## Assignment 1 Introduction

You will have 2 project assignments in this course, the first of which is introduced below and due at the end of Week 4. The second will be introduced in Week 5 and will be due at the end of Week 7.

Please read the information provided here to make sure you understand what is required and so that you can begin working toward the deadline. If you have questions, please post them in the discussion board.

**Note: This assignment can be done in either Python or MATLAB.**

## Assignment Instructions

In this project, we will attempt to develop a computing system that can **understand** human activities. You will be provided data for a given activity, specifically eating action mixed with other unknown activities. Your aim is to identify the eating activities amongst the noise. We will be working with real world wristband data and, in the process, we will use all the concepts that we will learn in the data mining course. **Note that this is only an attempt and we do not expect a full solution yet.**

**Submission Instructions:**

- The entire code to answer the questions for all three phases should be in **one file.**
- Create a PDF explaining your answers to the questions in all three phases.
- Create the graphs to answer the appropriate questions in Phases 2 and 3.
- Create a **single zip file** containing the code, the graphs, and PDF explaining your answers. **Due to the assignment configuration in Coursera, you will need to upload the *same zip file* to each submission space provided in week 4.**

**Assignment 1_Activity Recognition Project.pdf**
PDF File

**Assignment 1_Activity Recognition Project Rubric.pdf**
PDF File

**Assignment 1_Activity Recognition Sample Project.pdf**
PDF File

## Phase One: Data Cleaning and Organization [20 points]

Data_Mining_Assign1Data.zip

The data provided to you is collected using two sources:

A) **Wristband data**: Where the subject is wearing the wristband and performing eating actions periodically interspersed with non-eating unknown actions. The wristband provides you with i) accelerometer, ii) gyroscope, iii) orientation, and iv) EMG sensors. The sampling rate is 50 Hz for all the IMU sensors and 200 Hz for EMG sensors .

B) **Video recordings** of the person performing eating actions only used for establishing ground truth. In the data, you will be provided with the ground truth in the form of frame numbers where an eating action starts and ends. The actual video data will not be provided for privacy issues. The video data is taken at 30 frames per second. Hence you have to convert the frame numbers into sample numbers for the wristband sensors through some logic that you have to develop. The assumption that you can take is that the start frame and sample #1 of the wristband are synchronized. The output of this step will be a set of data snippets that are labelled eating actions and a set of data snippets that are non-eating.

The way to convert the frame numbers into sample numbers is as follows:

Consider the ground truth file, where there are three columns. Ignore the last column. The first column is the start frame of an eating action, the second column is the end frame. Each row is an eating action. The first frame number can be multiplied by 50 and divided by 30. This gives you the corresponding sample number that indicates the start of an eating action. The second frame number can also be multiplied by 50 divided by 30 to get the end sample of an eating action. Do this for every row to get the sample numbers for eating actions for a person.

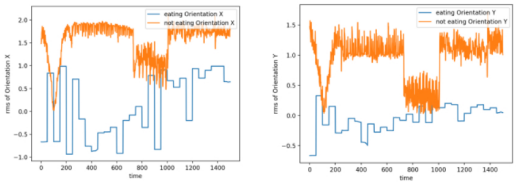
## Phase Two: Feature Extraction [50 Points]

In this task, you should select and implement five existing feature extraction methods such as Fast Fourier Transform, Discrete Wavelet Transform, a set of statistical features (min, max, avg, std, RMS, energy function), etc. The five types of feature extraction methods can be chosen by you. The aim is to use features that show clear distinction between the actions.

A potential way of figuring out features is to plot the raw data and try to visually understand differences.

For each type of feature extracted do the following things: a) Write an explanation on how the feature is extracted. b) Write an intuition on why you use such a feature. c) Write a MATLAB code to extract that feature from each time series created in task 1. d) Generate plots each corresponding to eating and non-eating activities. This will give you a better idea of potential patterns in the features. e) Discuss whether your initial intuition about the features that you selected holds true or not.

An example plot is shown below which shows clear difference between an eating and non-eating action.




## Phase Three: Feature Selection [50 Points]

This step involves reduction of the feature space and keeping only those features which show maximum distance between the two classes (eating and non-eating). We will use Principal Component Analysis technique discussed in class for this purpose. The PCA code is already available in MATLAB, hence there is no need to PANIC! Just use it.

- **Subtask 1: Arranging the feature matrix –** You know PCA only takes one matrix. How will you arrange all sensors and their corresponding features into a single matrix such that the eigenvectors of the covariance matrix directly makes sense to your data set? This means that if the PCA results gives you a eigen vector then the new feature matrix can be obtained by simply multiplying the eigen vector with the old feature matrix. Write your logic of feature matrix arrangement.
- **Subtask 2: Execution of PCA –**Use MATLAB's PCA function to run PCA on your feature matrix. Show all the eigen vectors in a plot.
- **Subtask 3: Make sense of the PCA eigen vectors –** Write an explanation on the reason why the eigen vectors turned out the way they did.
- **Subtask 4: Results of PCA –** Create the new feature matrix.
- **Subtask 5: Argue whether doing PCA was helpful or not –** May be compare the plots generated from subtask d of task 2 and subtask 4 of Task 3.

Mark as completed