HIGH-DIMENTIONAL ANALYSIS

Programming Exercise – 01

Deadline: 23h59 22/05/2021

PROBLEM:

Using the room temperature dataset in file room-temperature.csv:

a) Plot a time series plot for each type of temperature data: FrontLeft, FrontRight, BackLeft and BackRight.
b) Implement the PCA algorithm to the whole data set.
c) After observing these time series plot in a), how many latent variables should we use for the PCA algorithm to keep the most meaningful information? Apply the PCA for the temperature data by your chosen number of latent variables.

Dataset description:

- N = 144: number of samples
- K = 4 + 1 columns containing the date and time at which the 4 temperatures were recorded
- Temperature measurements from 4 corners of a room: front left, front right, back left and back right.

After finishing your task, please write a short report or a summary (pdf file) to explain your answers, ideas and the way your code works.

NOTICE

1. Please send the two files (coding and report file) before the due date. Or send the jupyter-notebook file (ipynb, html, pdf) or google-colab link.

2. The mail subject and the folder's name would be *[HDA2020_PEXX_Name_StudentCode]*, where PE means Programming Exercise. For example: *HDA2021_PE01_Le_Van_A_1711001* or *HDA2020_PE01_LeVanA_1711001*.

3. Inside the coding file, there should be a brief introduction (as example below).

```
"""
HIGH DIMENSIONAL ANALYSIS
Programming Exercise: 01
Name: Le Van A
Student code: 1711001
"""
```

4. There is **NO** acceptance for **cheating** or **copying**.

TUTORIAL

Export html file from jupyter-notebook