FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

SI 20/21 - Predicting the Output for a Multistage Factory Process using Machine Learning

Problem Description

A multistage manufacturing process refers to a system encompassing multiple components, stations or stages required to finish the final product, being very common in modern manufacturing. Typically, the quality of the final product depends on the complex interactions between different stages. Thus, the quality characteristics at one stage are not only influenced by local variations at that stage, but also by variations propagated from upstream stages.

In this context predictive models based on machine learning can be used in the development of real time process controllers, anomaly detection, quality control, etc.

Data

The dataset contains data sampled every minute from one production run of a multistage continuous flow manufacturing process spanning several hours. The focus is put on three machines operating in parallel at the first stage, which then feed their outputs into a step that combines the flows.

The output from the combiner step is measured in 6 different locations. The **goal** is to predict the measurements of the output from this stage, based on the data from the shopfloor environment and operations upstream.

Required Material

The assignment will be developed using **Google Colaboratory** to ensure that everyone has access to the same computational resources, without requiring additional configuration effort.

https://colab.research.google.com/

Google Colab provides a platform that allows you to write and execute Python notebooks in the browser, with minimal configuration required and free access to GPUs.

The .ipynb provided by the the teaching staff should be uploaded to the student's Google Drive (using the University's student account) and opened directly in Colab.

Recommended Packages: *Pandas* and *Numpy* (data structures and manipulation), *Scikit-Learn* (Machine Learning), *Matplotlib* and *Seaborn* (visualization).

Submission Guidelines and Deadline

- Completed projects should be submitted via the course's **Moodle** page before the end of the deadline.
- Projects should be executed in groups of 2 or 3 (maximum) students.
- The project should be submitted as a **single .ipynb** notebook file, named following the template "***studentNumber1_studentNumber2_studentNumber3.rar***" (e.g. *31444_31445_31446.rar*) containing:
    - The complete jupyter notebook contemplating the data analysis / machine learning part of the assignment. You can use the template provided in the CLIP platform which simultaneous serves as a guideline, project template and report.
- Deadline is **19 of December, 23:59 GMT**.

Evaluation Criteria

All of the goal/value pairs listed below are based on the assumption that a correct implementation is submitted.

Feel free to fill in the *Completed* column in accordance to your submission for the discussion (replace "-" with "*X*" when suitable).

Goal	Value	Completed
Loading and preparing train/test data	4	-
Training **at least 3** different regressors	6	-
Evaluating each regressor using adequate metrics	4	-
Plotting the results for comparison	3	-
Discussing the results	1	-
Additional features (Free choice)	2	-

Please refer to the lab staff for additional info regarding possible additional features. Examples include for instance performing feature extraction on the original dataset to check if newly created features improve performance or tuning the models' hyper-parameters (please refer to the documentation which includes several examples) to optimize performance.

Lab Planning

- Lab 1 (week of 30/11/20) - Intro, Data Ingestion and Exploration
- Lab 2 (week of 07/12/20) - Model Training and Evaluation
- Lab 3 (week of 14/12/20) - Visualization and Bonus Features

---