

Project 3 - Presentation

Yann Claes

Institut Montefiore, University of Liège, Belgium



ELEN062-1

Introduction to Machine Learning

November 2021

Goal: Human activity prediction

The goal of this project is to design an algorithm, using supervised learning techniques, in order to predict the **activity** performed by a given subject.

The experiments¹ have been conducted on **8** different subjects: 5 of them will be considered during training while the 3 others will be used for testing purpose.

¹<http://www.pamap.org/index.html>

Dataset - Overview

To achieve this task, you have access to **two** sets of data: *LS*, for training, and *TS*, for testing. Each dataset contains measurements obtained from **31** sensors (`SetName_sensor_Number.txt`).

Each such file contains **3 500** time-series (rows) of length **512** (columns), representing the 5-second periods of measurements. A time-series corresponds to a given subject, whose identifier is given in `subject_Id.txt`, for both sets.

For the training set only, you have access to the **label** of the activity corresponding to the given time-series (`activity_Id.txt`).

Dataset - List of sensors

- ▶ 2: heart beat rate (bpm)
- ▶ 3: hand temperature ($^{\circ}\text{C}$)
- ▶ 4,5,6: hand acceleration (3D, ms^{-2})
- ▶ 7,8,9: hand gyroscope (3D, rad/s)
- ▶ 10,11,12: hand magnetometer (3D, μT)
- ▶ 13: chest temperature ($^{\circ}\text{C}$)
- ▶ 14,15,16: chest acceleration (3D, ms^{-2})
- ▶ 17,18,19: chest gyroscope (3D, rad/s)
- ▶ 20,21,22: chest magnetometer (3D, μT)
- ▶ 23: foot temperature ($^{\circ}\text{C}$)
- ▶ 24,25,26: foot acceleration (3D, ms^{-2})
- ▶ 27,28,29: foot gyroscope (3D, rad/s)
- ▶ 30,31,32: foot magnetometer (3D, μT)

Dataset - List of activities

1. lying
2. sitting
3. standing
4. walking very slow
5. normal walking
6. nordic walking
7. running
8. ascending stairs
9. descending stairs
10. cycling
11. ironing
12. vacuum cleaning
13. rope jumping
14. playing soccer

Output format

Each **submission** to the platform should be formatted as performed in the example submission (see [Kaggle platform](#)).

Organization

- ▶ Guidelines are similar to the previous projects:
 - ▶ **Written report**
 - ▶ Your reproducible codes
- ▶ This project is a **competition** (see rules).
- ▶ Two deadlines:
 1. On **08/12 (11:59PM)**: end of the competition
 2. By **12/12**: submission of the project (codes + report) on the [submission platform](#).
- ▶ Questions? Prefer posting them on the **forum**.

Instructions

1. Create a [Kaggle](#) account, with your real name (please).
2. Create/join a team on Kaggle and on the submission platform.
3. Download the data and the python toy script (you will need to install the Kaggle package).
4. Submit a toy submission (see `toy_script.py`)
5. Think about new ideas and test them
6. Submit your tries
7. Repeat step 5 and 6

Important: write your report in parallel!

Rules

- ▶ Privately sharing code or data outside of teams is not permitted.
- ▶ The maximum size of a team is 3 participants.
- ▶ You may submit a maximum of **5** entries per day.
- ▶ You may select up to **1** final submission for judging.
- ▶ You can not use **external data**.
- ▶ No plagiarism (give your sources).

Rankings and presentations

- ▶ Public and private leaderboards:
 - ▶ **Public** leaderboard composed of **350** samples (time-series).
Directly available for every submission to give you an idea of your model's performance. You should not overfit this ranking.
 - ▶ **Private** leaderboard composed of **3 150** samples (time-series),
computed on your **selected** submission. This final ranking is the only one that matters.
- ▶ 15/12/21 (TBC): presentations.

Take-away messages

We **strongly** advise to start the report before the end of the competition. Finishing as the leader does **not** mean that you will get the best marks!

Have fun!