Mobile and Wearable Computing

Assignment 04 – Analysis of Physiological Data

Lecturer: Martin Gjoreski
Teaching Assistant: Nouran Abdalazim, Leonardo Alchieri

Lugano, November 27, 2024

Assignment 04

- Deadline
  - Tuesday, December 06, 23:59pm (Lugano time)

- Submission
  - One PDF file containing the following
    - Link to the GitHub repository
    - Screenshots for the results obtained, with a description of them.
  - Uploaded in iCorsi by the deadline

The PDF file

- One (1) PDF file with all your answers

- The PDF must contain
  - Your name
  - Link to the GitHub repository
  - Screenshots for the results obtained, with a description of them.

- The name of the PDF file should be:
  YourName_YourSurname_Assignment04.pdf

- How you create the PDF file is up to you
  - Using a text editor is highly recommended
  - The TA will not correct poorly readable files!
    - E.g.., low-quality pictures are poorly readable!

Exercise 0 – Re-run Tutorial Code

- Re-run the code from Tutorial 07, but with the data provided for this assignment (“assignment_data.csv”).

- **Do not use online tools or upload/share the data with people outside the University.**

- Report the new *balanced accuracy* for the 5-fold and Leave One User Out cross validation. Did the results change?

Exercise 1 – Leave One Day (per user) Out

- Now, implement a Leave One Day (per user) Out cross validation. The day is identified by the date column. Same dates from different users should be considered different “days”.

- In the validation paradigms, you’ll have to leave out a single session from a single user.

- Write the code to run using the models implemented in class, i.e., XGBoost and the DummyClassifier (with uniform strategy).

- Report the *balanced accuracy* for both models with their standard deviation. Did you obtain better or worse results than in the other validation paradigms? Why do you think it could be the case?

Exercise 2 – More models

- Implement the 2 validation paradigms shown in class and the one from Exercise 1 for the following models:
  - Support Vector Machine
  - Random Forest Classifier
  - Naïve Bayes Classifier

- Report the results from this models using balanced accuracy ± standard deviation. How do these models compare to the previous ones?

- You can find their implementation in scikit-learn.