Time series forecasting project

Oussama DJEDIDI

November 30, 2021

This project focus on the piratical aspect of time series forecasting. During the courses, several techniques were explaining in order of utility and complexity. In this project, you will have to apply this knowledge on an industry dataset.

Target forecasting

1. **Prepare the data**

   - Load the "`TimeSeriesForecasting.csv`" file. It contains three names columns and 24 unnamed ones (with the names Column 1 to 24). the time index is called `time`, whereas the variable to predict is called `targets`. The column `forecastedTagets` are predictions made by the the industrial against whom you will compare your predictions.

   - Perform all the necessary data analysis and cleaning on your dataset to ready it for modeling an prediction.

   - Split the data into training (75%) and test (25%) data.(data is ordered, be careful how you split it).

2. **Analyze your times series**

3. **Start predicting**

   - **Smoothing techniques:** Use Moving average and exponential smoothing to predict your targets. Discuss your results.

   - **AutoRegressive Models:** Use the AR, ARMA, and ARIMA models to predict your targets.

     - These model have a *Look back* parameter. Create a function to reshape you data according to a look back parameter (*Default =1*).

     - Train and test your model.

     - Discuss your results.

   - **Optional - Deep learning:** Use LSTM neural networks.

     - Be careful: These model too have a *Look back* parameter.

     - Train and test your model and discuss your results.

Comparing Models

1. Compare the predictions from your models with the ones present in **forecastedTagets**.

2. Would using the other columns (Column 1 to 24) enhance your results? Can you test with ARMAX and ARIMAX models?

3. What type of models do you think the industrial engineers used?

4. **Optional:**

   - Can you guess which dataset is this ?

   - Can you propose a model—not proposed here—that can achieve better results?

Remarks

- The final report with results should be submitted before the deadling as Jupyter Notebook named: **name_firstname.ipynb** (or **name1_firstname1-name2_firstname2.ipynb**, if you are doing this project in groups of two).

- The report (Jupyter file) should be structured, have an easily readable code with explanatory comments. It also should contain markdown cells explaining and discussing the steps you're undertaking.

- In every training and test you should:

  - Plot the targets and predictions

  - Plot and analyze the prediction errors (Study the distribution, the characteristics (mean, min, max...), test for heteroscedasticity...)

Deadline for submission: December 10^{10}, 2021.