

Mobile and Wearable Computing

Assignment 04 – Analysis of Physiological Data

Lecturer: Silvia Santini

Teaching Assistant: Nouran Abdalazim, Leonardo Alchieri

Lugano, November 24, 2023

Assignment 04

- Deadline
 - Tuesday, December 05, 07:00pm (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 08, 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. Each user had data for multiple “sessions” (identified by a date).
- 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 – Statistical test on results

- In class, we have calculated the *balanced accuracy*, in two validation paradigms (5-fold and Leave One User Out cross validation), for two models.
- However, just looking at the results of the *balanced accuracy* is not enough. Perform a statistical test (t-test) to compare the results obtained from the XGBoost model to those obtained from the DummyClassifier: **are they statistically different?**
- Do this calculation on both 5-fold, Leave One User Out and Leave One Day (per user) Out cross validation.
- Report the p-value and their statistical significant (with $\alpha=0.05$) for all 3 validation paradigms, and comment on them.

Exercise 3 – 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
- You can find their implementation in [scikit-learn](https://scikit-learn.org).
- Finally, re-run Exercise 2 on the results from these new models, checking with ones are statistically different from XGBoost *and* the DummyClassifier. Only compare the new models to these 2: you do not have to perform all possible comparisons.
- Report the *balanced accuracy*, the p-value for the t-tests and their significance for the models mentioned. Comments this results.

Some hints

- Try to implement functions that can be reused in your assignment: avoid boilerplate code.
- We updated the code used for Tutorial 08, especially the Machine Learning prediction. We added shuffling, which aids in the prediction of the model.