Project Specifications

Instructions

- Project tasks:
  - Task1: Next Value Prediction (supervised)
    - What is the next value in a sequence of values?

  - Task2: Anomaly Detection (unsupervised)
    - Identify anomalous behaviour in a set of observations

- Create deep learning models to solve the two tasks

- In both tasks we have data collected in the study of Parkinson Disease

Next Value Prediction Task (1)

- We have 3 time series (X, Y, Z) recorded each 10 seconds
- Given a sequence of 5 minutes every one minute
  - **The goal is to predict the next value of the sequence for each time series**

  - E.g. Suppose that x has these values: x = [0, 1, 2, 3, 4]
    - Assume that window_size is 2 and window_shift is 1
    - We create the following sequences: data = [[0, 1], [1, 2], [2, 3]], , label = [[2], [3], [4]]
    - In this case, the goal is to predict for the sequence [0, 1] the value 2, for [1, 2] the value [3] and fot the sequence [2, 3] the value [4]

- Evaluation Metric: **Mean Absolute Error**

Next Value Prediction Task (2)

- Choose a time series (X, Y or Z)
  - **How do the window_size and the window_shift affect the result?**
    - Try with different values and compare your results

Anomaly Detection Task (1)

- **Given data from wearable, the goal is to identify anomalous events, i.e. tremors, in a set of observations**

- Data are collected by patients with and without Parkinson' Desease
- Features include:
  - Identification of patient
  - Accelerometer readings in the three axes (x, y, z)
  - Heart Rate
  - Date and timestamp

Anomaly Detection Task (2)

- The training set is composed by control patient, i.e. volunteers without Parkinson's Desease
  - *Each 1 seconds there is a record*
  - *Missing value on heart rate attribute (labeled with -1)*

- The test set is composed by patient with Parkinson's Desease
  - *Each 10 seconds there is a record*

Output

- Documentation of the solution, max 25 pages

- Code of the solution

- Presentation of the solution
  - A summary of what you did
  - It must last 15 minutes at most

Output

- For the documentation you can follow these steps:
  - Introduction
  - Data Understanding and Preparation
  - Modeling
  - Evaluation
  - Conclusions

General Information

- The project will last 1 year
  - You can deliver it whenever you are ready (however, two weeks before the exam)
  - Official notifications will be provided in official exam periods

- The dataset is available on:
  https://drive.google.com/drive/folders/1vBMbCz2lYNT212JnztVjGAnlUB3KeK2u?usp=sharing