Package 'MarkovLS'

June 1, 2022

Type Package

Title Suboptimal learning via the variable order Markov Chain Learning Space

Version 0.1

Author Diego Marcondes

Maintainer Diego Marcondes <dmarcondes@ime.usp.br>

Description Performs a suboptimal learning via Learning Spaces to solve the problem of forecasting a binary sequence. More details can be found at PhD thesis ``A data-driven systematic, consistent and feasible approach to Model Selection'', Universidade de São Paulo,2022.

License GPL v3.0

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

Depends R (>= 2.10)

Imports ggplot2 (>= 3.3.6),
tidyverse (>= 1.3.1)

References Diego Marcondes. A data-driven systematic, consistent and feasible approach to Model Selection. \{ }emph{PhD thesis}. Universidade de São Paulo, 2022

R topics documented:

---

bitcoin   *Bitcoin history*

Description

Data set containing the bitcoin value history from April 30th 2013 to April 6th 2022.

Usage

    bitcoin

Format

An object of class data.frame with 3264 rows and 7 columns.

Details

The data is divided into training, validation and test samples.

Value

date	The date
open	Open bitcoin value in US dollars in the respective date
close	Close bitcoin value in US dollars in the respective date
variation	Percentage variation from open to close value
y	Variation positive (1) or negative (0)
sample	If the respetive date is part of the training, validation or test sample
x	The past variation (positive or negative) on the prior 30 days

References

Diego Marcondes. A data-driven systematic, consistent and feasible approach to Model Selection. *PhD thesis*. Universidade de São Paulo, 2022

ERM_markov   *ERM estimative of Markov model of length k*

Description

Estimate the hypothesis that minimizes the empirical classification error under the Markov model of length k.

Usage

```
ERM_markov(train, k)
```

Arguments

train	A data frame with the training data. Should have columns named x and y representing the input and output strings.
k	Length of Markov Chain.

Details

Receives the sequence values y and its past x, and returns the predictor of length k that minimizes the classification error in the sample x,y.

Value

A list containing the table with the estimated hypothesis and a function to predict the output of a given input.

References

Diego Marcondes. A data-driven systematic, consistent and feasible approach to Model Selection. *PhD thesis*. Universidade de São Paulo, 2022

Examples

```
ERM_markov(bitcoin,2)
```

---

error_models   *Predict output from multiple models*

Description

Predict the output for each observation in a sample and calculate the error for multiple models.

Usage

```
error_models(sample, predict, names, simulate = F, init_value = 1000)
```

Arguments

sample	A data frame with the sample data. Should have columns named x and y representing the input and output strings, and columns variation and date containing the daily variation and date for simulation.
predict	A list of functions which, given an input string, returns a predicted output.
names	A vector containing the names of the models which generated the predict functions.
simulate	Logical. Wheter to simulate the data and plot it.
init_value	Initial value of the account for simulation.

Details

Given a list of functions from models estimated via search_LS function, compare them by calculating errors and ploting the results.

Value

A list containing the sample data frame with new columns for prediction and simulation, the error of each model, a plot with the simularion and the data used in the plot.

References

Diego Marcondes. A data-driven systematic, consistent and feasible approach to Model Selection. *PhD thesis*. Universidade de São Paulo, 2022

Examples

```
mod1 <- search_LS(train = bitcoin[bitcoin$sample == "Train",],
val = bitcoin[bitcoin$sample == "Validation",],
k = 30,max_leaves = 8,max_init = log(8,2),verbose = FALSE)

mod2 <- search_LS(train = bitcoin[bitcoin$sample == "Train",],
val = bitcoin[bitcoin$sample == "Validation",],
k = 30,max_leaves = 16,max_init = log(16,2),verbose = FALSE)
```

```
mod3 <- search_LS(train = bitcoin[bitcoin$sample == "Train",],
val = bitcoin[bitcoin$sample == "Validation",],
k = 30,max_leaves = 32,max_init = log(32,2),verbose = FALSE)

e <- error_models(sample = bitcoin[bitcoin$sample == "Test",],
predict = list(mod1$predict,mod2$predict,mod3$predict),names = c("d8","d16","d32"),
simulate = TRUE)
```

---

search_LS   *Search the Markov Boolean Learning Space*

Description

Performs an stochastic search on the Markov Boolean Learning Space.

Usage

```
search_LS(
  train,
  val,
  k,
  max_leaves = 50,
  max_init,
  verbose = T,
  error = "simple"
)
```

Arguments

train	A data frame with the training data. Should have columns named x and y representing the input and output strings.
val	A data frame with the validation data. Should have columns named x and y representing the input and output strings.
k	Maximum length of Markov Chain.
max_leaves	Maximum number of leaves in a node.
max_init	Maximum length of initial node to try.
verbose	Logical. Whether to print algorithm progress.
error	Wheter to consider the simple error or the weighted error

Details

Performs the suboptimal algorithm discussed in the reference.

Value

A list containing the validation error of the estimated node, a table with the estimated hypothesis, a function to predict the output of a given input and the algorithm processing time.

References

Diego Marcondes. A data-driven systematic, consistent and feasible approach to Model Selection. *PhD thesis*. Universidade de São Paulo, 2022

Examples

```
mod <- search_LS(train = bitcoin[bitcoin$sample == "Train",],
val = bitcoin[bitcoin$sample == "Validation",],
k = 30,max_leaves = 8,max_init = log(8,2),verbose = FALSE)
```

Index