



---

## Assignment #2: Daily and Sports Activity Detection

### Problem Statement

In today's era of data-driven decision-making, the ability to identify patterns, group similar entities, and extract insights from large datasets has become indispensable across various domains. Clustering a fundamental technique in unsupervised machine learning plays a key role in organizing data into meaningful groups based on similarities. The objective of this assignment is to help students understand how K-Means and Normalized Cut algorithms can be used for the detection of daily and sports activities captured using motion sensors.

## 1 Download Dataset and Understand the Format

We will use the "Daily and Sports Activity" dataset for this assignment. The dataset comprises motion sensor data of 19 daily and sports activities each performed by 8 subjects in their style for 5 minutes. Five Xsens MTx units are used on the torso, arms, and legs. The data is available at the following link

The dataset consists of 19 activities, each activity is done by 8 subjects for 5 minutes. The 5-minute signals are divided into 5-second segments so that  $480(=60 \times 8)$  signal segments are obtained for each activity. We will treat each of these segments as a data point.

So our dataset now consists of 19 activities  $\times$  8 subjects  $\times$  60 segments resulting in 9120 data points. From each subject, we will take the first 48 segments for training and the rest for evaluation.

Analyze the dataset and preprocess the dataset to be ready for clustering.

## 2 Clustering Using K-Means and Normalized Cut (Your implementation)

We will use K-Means to cluster the daily and sports activity data and cluster them. In each text file, there are 5 units  $\times$  9 sensors = 45 columns and 5 sec  $\times$  25 Hz = 125 rows. Each column contains the 125 samples of data acquired from one of the sensors of one of the units over 5 sec.



We will work on two solutions.

1. Taking the mean of each column in each segment resulting in 45 features for each data point.
2. Flattening all the features together in  $45 \times 125 = 5625$  features for each data point. For this solution, you are required to reduce the number of dimensions and work on the projected data after reducing the dimensions.

Requirements:

- We will change the K of the K-means algorithm between {8, 13, 19, 28, 38} clusters. You will produce different clusters.
- We will compare between the two solutions mentioned above.

### 3 Normalized Cut (Your implementation)

Here, we will use the Normalized cut algorithm to cluster the daily and sports activities. For this experiment.

- Apply Normalized Cut algorithm to the two solutions to cluster the data into **19** clusters.
- Compare the results of K-Means and Normalized Cut clustering regarding the number of detected activities and their characteristics.

### 4 Evaluation

We will evaluate models based on their ability to detect daily and sports activities accurately. You will be required to use the following metrics to evaluate the quality of their models:

- Precision
- Recall
- F1 score
- Conditional Entropy



---

## 5 New Clustering Algorithm (Your implementation)

Your goal is to get to know to other clustering techniques and how they are working and differ than K-Means and Normalized cut. Choose any clustering technique of your own choice, implement it and repeat the above experiments.

## 6 Submission Notes

- Work in groups of 3 students.
- You are required to submit a clear and detailed report [in PDF format] illustrating every step in the assignment