Project 3: Competition

Predicting a human activity from body sensors.

ELEN0062-1 | Introduction to Machine Learning

Academic Year 2024-2025

Context

Human activity recognition is an important field, with applications like rehabilitation, illness detection, and healthy lifestyle help. This is heavily used in smart watches, for example.

In this project, you will explore this challenge using measurements from different subjects equipped with 31 different sensors, and performing an activity among 14.

Goal: Using machine learning, predict the *activity* being performed from the 31 time series corresponding to each sensor.

Problem Statement

More formally,

- **Input:**
  1. 31 sensor time series of size 512: $\{(s_1^1, ..., s_{512}^1), ..., (s_1^{32}, ..., s_{512}^{32})\}$.
  2. Subject ID (integer).
- **Output:** Activity ID.

Each sensor has some specific scale and unit.

Sensors

- 2: Heart beat rate (bpm)
- 3: Hand temperature (°C)
- 4, 5, 6: Hand acceleration (3D, ms^{-2})
- 7, 8, 9: Hand gyroscope (3D, $\mathrm{rad/s}$)
- 10, 11, 12: Hand magnetometer (3D, $\mu\mathrm{T}$)
- 13: Chest temperature (°C)
- 14, 15, 16: Chest acceleration (3D, ms^{-2})
- 17, 18, 19: Chest gyroscope (3D, $\mathrm{rad/s}$)
- 20, 21, 22: Chest magnetometer (3D, $\mu\mathrm{T}$)
- 23: Foot temperature (°C)
- 24, 25, 26: Foot acceleration (3D, ms^{-2})
- 27, 28, 29: Foot gyroscope (3D, $\mathrm{rad/s}$)
- 30, 31, 32: Foot magnetometer (3D, $\mu\mathrm{T}$)

Activites

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

Data (1/2)

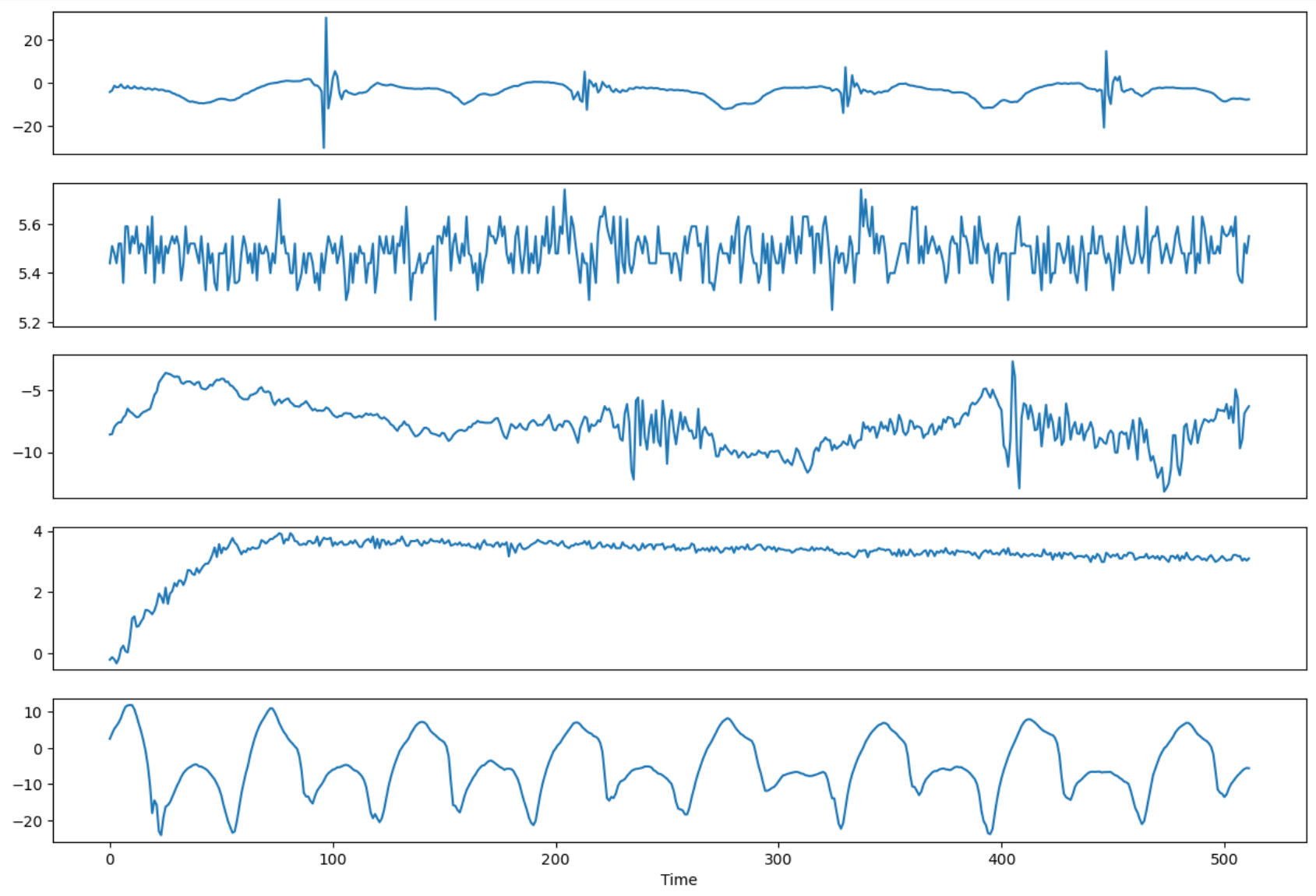
To design a solution, you have access to two datasets, a learning set and a test set. Each one contains 3,500 time series for each sensor. Pay attention, some missing values are present (encoded as -999999.99).

The learning set is measured on 5 subjects, while the test set is measured on 3 other subjects.

Subject ID	s_1	s_2	s_3	...	s_{512}
2	6.85	7.18	7.40	...	7.48
3	2.00	2.00	1.99	...	2.03
2	0.39	1.29	1.80	...	1.91
3	5.51	6.92	8.08	...	5.06
2	5.19	4.22	3.77	...	3.25

Data Visualisation

Sample measurements from 5 sensors



The challenge

Time series should not be treated as typical tabular features, i.e., a vector with 15,872 (= 512 $\times$ 31) features.

The main challenge of the project is to come up with strategies to process this data so that it can be effectively fed to the machine learning methods we studied in class.

There are *many* ways to be very creative with this project, so try many things!

Evaluation Metric

The quality of your predictions will be computed simply based as the *accuracy*, that is the ratio of correct predictions over total predictions.

The Competition

This project is organized as a *competition*.

During the competition: You can submit at most 3 times a day your predictions for the 3,500 test samples. A *public* score is displayed, which is computed on 10% of the test set.

After the competition: Once it is over, your final selected submission will be compared to the other 90% of the test set. The resulting *private* score will be the final score.

Report

We expect more than a good score!

Shortly after the end of the competition, you are asked to submit a report which contains:

- A detailed description of all the approaches you have investigated.
- A detailed description of your approach to select and assess your model.
- A table summarising the results of your different approaches.
- All tables, figures and results should be analysed in-depth while avoiding unnecessary redundancies.
- Any complementary information that you want to analyse.

Guidelines

- 3 submissions per group per day.
- You can use any techniques or algorithms you want.
- You can use any library, but your results must be reproducible.
- You can not use external data, unless you get our approval first.
- Reference every external code, tool, algorithm that you use (e.g., even Scikit-Learn).
- Obviously, do not use ready-made solutions.

Resources

All the resources are available on the website.

https://iml.isach.be

Files:

- `statement.pdf`: More detailed statement.
- `LS`: Learning set folder (see file format above).
- `TS`: Test set folder (see file format above).
- `example_submission.csv`: Example submission CSV file.
- `toy_script.py`: Contains example code to read the train and test sets, process the data, train a simple model, and produce an example submission.

Questions?