Master's degree ICT Internet Multimedia Engineering

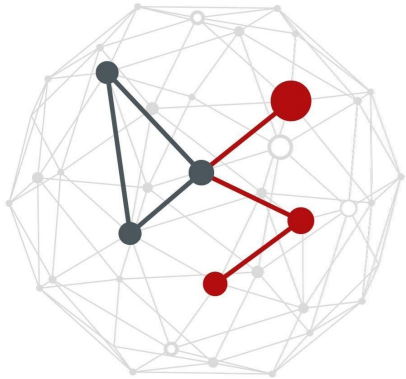
LAB 1 - MACHINE LEARNING TOOLS FOR DIMENSIONALITY REDUCTION: PCA AND CLUSTERING

Eleonora Cicciarella, Cesare Bidini

eleonora.cicciarella@phd.unipd.it

cesare.bidini@phd.unipd.it

University of Padova, IT



DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DIPARTIMENTO MATEMATICA

UNIVERSITÀ DEGLI STUDI DI PADOVA

Lab classes

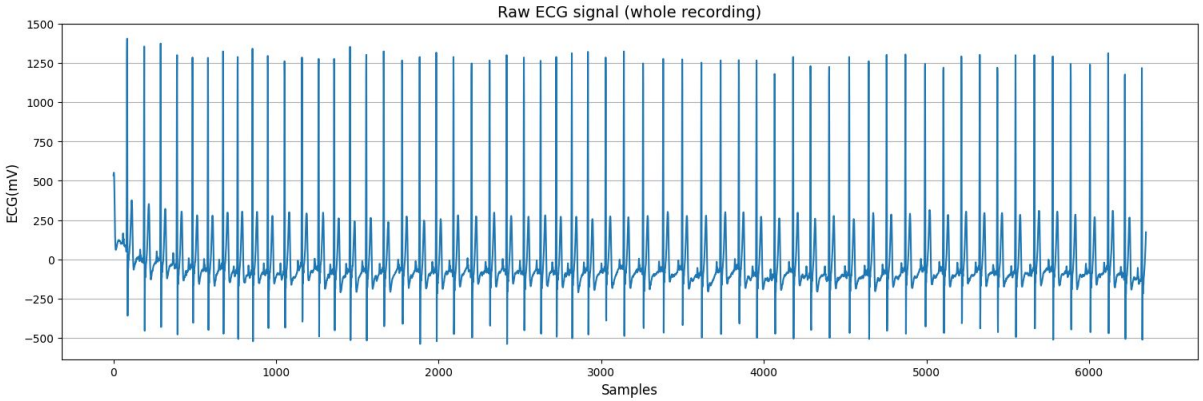
- In this lab we will start with some basic tools

- We suggest you use Google Colab but if you are familiar with your local Python editor feel free to use it

Note: Google Colab does not work with the University's Google accounts. You need to use your own Google account or create a new one.

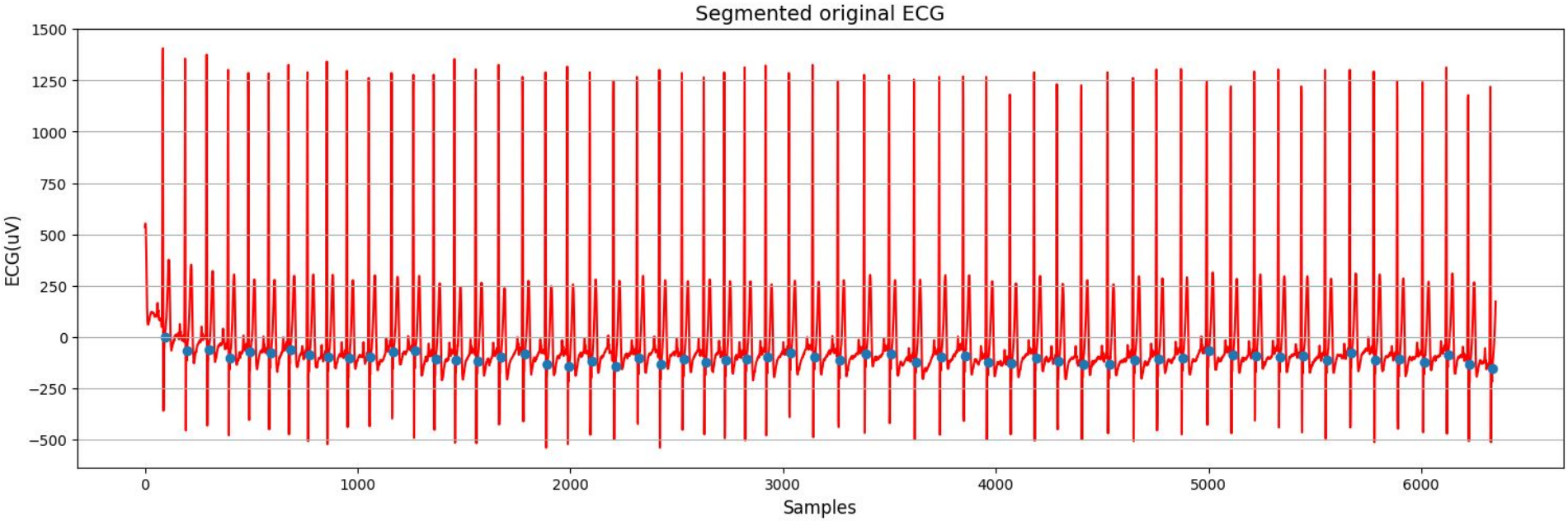
Lab 1 – Oct. 28

- Machine learning tools for dimensionality reduction: PCA and clustering
  - The challenge:
    - ECG signals dimensionality reduction
  - You will learn to:
    - implement a dimensionality reduction algorithm based on PCA from scratch and using the implementation from Python libraries
    - implement dimensionality reduction based on clustering
    - use Scikit-learn Python library



Preprocessing

1. Split the ECG signal into the beats dimensionality reduction will be applied to each beat
2. Normalize the data very important step to be applied at the input of machine learning algorithms!

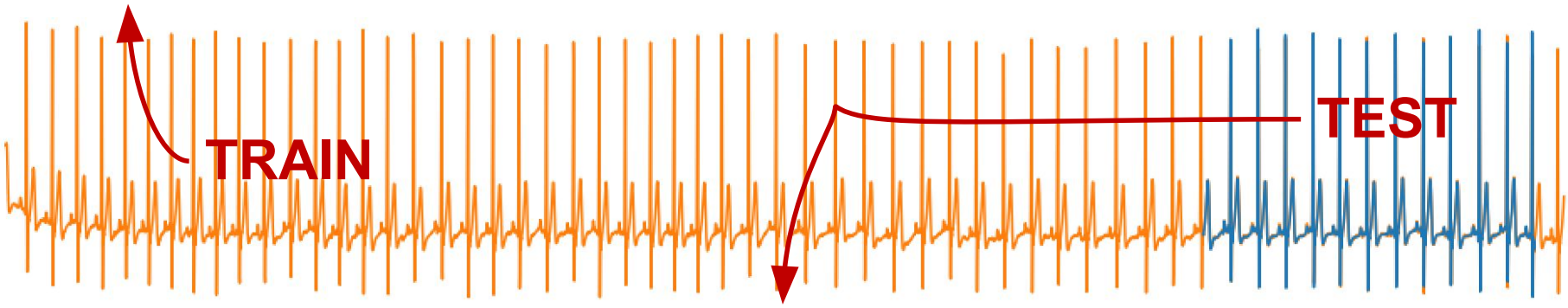


PCA

Obtain the principal components through two approaches:

- Implement PCA from scratch following the procedure detailed in **L01_PCA** (slides from Prof. Rossi) will allow you to digest the procedure seen during the lectures
- Use the PCA algorithm already implemented in Python and offered as part of the Scikit-learn library

https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html#sklearn.decomposition.PCA



Evaluate the effectiveness of the components in representing the ECG beats (test on data not used during training!)

Clustering: DBSCAN

- You will use a different ECG signal composed of
  1. A first part collected from a patient with atrial fibrillation
  2. A second part with normal beats
- Implement DBSCAN using the Scikit-learn library
- Find the optimal **Eps** to separate the two classes (and outliers) using the procedure explained during the classes (see L03_DBSCAN)