

Package ‘jenga’

October 21, 2021

Type Package

Title Fast Extrapolation of Time Features using K-Nearest Neighbors

Version 1.0.0

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 (>= 3.6)

Imports purrr, abind, ggplot2, readr, stringr, lubridate, narray, imputeTS, scales, tictoc, modeest, moments, philentropy

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

NeedsCompilation no

Repository CRAN

Date/Publication 2021-10-21 17:40:02 UTC

R topics documented:

covid_in_europe	2
engine	2
hood	3
jenga	4
Index	6

covid_in_europe	<i>covid_in_europe data set</i>
-----------------	---------------------------------

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

engine	<i>support functions for jenga</i>
--------	------------------------------------

Description

support functions for jenga

Usage

```
engine(ts, seq_len, k, method, kernel, error_measurement, deriv, mode)
```

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.
method	Positive integer. Distance method for calculating neighbors. Possible options are: "euclidean", "manhattan", "chebyshev", "sorensen", "gower", "soergel", "kulczynski_d", "canberra", "lorentzian", "intersection", "non-intersection", "wave-hedges", "czekanowski", "motyka", "kulczynski_s", "tanimoto", "ruzicka", "inner_product", "harmonic_mean", "cosine", "hassebrook", "jaccard", "dice", "fidelity", "bhattacharyya", "squared_chord", "squared_euclidean", "pearson", "neyman", "squared_chi", "prob_symm", "divergence", "clark", "additive_symm", "taneja", "kumar-johnson", "avg".

kernel	String. DIstribution used to calculate kernel densities. Possible options are: "norm", "cauchy", "logis", "unif", "t", "exp", "lnorm".
error_measurement	Logical. TRUE for measuring validation error. FALSE otherwise.
deriv	Integer vector. Number of differentiation operations to perform for each original time feature. 0 = no change; 1: one diff; 2: two diff.
mode	String. Sequencing method: deterministic ("segmented"), or non-deterministic ("sampled").

Author(s)

Giancarlo Vercellino <giancarlo.vercellino@gmail.com>

hood *support functions for jenga*

Description

support functions for jenga

Usage

```
hood(ts, seq_len, k, method, kernel, ci, deriv, n_windows, mode, dates)
```

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.
method	Positive integer. Distance method for calculating neighbors. Possible options are: "euclidean", "manhattan", "chebyshev", "sorensen", "gower", "soergel", "kulczynski_d", "canberra", "lorentzian", "intersection", "non-intersection", "wave-hedges", "czekanowski", "motyka", "kulczynski_s", "tanimoto", "ruzicka", "inner_product", "harmonic_mean", "cosine", "hassebrook", "jaccard", "dice", "fidelity", "bhattacharyya", "squared_chord", "squared_euclidean", "pearson", "neyman", "squared_chi", "prob_symm", "divergence", "clark", "additive_symm", "taneja", "kumar-johnson", "avg".
kernel	String. DIstribution used to calculate kernel densities. Possible options are: "norm", "cauchy", "logis", "unif", "t", "exp", "lnorm".
ci	Confidence interval.
deriv	Integer vector. Number of differentiation operations to perform for each original time feature. 0 = no change; 1: one diff; 2: two diff.
n_windows	Positive integer. Number of validation tests to measure/sample error.
mode	String. Sequencing method: deterministic ("segmented"), or non-deterministic ("sampled").
dates	Date. Vector with dates for time features.

Author(s)

Giancarlo Vercellino <giancarlo.vercellino@gmail.com>

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,
  deriv = NULL,
  n_windows = 10,
  mode = NULL,
  n_sample = 30,
  search = "random",
  dates = NULL,
  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. Possible options are: "euclidean", "manhattan", "chebyshev", "sorensen", "gower", "soergel", "kulczynski_d", "canberra", "lorentzian", "intersection", "non-intersection", "wave-hedges", "czekanowski", "motyka", "tanimoto", "ruzicka", "inner_product", "harmonic_mean", "cosine", "hassebrook", "jaccard", "dice", "fidelity", "bhattacharyya", "squared_chord", "squared_euclidean", "pearson", "neyman", "squared_chi", "prob_symm", "divergence", "clark", "additive_symm", "taneja", "kumar-johnson", "avg". Default: NULL (automatic selection).
kernel	String. Distribution used to calculate kernel densities. Possible options are: "empirical", "norm", "cauchy", "logis", "unif", "t", "exp", "lnorm". Default: NULL (automatic selection).

ci	Confidence interval. Default: 0.8
deriv	Integer vector. Number of differentiation operations to perform for each original time feature. 0 = no change; 1: one diff; 2: two diff.
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.
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: results for the best model in term of weighted average rank, including:
 - errors: training and testing errors for one-step and sequence for each ts feature (rmse, mae, mdae, mpe, mape, smape)
 - predictions: min, max, q25, q50, q75, quantiles at selected ci, mean, sd for each ts feature
 - pred_stats: for each predicted time feature, IQR to range, Kullback-Leibler Divergence (compared to previous point in time), upside probability (compared to previous point in time), both averaged across all points in time and compared between the terminal and the first point in the prediction sequence.
- 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 = 3)
jenga(covid_in_europe[, c(4, 5)], n_sample = 3)
```

Index

* datasets

covid_in_europe, [2](#)

covid_in_europe, [2](#)

engine, [2](#)

hood, [3](#)

jenga, [4](#)

jenga-package (jenga), [4](#)