Data pretreatment Lab Assignment

In this assignment, you will perform the pretreatment steps described in chapter 2.6 [R1] of the course literature on given sets of process data.

For implementing the lab MATLAB or Python could be used.

Task 1: ----------

Goal: Reduce the size of a large dataset, then rescale it.

Instructions:

1. **Examine the data**
    - Load the file named “large_data.mat” and “imu_data_stationary.mat”
    - Plot the data as a time series
    - Plot a histogram of the data. What can you say about the distribution?
    - What is the mean and standard deviation of the data?
2. **Size reduction**
    - Reduce the size of the data of “large_data.mat” to 64000 data points by undersampling.
3. **Scaling**
    - Rescale the data by standardization.
    - What is the mean and standard deviation of the standardized data?
4. **Evaluation**
    - Show your results and how you produced them.
    - Be prepared to answer the questions listed under each step.

Task 2: ----------

Goal: Perform data cleaning.

Instructions:

1. **Examine the data**
    - Load the file named “corrupted_data.mat”
    - Plot the data as a time series
    - Plot a histogram of the data. What does this tell you?
2. **Missing data approximation**
    - The dataset is missing some values, represented by NaNs.

- Fill in the missing values through your choice of interpolation method.

3. **Outlier removal**
    - Detect and remove outliers using the z-score method.
    - Remember that the removed outliers should be replaced by generated values. You can use the same method as in the missing data approximation.
    - Detect and remove outliers using the z-score method on “imu_data_stationary.mat”.
4. **Evaluation**
    - Show your results and how you produced them.
    - Be prepared to answer questions.

Reflections (optional)

References

[R1]. AI in Chemical Engineering Unlocking the Power within Data. Authors: José A. Romagnoli, and Luis Briceno-Mena.