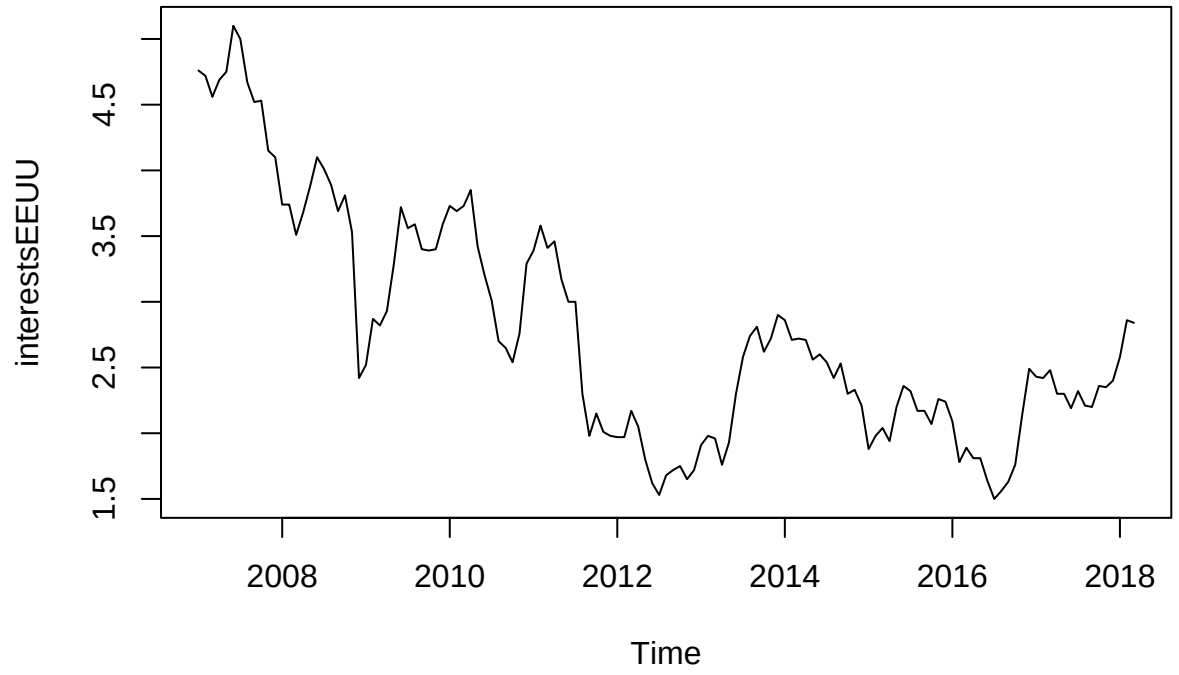
Time series forecasting with classical Machine Learning algorithms

Usue Mori

3 de noviembre de 2022

In this tutorial we will learn to perform forecasts on time series using classical regression algorithms from machine learning. As an example, we are going to use a time series that represents the long-term interest rates in E.E.U.U from January 2007 to March 2018. [1]

```
interestsEEUU <- read.table("./Data/MEI_FIN_23092019161922426.csv", sep=",", header=TRUE)
interestsEEUU <- ts(interestsEEUU[interestsEEUU$Country=="United States"
                    & interestsEEUU$Tipo=="IRLT", c("Value")],
                    start=2007, frequency=12)
plot(interestsEEUU)
```



1. Exercises

- In order to evaluate our models in a honest way, we will use a hold-out method, but we will respect the temporal order to the data. This means that we will use the first 70 % data points (first 8 years) of the time series to train the models are the remaining 30 % to test them.

[1] Time series extracted from `http://rwanjohi.rbind.io/2018/04/05/time-series-forecasting-using-lstm-in-r/`).

NOTE: The `caret` package proposes specific methods to generate train/test splits in time series. It also proposes ways to perform cross-validation with temporal data. For more information go to `https://otexts.com/fpp2/accuracy.html`.

- As we have seen in the theory, classical regression algorithms are not designed to perform predictions on time series, they are designed for tabular data. So, in order to apply them to temporal data, we must transform our time series into a tabular form, using a sliding window. Program a function in R that will format the data in R so that it is suitable for classical regression algorithms. This function will take two parameters as input: the time series and the size of the sliding window.
- Apply different regression algorithms from the `caret` package to the formatted time series (use only the training set for fitting the models).
- Evaluate the models by using the test set and computing the RMSE and represent the predictions of the best model together with the time series.

2. Extra work

- Repeat the previous exercise to perform 3-step ahead predictions.

3. References

- Kuhn, M. (2008). Caret package. Journal of Statistical Software, 28(5)