Package 'jenga'

May 23, 2022

Type Package

Title Fast Extrapolation of Time Features using K-Nearest Neighbors

Version 1.2.0

Author Giancarlo Vercellino

Maintainer Giancarlo Vercellino <giancarlo.vercellino@gmail.com>

Description Fast extrapolation of univariate and multivariate time features using K-Nearest Neighbors. The compact set of hyper-parameters is tuned via grid or random search.

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

Depends R (>= 4.1)

Imports purrr (>= 0.3.4), abind (>= 1.4-5), ggplot2 (>= 3.3.5), readr (>= 2.1.2), lubridate (>= 1.4.0), narray (>= 0.4.1.1), imputeTS (>= 3.2), scales (>= 1.1.1), tictoc (>= 1.0.1), modeest (>= 2.4.0), moments (>= 0.14), philentropy (>= 0.5.0), greybox (>= 1.0.1), Rfast (>= 2.0.6), dplyr(>= 1.0.7)

URL https://rpubs.com/giancarlo_vercellino/jenga

NeedsCompilation no

Repository CRAN

Date/Publication 2022-05-23 19:40:02 UTC

R topics documented:

covid_in_europe *covid_in_europe data set*

Description

A data frame with with daily and cumulative cases of Covid infections and deaths in Europe since March 2021.

Usage

```
covid_in_europe
```

Format

A data frame with 5 columns and 163 rows.

Source

www.ecdc.europa.eu

jenga *jenga: automatic projections of time features using KNN*

Description

Automatic projections of time features using KNN

Usage

```
jenga(
  ts,
  seq_len = NULL,
  k = NULL,
  method = NULL,
  kernel = NULL,
  ci = 0.8,
  n_windows = 10,
  mode = NULL,
  n_sample = 30,
  search = "random",
  dates = NULL,
  error_scale = "naive",
  error_benchmark = "naive",
  seed = 42
)
```

Arguments

ts	A data frame with time features on columns
seq_len	Positive integer. Time-step number of the projected sequence
k	Positive integer. Number of neighbors to consider when applying kernel average. Min number is 3. Default: NULL (automatic selection).
method	Positive integer. Distance method for calculating neighbors. Possibile options are: "euclidean", "manhattan", "canberra1", "minimum", "maximum", "minkowski", "bhattacharyya", "kullback_leibler", "jensen_shannon". Default: NULL (automatic selection).
kernel	String. Distribution used to calculate kernel densities. Possible options are: "norm", "cauchy", "logis", "unif", "t". Default: NULL (automatic selection).
ci	Confidence interval. Default: 0.8
n_windows	Positive integer. Number of validation tests to measure/sample error. Default: 10.
mode	String. Sequencing method: deterministic ("segmented"), or non-deterministic ("sampled"). Default: NULL (automatic selection).
n_sample	Positive integer. Number of samples for grid or random search. Default: 30.
search	String. Two option available: "grid", "random". Default: "random".
dates	Date. Vector with dates for time features.
error_scale	String. Scale for the scaled error metrics. Two options: "naive" (average of naive one-step absolute error for the historical series) or "deviation" (standard error of the historical series). Default: "naive".
error_benchmark	String. Benchmark for the relative error metrics. Two options: "naive" (sequential extension of last value) or "average" (mean value of true sequence). Default: "naive".
seed	Positive integer. Random seed. Default: 42.

Value

This function returns a list including:

- exploration: list of all not-null models, complete with predictions, test metrics, prediction stats and plot
- history: a table with the sampled models, hyper-parameters, validation errors, weighted average rank
- best_model: results for the best model in term of weighted average rank, including:
  - predictions: min, max, q25, q50, q75, quantiles at selected ci, mean, sd, mode, skewness, kurtosis, IQR to range, risk ratio, upside probability and divergence for each point fo predicted sequences
  - testing_errors: training and testing errors for one-step and sequence for each ts feature (me, mae, mse, rmsse, mpe, mape, rmae, rrmse, rame, mase, smse, sce, gmrae)
  - pred_scores: a measure of prediction interval fit for each point in predicted sequence (value range from 0, out of boundaries, to 1, close to the median)
- time_log

Author(s)

Giancarlo Vercellino <giancarlo.vercellino@gmail.com>

See Also

Useful links:

- https://rpubs.com/giancarlo_vercellino/jenga

Examples

```
jenga(covid_in_europe[, c(2, 3)], n_sample = 1)
jenga(covid_in_europe[, c(4, 5)], n_sample = 1)
```

Index