Stepwise GLMs via Train and Test Approach

Version 0.2.0

Author Aaron England <aaron.england24@gmail.com>

Maintainer Aaron England <aaron.england24@gmail.com>

Description Randomly splits data into testing and training sets. Then, uses stepwise selection to fit numerous multiple regression models on the training data, and tests them on the test data. Returned for each model are plots comparing model Akaike Information Criterion (AIC), Pearson correlation coefficient (r) between the predicted and actual values, Mean Absolute Error (MAE), and R-Squared among the models. Each model is ranked relative to the other models by the model evaluation metrics (i.e., AIC, r , MAE, and R-Squared) and the model with the best mean ranking among the model evaluation metrics is returned. Model evaluation metric weights for AIC, r , MAE, and R-Squared are taken in as arguments as `aic_wt`, `r_wt`, `mae_wt`, and `r_squ_wt`, respectively. They are equally weighted as default but may be adjusted relative to each other if the user prefers one or more metrics to the others, Field, A. (2013, ISBN:978-1-4462-4918-5).

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 6.1.0

Depends caret, formula.tools

NeedsCompilation no

Repository CRAN

Date/Publication 2018-11-14 11:20:06 UTC

R topics documented:

<code>backwd_stepwise_glm</code>	2
<code>fwd_stepwise_glm</code>	3

Index	5
--------------	---

backwd_stepwise_glm *Automated Backward Stepwise GLM*

Description

Takes in a dataframe and the dependent variable (in quotes) as arguments, splits the data into testing and training, and uses automated backward stepwise selection to build a series of multiple regression models on the training data. Each model is then evaluated on the test data and model evaluation metrics are computed for each model. These metrics are provided as plots. Additionally, the model metrics are ranked and average rank is taken. The model with the best average ranking among the metrics is displayed (along with its formula). By default, metrics are all given the same relative importance (i.e., weights) when calculating average model metric rank, but if the user desires to give more weight to one or more metrics than the others they can specify these weights as arguments (default for weights is 1). As of v 0.2.0, only the family = gaussian(link = 'identity') argument is provided within the glm function.

Usage

```
backwd_stepwise_glm(data, dv, aic_wt = 1, r_wt = 1, mae_wt = 1,
  r_squ_wt = 1, train_prop = 0.7, random_seed = 7)
```

Arguments

data	A dataframe with one column as the dependent variable and the others as independent variables
dv	The column name of the (continuous) dependent variable (must be in quotes, i.e., 'Dependent_Variable')
aic_wt	Weight given to the rank value of the AIC of the model fitted on the training data (used when calculating mean model performance, default = 1)
r_wt	Weight given to the rank value of the Pearson Correlation between the predicted and actual values on the test data (used when calculating mean model performance, default = 1)
mae_wt	Weight given to the rank value of Mean Absolute Error on the test data (used when calculating mean model performance, default = 1)
r_squ_wt	Weight given to the rank value of R-Squared on the test data (used when calculating mean model performance, default = 1)
train_prop	Proportion of the data used for the training data set
random_seed	Random seed to use when splitting into training and testing data

Value

This function returns a plot for each metric by model and the best overall model with the formula used when fitting that model

Examples

```
dt <- mtcars
stepwise_model <- backwd_stepwise_glm(data = dt,
                                     dv = 'mpg',
                                     aic_wt = 1,
                                     r_wt = 0.8,
                                     mae_wt = 1,
                                     r_squ_wt = 0.8,
                                     train_prop = 0.6,
                                     random_seed = 5)

stepwise_model
```

fwd_stepwise_glm

*Automated Forward Stepwise GLM***Description**

Takes in a dataframe and the dependent variable (in quotes) as arguments, splits the data into testing and training, and uses automated forward stepwise selection to build a series of multiple regression models on the training data. Each model is then evaluated on the test data and model evaluation metrics are computed for each model. These metrics are provided as plots. Additionally, the model metrics are ranked and average rank is taken. The model with the lowest average ranking among the metrics is displayed (along with its formula). By default, metrics are all given the same relative importance (i.e., weights) when calculating average model metric rank, but if the user desires to give more weight to one or more metrics than the others they can specify these weights as arguments (default for weights is 1). As of v 0.2.0, only the family = gaussian(link = 'identity') argument is provided within the glm function.

Usage

```
fwd_stepwise_glm(data, dv, aic_wt = 1, r_wt = 1, mae_wt = 1,
                 r_squ_wt = 1, train_prop = 0.7, random_seed = 7)
```

Arguments

data	A dataframe with one column as the dependent variable and the others as independent variables
dv	The column name of the (continuous) dependent variable (must be in quotes, i.e., 'Dependent_Variable')
aic_wt	Weight given to the rank value of the AIC of the model fitted on the training data (used when calculating mean model performance, default = 1)
r_wt	Weight given to the rank value of the Pearson Correlation between the predicted and actual values on the test data (used when calculating mean model performance, default = 1)
mae_wt	Weight given to the rank value of Mean Absolute Error on the test data (used when calculating mean model performance, default = 1)

Index

`backwd_stepwise_glm`, 2

`fwd_stepwise_glm`, 3