

Lab session: Regression and Classification

General information

Before starting the lab session, please note the following information:

- Please submit your report as a single PDF file.
- We recommend working in groups of 3 students (the last group can be a group of 2 or 4)
- You can prepare one single report for the group (name1_name2_name3_lab_PCA.pdf), but every member needs to upload the same file individually.
- There are two experiments in this lab session. A python library has been developed based on Jupyter notebook which is accessible via the Moodle of the course.

Introduction

In this lab session, we focus on the practical usage of the conventional regression and classification methods to design solutions for real-life biomedical problems. The main focuses are as follows:

- Features extraction and selection
- Models definitions and training
- Models testing and validation
- Interpretation of the validation result

Experiment 1: Heart rate estimation using regression techniques

The goal of this exercise is to estimate the heart rate from PPG and acceleration signals using regression methods.

1.1 Code

In the python library, please refer to the following file:

- regression_and_classification_labs\notebooks\ hr_estimation.ipynb

1.2 Data

We use data from the PPG-DaLiA dataset which is accessible via the following link:

- <https://archive.ics.uci.edu/ml/datasets/PPG-DaLiA>

The dataset includes PPG and acceleration signals as well as the reference heart rate computed from an ECG signal. These signals were collected during various activities. Here, we focus only on sitting and walking. The processed data is also available in the python via the following link:

- regression_and_classification_labs\data\ ppg_dalia.pkl

1.3 Problems

- P1. If you want to select a set of features manually, which features would you choose and why?
- P2. Do the automatically selected features match your manually selected features? Explain the reasons for any similarities and/or differences.
- P3. Do you think that the feature selection was useful in this exercise? Why?
- P4. How do you interpret mean, std, and rmse errors of the models?
- P5. Considering all conditions, which model will you finally choose to estimate heart rate? Why?

Experiment 2: Atrial fibrillation classification using conventional classification techniques

The goal of this exercise is to train different conventional classification models to discriminate between atrial fibrillation and normal sinus rhythm from a sequence of inter-beat intervals.

1.4 Code

In the python library, please refer to the following file:

- regression_and_classification_labs\notebooks\ af_classification.ipynb

1.5 Data

We use inter-beat intervals extracted from the publicly available dataset Long Term AF which is accessible via the following link:

- <https://physionet.org/content/ltafdb/1.0.0>

The processed data can be found in the following directory in the python package.

- regression_and_classification_labs\notebooks\ ltafdb_intervals.npz

1.6 Problems

- P6. If you want to select a set of features manually, which features would you choose and why?
- P7. Do the automatically selected features match your manually selected features? Explain the reasons for any similarities and/or differences.
- P8. Do you see any signs of overfitting and/or underfitting of the models? Why?
- P9. Considering all conditions, which model will you finally choose to detect atrial fibrillation? Why?