Assignment 1- Human Activity Recognition

CSE 572: Introduction to Data Mining (Fall 2018)

To:
Prof. Ayan Banerjee
I.R.A. Fulton School of Engineering
Arizona State University

By:
Group - 4
Jayashree Subramanian (1214347796)
Mahathi Srinivasan (1215187155)
Mahima Chhalani (1215161558)
Arpan Vitthalbhai Patel (1215104930)
Vivek Ranjan Panigrahy (1215218901)

Abstract:

This project involved 3 phases i.e. Data Collection. Feature extraction and Feature Selection, in which sensor data from Myo sensors that records 5 different kinds of data for each activity performed, namely: Accelerometer, EMG, Gyrometer, Orientation and Orientation-Euler are collected. The data from these sensors are extracted and analyzed by selecting features which gave us maximum variation and plotted scatter plots to support our hypothesis. Secondly, we built the feature matrix which had a set of all features that were analysed in phase 2 and performed Principal Component Analysis (PCA) on the matrix in order to reduce the dimensionality of the data and to obtain the informative features.

Keywords:

Feature, Myo Armband, Accelerometer, EMG Sensor, Gyrometer, feature matrix, PCA

1. Introduction

1.1 Terminologies:

a. ***Feature:*** A feature is any term in the document/dataset which is specific to it and helps in distinguishing the document/dataset from others.

b. *Gesture:* A gesture is a form of nonverbal communication or non-vocal communication in which visible bodily actions communicate particular messages, either in place of, or in conjunction with, speech. Gestures include movement of the hands, face, or other parts of the body.

c. *Gesture Recognition*: Gesture recognition is the mathematical interpretation of a human motion by a computing device. It is the ability of a computer to understand gestures, interpret them and analyze it.

d. *Human Activity Recognition (HAR)*: Human activity recognition is a key sub area of Human Computing Interaction (HCI) where a subject (usually a human) wears an array of sensors and these sensors record various input signals from the body which can be interpreted as gestures. Once the system has enough input signals with labels (Gestures) to differentiate one signal from the other, we can use the system/model to predict the next input signal without a label.

e. *Myo Armband*: It is a wearable gesture control and motion control device that records 8 different kinds of data like Accelerometer, Gyro meter, Electromyography, Orientation, Orientation-Euler etc.

Fig 1. Myo Armband

f. *Accelerometer:* A sensor which measures proper acceleration of the object (in our case, hand) with respect to the x, y and z axis. Proper acceleration means the acceleration a body achieved by its own with respect to the instantaneous rest frame.

g. *Gyrometer*: This is used to measure the angular velocity i.e. the change in rotational angle per unit of time. These sensors collect input from the coriolis force applied to a vibrating element. For this reason, the accuracy of the angular velocity measured differs significantly depending on the element.

h. *EMG Sensor*: Electromyography sensor records the electrical activity produced by skeletal muscles in the body. Myoelectrical signals are formed by physiological variations in the muscle fibers. There are total of 8 EMG signals namely EMG1, EMG2, EMG3, EMG4, EMG5, EMG6, EMG7 and EMG8.

i. *Clustering:* It is a process where we club together similar data points having a common attribute or common goal which differentiates them from all other sets of samples/data points.

j. *PCA:* Principal Component Analysis is a statistical procedure that uses orthogonal transformation to convert a set of observations of linearly uncorrelated variables into principal components along maximum variance.

k. *Feature Matrix*: It is a (Sample x Matrix) matrix which has a list of all features with respect to their sensors. This feature matrix is the input to the PCA algorithm which then outputs a set of Eigenvectors and Eigenvalues corresponding to the best features across which we have maximum variance.

l. *Roll, Pitch and Yaw*: These are Euler angles used to measure orientation of an object. For example, if we consider a 3 straight orthogonal lines (intersecting each other at 90 degrees) along a central plane, then rotation along the front-to-back axis is called roll, rotation along the side-to-side axis is called pitch and rotation along the vertical axis is called yaw.

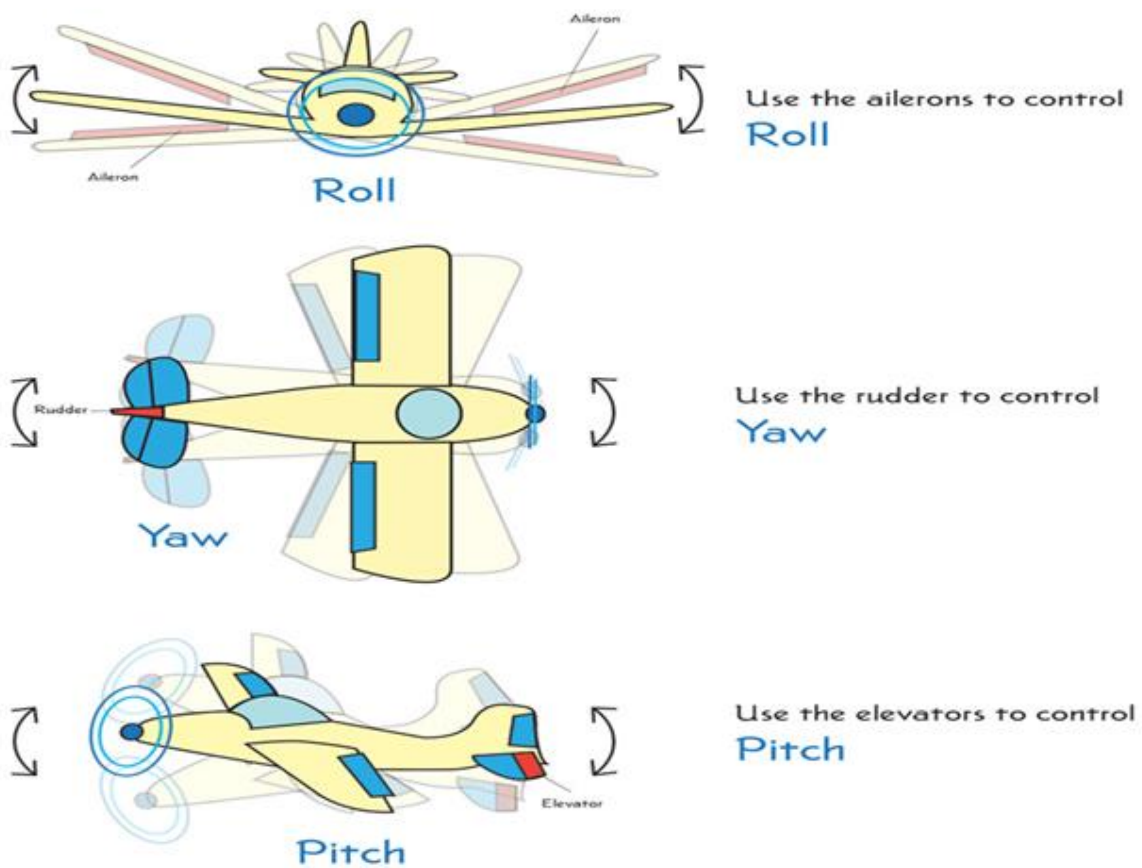


Fig 2. Representation of Pitch, Yaw and Roll

1.2 Goal description:

Phase 1: **Data Collection**

During this phase, 2 students wore the Myo band on their dominant hand for a period of 24 hours each. They were free to perform any activity they wished to. But they were requested to note down the start and end of each activity being performed like eating, cycling, swimming, cooking, driving, playing guitar, etc. Our aim for this phase was to collect data for at least two activities i.e. Eating and Non-Eating Activity. The data related to the rest of the activities measured could be used to perform an in-depth analysis to increase the accuracy.

Phase 2: **Feature Extraction**

In this phase we needed to extract meaningful features using feature extraction techniques like, mean, standard deviation, root mean square, min, max, fast fourier transform (FFT) etc. Our aim was to use the best features that showed a clear distinction between the two types of actions. We had 5 subtasks:

a) Writing an explanation on how the feature was extracted
b) Intuition for selecting the feature
c) Writing a MATLAB code to compute these features from each time series
d) Generating plots corresponding to each activity
e) Discussing whether initial hypothesis/intuition behind these features was true or not

Phase 3: **Feature Selection**
This phase involved dimensionality reduction of the feature space by creating the feature matrix and keeping only those features which retained the maximum variance between the two activities. We have four subtasks:

a) Arranging the feature matrix
b) Execution of PCA
c) Making sense of the obtained Eigenvectors and mapping them back to our initial features
d) Results of PCA: Conclusion on which features selected in the hypothesis stage (Phase 2) still holds true after doing PCA

2. Implementation:

The initial timestamp of when the activity was started, corresponds to the first timestamp in the sensor files. It is constant for all the sensors for one activity.

a) **Sampling frequency**: The sampling frequency of the data is given as 200 Hertz. Hence, we divided the timestamps into samples of 1 second each. On division of timestamps of 1 second, we got 50 timestamps for each second that corresponds to one sample. We then labeled these samples corresponding to their timestamps. We computed the sampling frequency (SF) of each sensor and found that the SF of the accelerometer is 73.935 and that of the gyro meter is 81.92 (averaged over 1000 milliseconds).

Hence the sampling rate is:
Sampling rate = (No of timestamps in one sample)/sampling frequency

$$S_r = T_n / S_v$$

S_r is the sampling rate
T_n is the number of timestamps in one sample
S_v is the sampling frequency of the sample data

b) **Segmentation**: Once we obtained the real world time for the activities in seconds, we segregating them into samples (each consisting of 50 timestamps) and extracted features like Mean, Min, Max, Standard Deviation, Root Mean Square and Fast Fourier Transform for each of the activities in their respective directions (x,y,z).

c) **Outliers**: They correspond to data points whose characteristics are very different from the other data points in the particular activity. These occurrences of outliers are aberrations which can be used due to multiple factors such as, miscalibration of the sensor, over stretching of the arms or legs.

d) **Feature Extraction and Feature Selection**: Five distinct features are selected as a result of feature extraction process for the sensors across all the dimensions (x, y, z and w). The number of features before performing feature reduction (PCA) is compared with the features obtained after PCA. The percentage of variance retained across the dimensions are measured, analysed and reported.

3. TASK 2: FEATURE EXTRACTION

a) Arranging the feature matrix: In feature extraction, we first analysed the data collected from the sensor for the eating and non-eating activities and then chose the following for pattern analysis**:**

1) Mean of the Gyro velocity x,y,z
2) Mean of the Accelerometer x,y,z
3) Mean of the Orientation x,y,z,w
4) RMS of the Gyro velocity x,y,z
5) RMS of the Accelerometer x,y,z
6) RMS of the Orientation x,y,z,w
7) Minimum of the Gyro velocity x,y,z
8) Minimum of the Accelerometer x,y,z
9) Minimum of the Orientation x,y,z,w
10) Maximum of the Gyro velocity x,y,z
11) Maximum of the Accelerometer x,y,z
12) Maximum of the Orientation x,y,z,w
13) First Fourier Transform of the Orientation x,y,z,w

In 1,2,3 we calculated the Mean of the various samples taken in the x, y and z axes. The mean is given by the following formula:

- Mean(J)= $(\sum_{i=1}^{n} J)\ /$n

In 4,5,6 we calculated the Root Mean Square of the various samples taken in the x, y and z axes. The Root Mean Square is given by the following formula:

- $x_{rms}=\sqrt{\frac{((x1)^2+(x2)^2+\cdots\ldots\ldots+(xn)^2)}{n}}$

In 7,8,9 we calculated the Minimum value of the various samples taken in the x, y and z axes. The minimum is given by the following formula:

- Minimum(X)= min (x_1, x_2, x_3…….., xn)

In 10,11,12 we calculated the Maximum value of the various samples taken in the x, y and z axes. The maximum is given by the following formula:

- Maximum(X)= max (x_1, x_2, x_3……., xn)

In 13 we used a combination of mean and FFT since MATLAB's fft () gives both real and complex values. In order to obtain the real values and plot them, the following formula is used:

- value=mean(abs((fft(x_1,x_2,...,x_n))^2))

b) Our Intuition, given five features:

1. Mean

- We first clubbed the several time stamps in the file corresponding to a single sample (for a particular second) and then we took the mean of the total number of the entries corresponding to the particular second after clubbing the entries.
- When taking mean of a series of data points we normalise the sample space into one value, hence getting rid of unnecessary noise and extraneous values. Hence mean was selected as a feature for differentiating between data points of the 2 activities, Eating and Non-Eating Activities, i.e. Eat Food v/s No Movement.

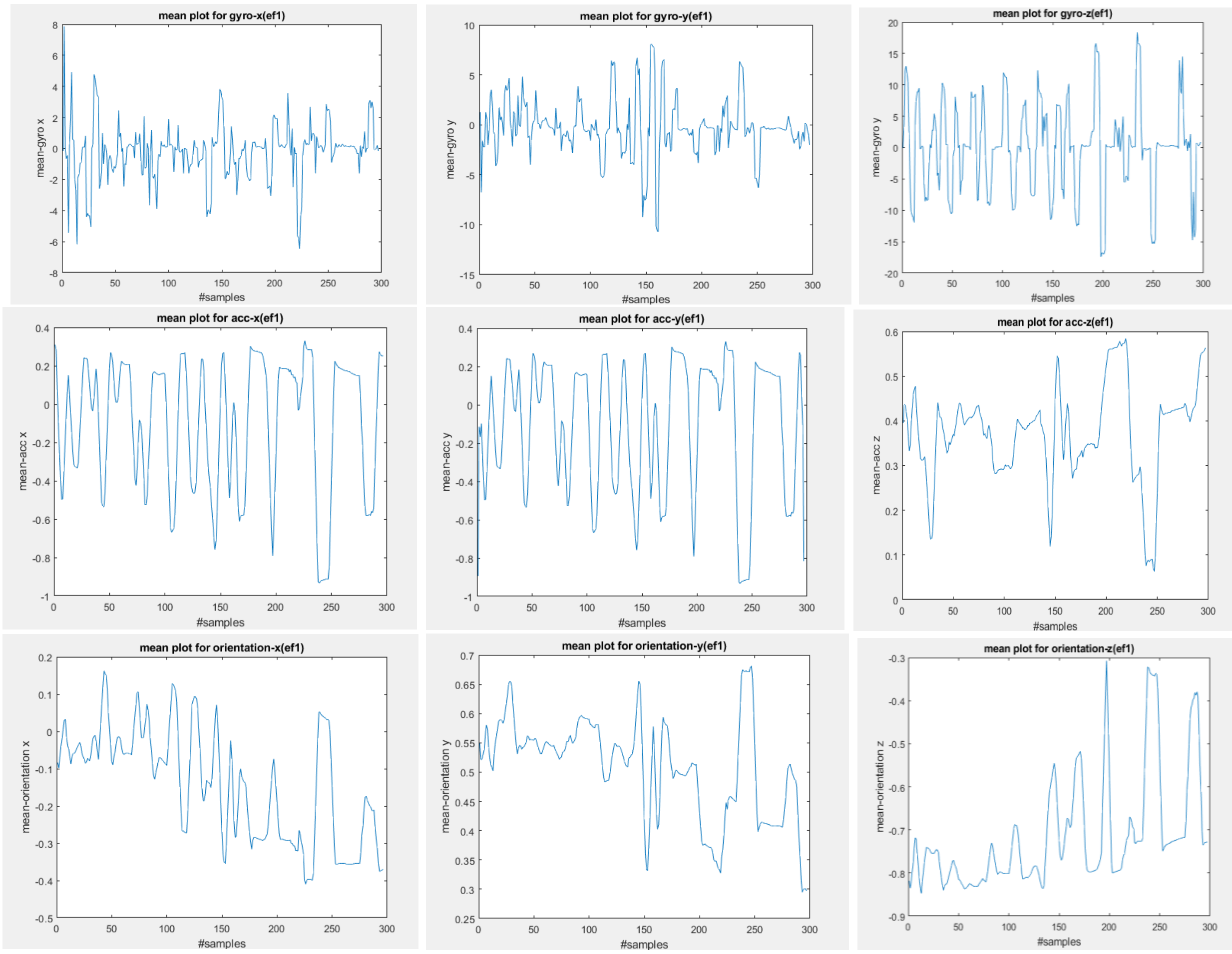


Fig 3. Mean plots for samples (accelerometer,gyrometer,orientation) in all directions EatFood(ef)

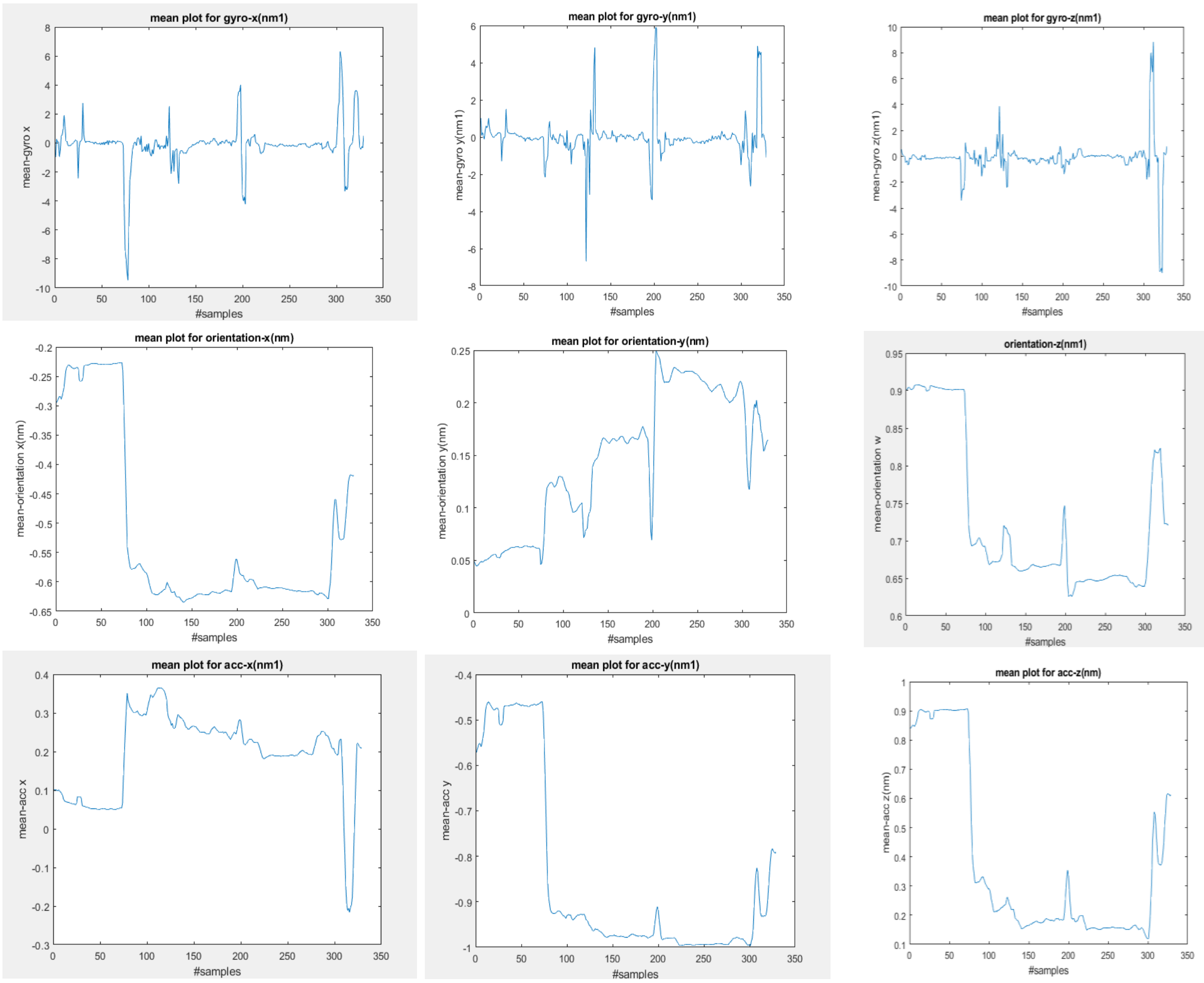


Fig 4. Mean plots for samples (accelerometer,gyrometer,orientation) in all directions for No Movement(nm)

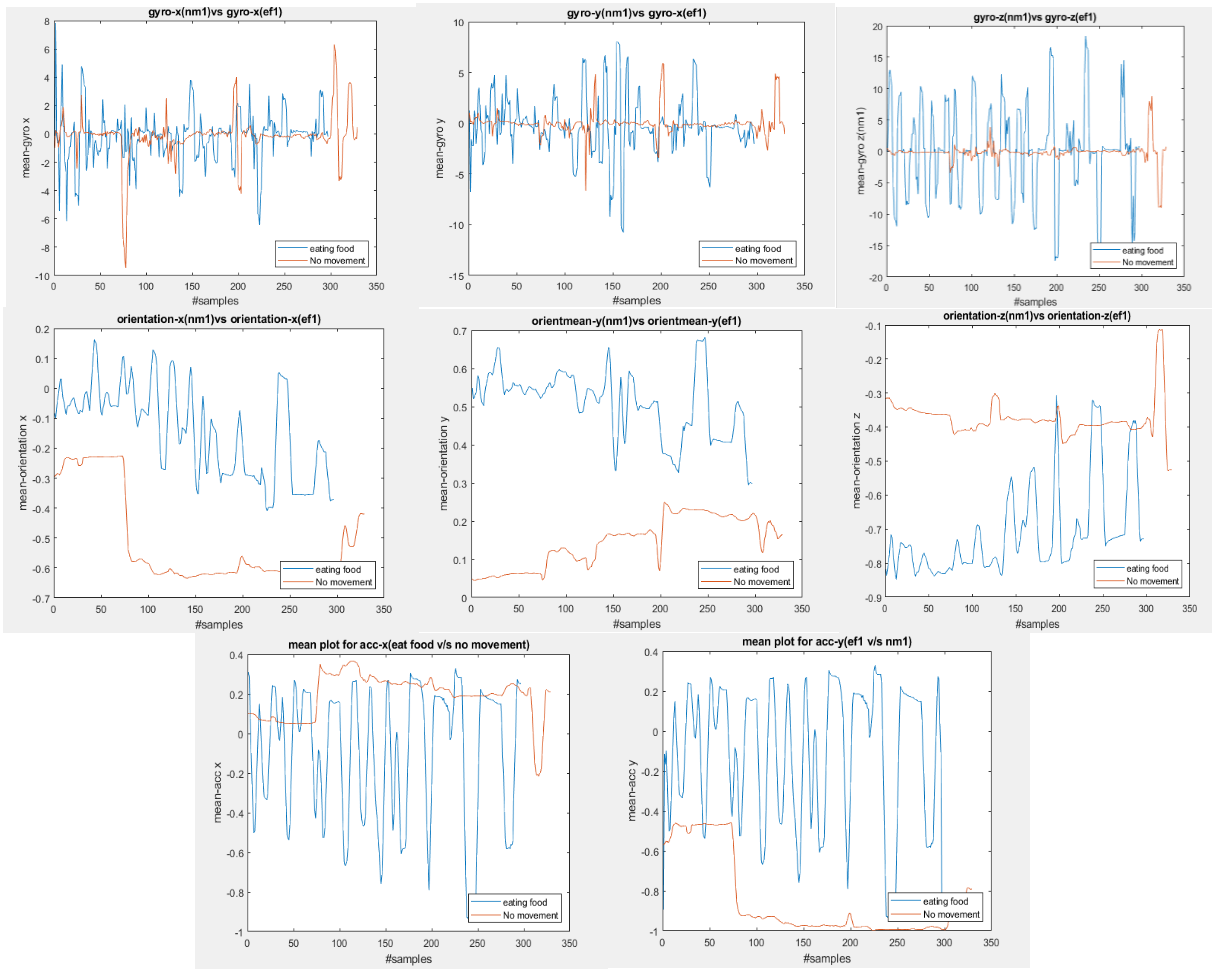


Fig 5. Comparison of mean plots between the two activities Eat Food and No Movement

2. Root Mean Square

- In calculating the Root Mean Square, we used of the samples of the mean that we had calculated earlier.
- Root Mean Square over a sample gives us the effective value in the series which best corresponds to the entire sample space, hence it was considered to be a feature which could be used in segregating the two activities.

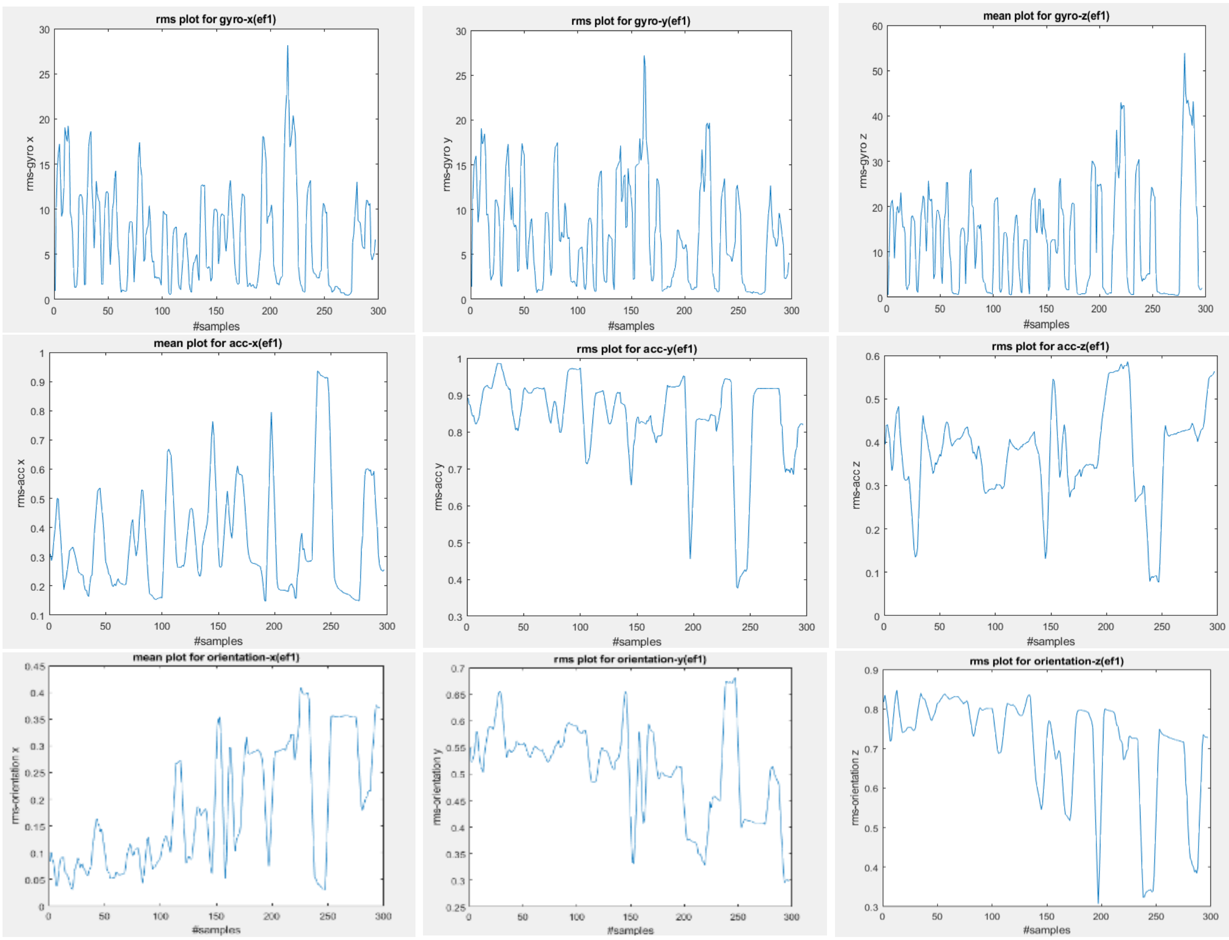


Fig 6. RMS plots for samples (accelerometer,gyrometer,orientation) in all directions for EatFood(ef)

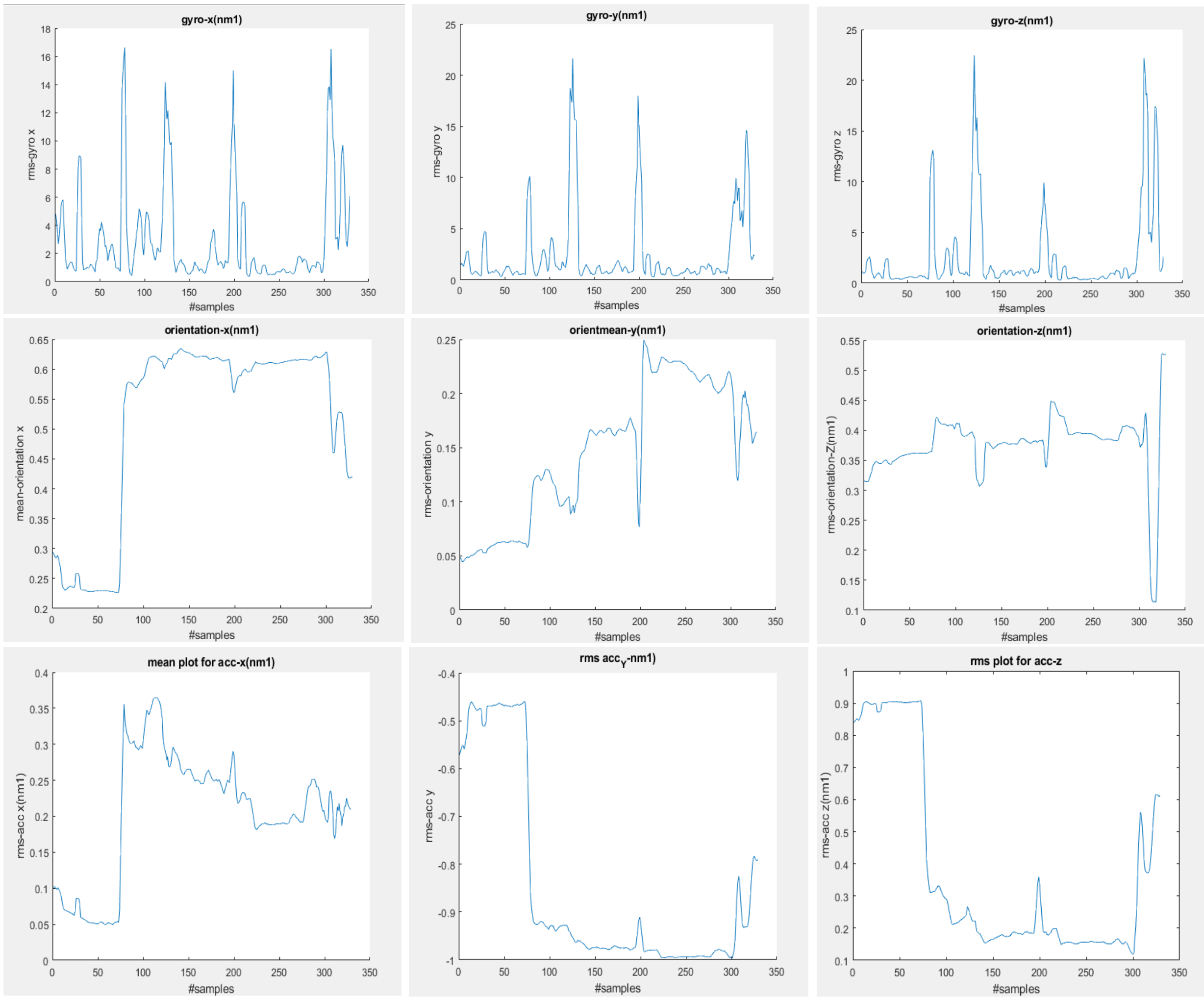


Fig 7. RMS plots for samples (accelerometer,gyrometer,orientation) in all directions for No Movement(nm)

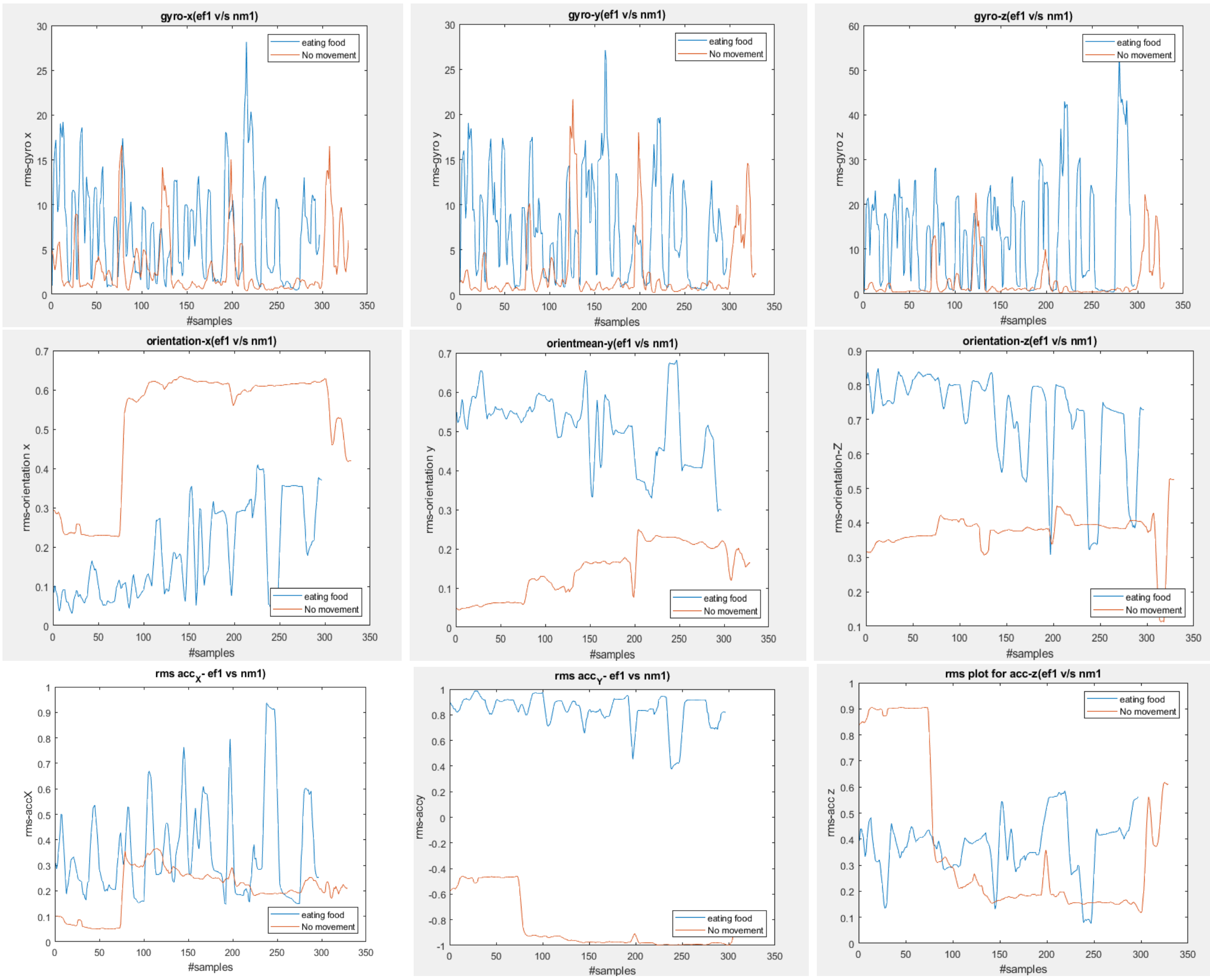


Fig 8: Comparison of RMS plots between the two activities Eat Food and No Movement

3. Minimum

- For finding the minimum value, here also we have clubbed the time stamps into the samples and have found the minimum value for all the corresponding readings in X, Y and Z.
- The intuition for finding the minimum value is that by calculating it, we obtain the data points across the samples having the least variance.

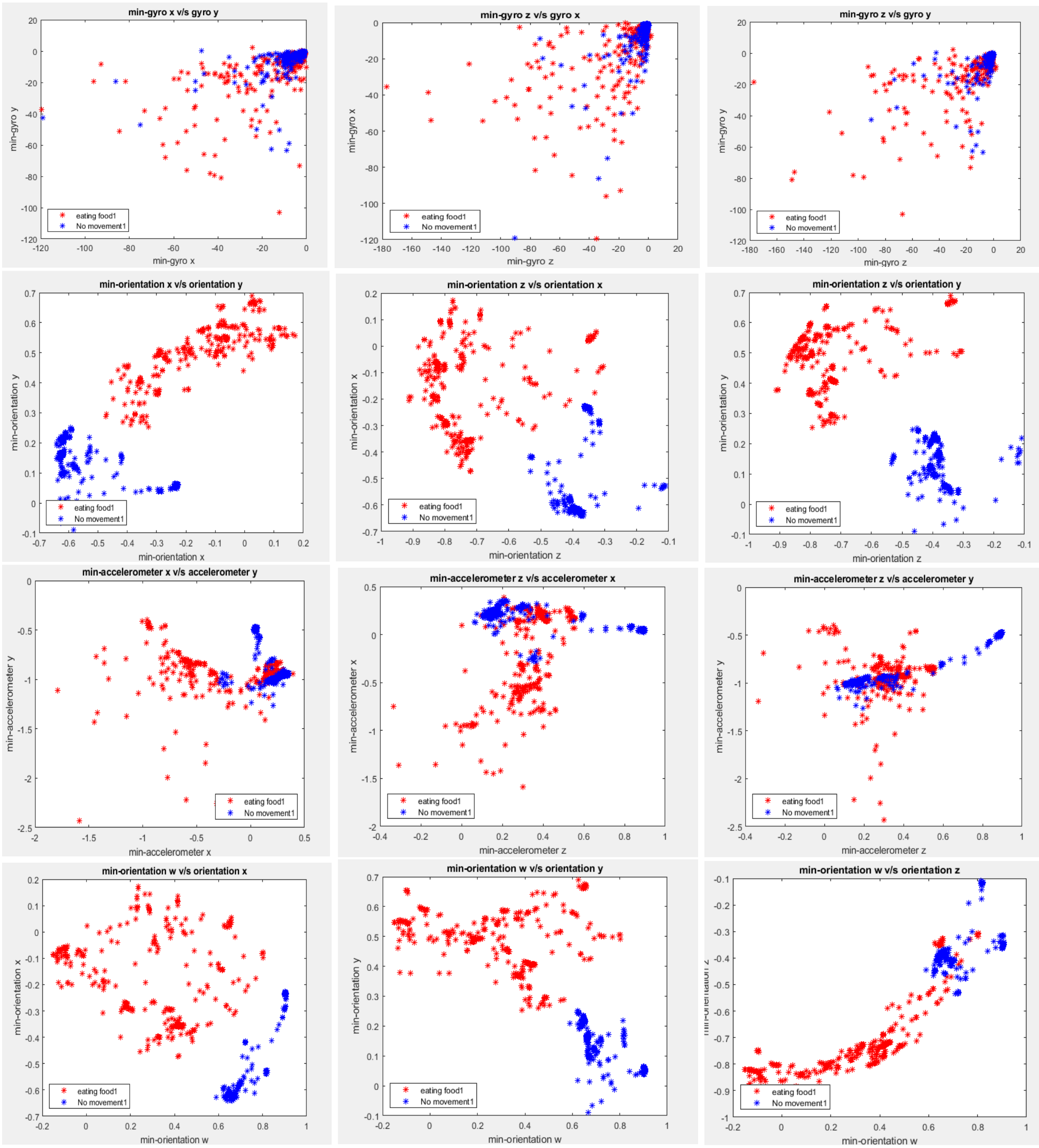


Fig 9. Comparison of min scatter plots between the two activities Eat Food and No Movement

4. Maximum

- Here, for finding the maximum value we have done the sampling in the same way as we have done while finding the minimum value and have calculated the maximum value for the samples.
- The intuition behind finding the maximum value is that by calculating it, we obtain the data points across the samples having the highest variance.

- Thus obtaining minimum (previous feature) and maximum now we tend to have a better comparison of the features for each sample considered.

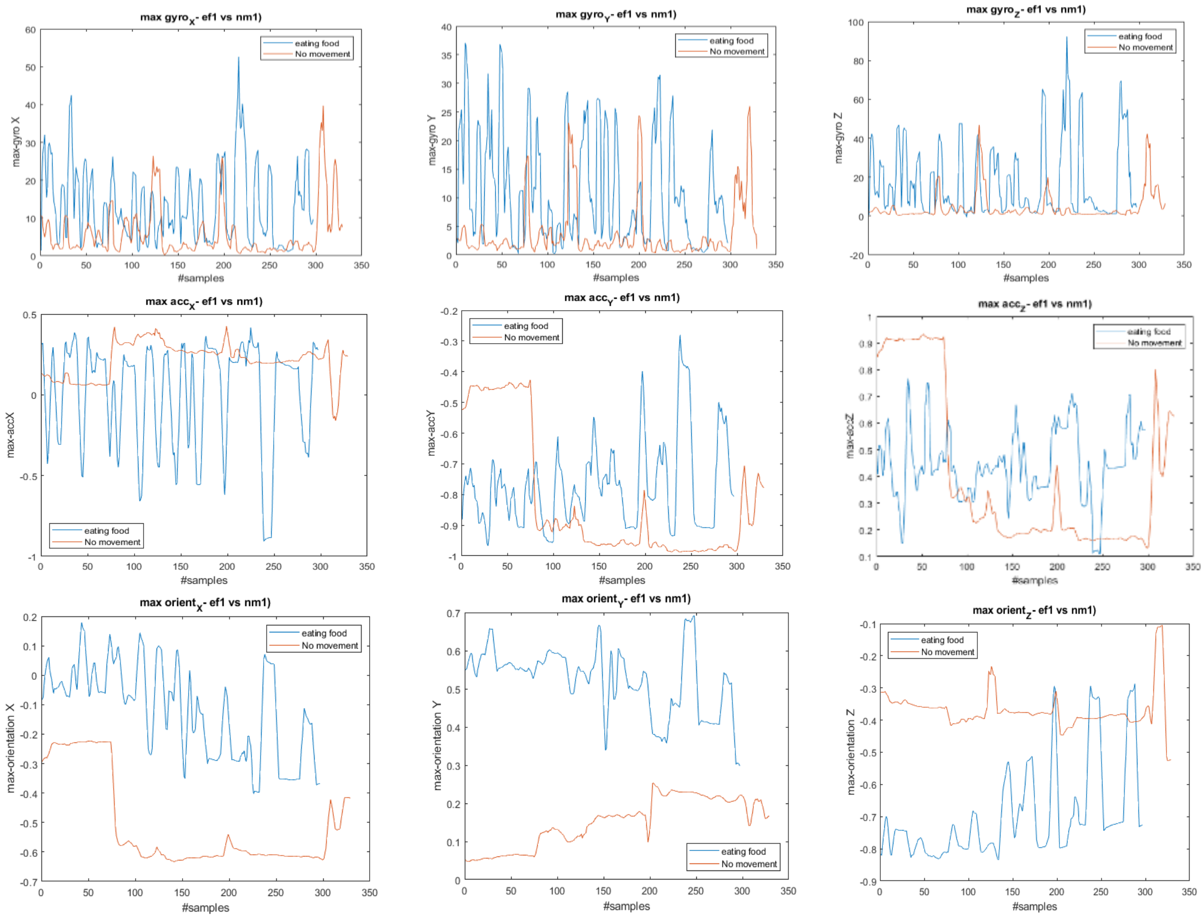


Fig 10. Comparison of max plots between the two activities Eat Food and No Movement

5. Fast Fourier Transform:

- It is a technique which transforms the data points in our sample space over a period of time and divides it into frequency components.
- Choosing FFT as a feature helped us get a better spread of the data across the time domain with its frequency components intact hence ensuring that the data preserves all the properties even after transformation.

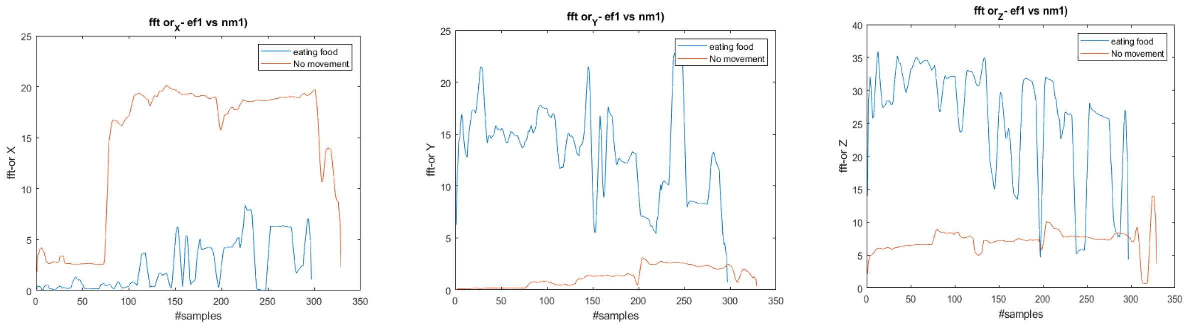


Fig 11. Comparison of FFT plots between the two activities Eat Food and No Movement

c) Initial hypothesis/intiuition held for choosing the features:

After performing feature selction on all the data points, we observed that the plots corresponding to orientation for all five features were able to distinguish between the eating and non eating activity more effectively. In particular, FFT across all dimensions, mean and RMS were found to be the most interesting features amongst other features. Maximum and Minimum features for gyrometer and accelerometer did not prove to be a good measure in distinguishing the two activites.

4. TASK 3: FEATURE SELECTION

The aim of this task was to change the original feature space into a reduced feature space by keeping only those features which showed maximum distance between the two classes. To achieve this, we used the Principle Component Analysis (PCA) technique.

Principal component analysis (PCA) is a dimension-reduction algorithm that simplifies the complexity of high-dimensional data by performing an orthogonal linear transformation of data into fewer dimensions. It reduces data by geometrically projecting them onto lower dimensions called principal components, with the goal of preserving the variance of the data using a limited number of principal components. The first principal component is chosen to maximize the variance of the projected points. The second and subsequent principal components are selected similarly, with the additional requirement that they be uncorrelated with all previous principal components.

The feature matrix is arranged by building a n-by-d matrix, where each row corresponds to a sensor observation (each recorded sample) and each column corresponds to their corresponding features. To obtain the new feature matrix, we need to find the eigenvector matrix of the covariance matrix and multiply the eigenvector with the old feature matrix in that order. We can find the new feature matrix directly by using MATLAB's PCA function by passing the original feature matrix.

Subtask 1: **Arranging the feature matrix**
We selected the sensors and their corresponding features which showed clear distinction between the two activities and arranged the feature matrix to perform PCA. We took Fast Fourier Transform, Mean, Root Mean Square, Minimum, Maximum of the Z, Y, W and X axes of the Orientation Dataset and the Root Mean Square, Mean, Minimum, Maximum of the X and Y axes respectively of the accelerometer dataset. We then normalized the data by finding the covariance matrix before executing PCA.

Subtask 2: **Execution of PCA**
We wrote the code in MATLAB, which is named, *featureselection_for_PCA.m*
The pca function of MATLAB defined as [coeff,score,latent,tsquared,explained,mu] = pca(X), returns:

- The coefficient matrix (28-by-28 matrix) as **'coeff'**. For our n-by-d feature matrix, i.e. 626 x 28 feature matrix X, the corresponding coefficient 28-by-28 matrix is calculated and stored in the variable 'coeff'. Each column of coeff contains loading/weights for one principal component, and the columns are in descending order of component variance, latent.
- The principal component scores as **'score'**. Principal component scores are the representations of X in the principal component space, where rows of score correspond to observations, and columns corresponds to components.
- The principal component variances as **'latent'**, which stores the eigenvalues of the covariance matrix of X.
- The Hotelling's T-squared statistic for each observation X as **'tsquared'**, which is the sum of squares of the standardized scores for each observation.
- The information regarding the percentage of the total variance retained by each principal component as **'explained'**.
- The estimated mean of each variable in X as **'mu'**.

After applying PCA, we obtained 79.07% variance for Principal Component 1, 15.31% variance for Principal Component 2, 5.11% variance for Principal Component 3, 0.44% variance for Principal Component 4 and 0.026% for Principal Component 5. Thus, we selected the first five principal components and retained 99.96% variance.

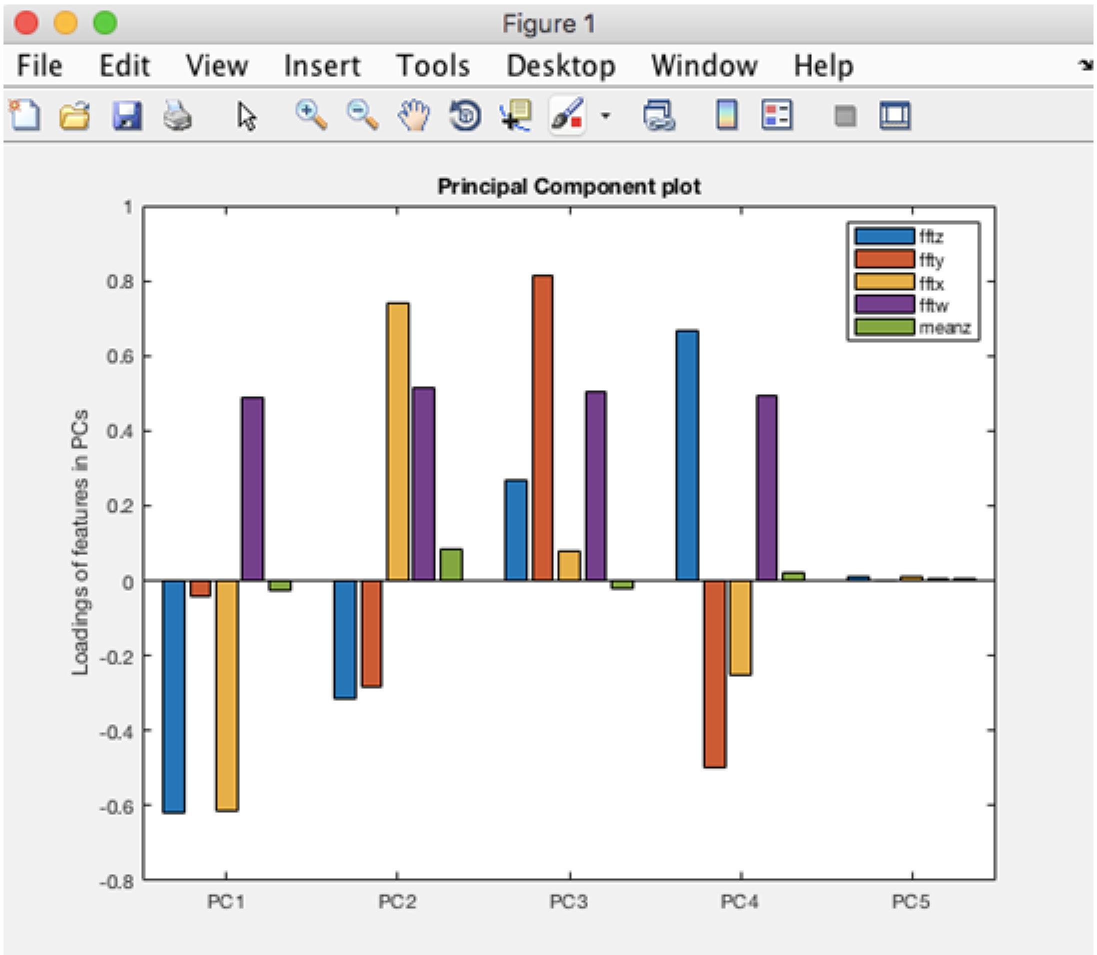
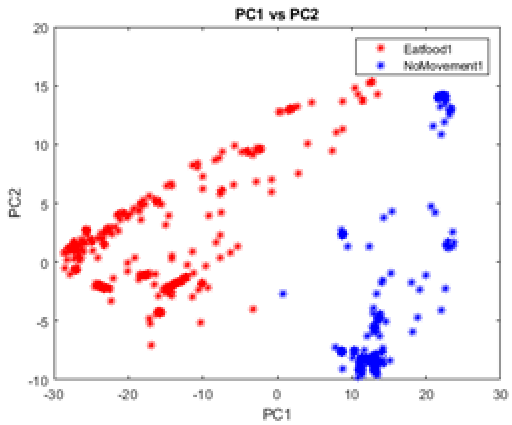
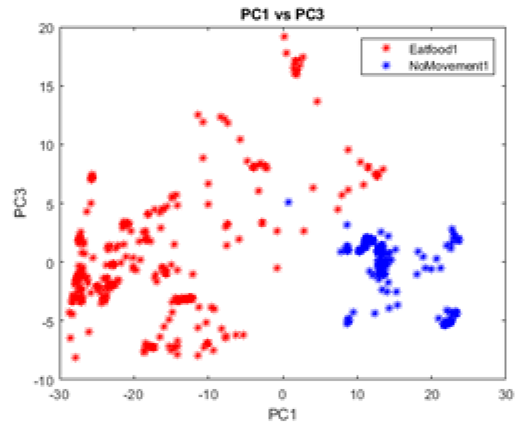


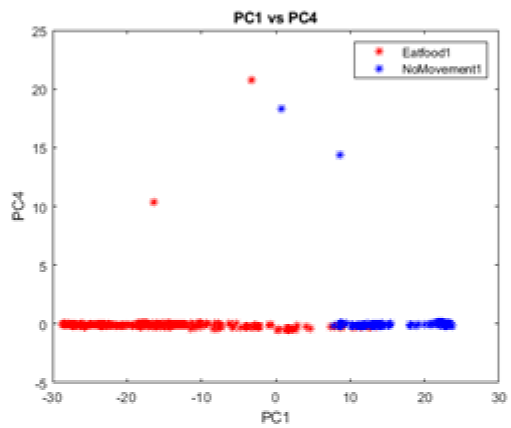
Figure 12. Plot showing the contribution of each feature to the Eigen vectors.

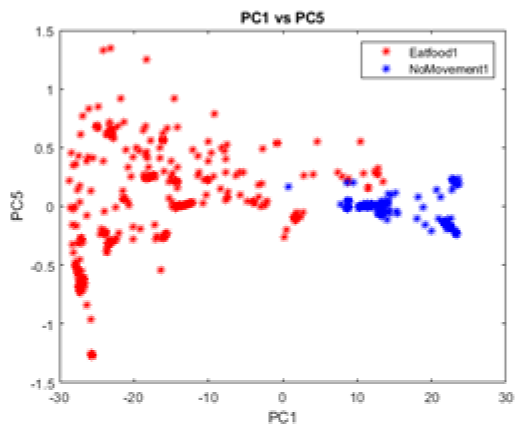
Subtask 3: **Making sense of the PCA's Eigenvectors**

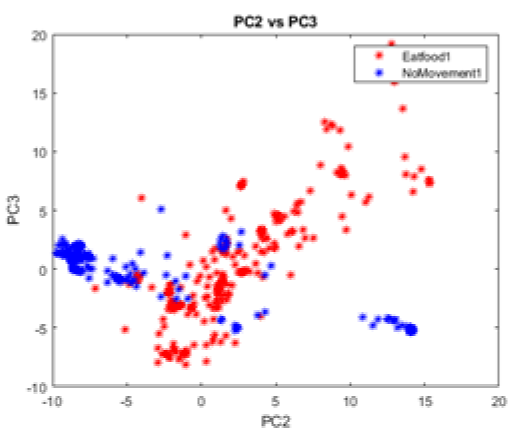
The above figure shows that the Fast Fourier Transform (FFT) feature in the z,x,y and w directions (taken from the orientation dataset of the two activities) play a significant role in influencing all the principal components chosen. The FFT in the w-direction stays the same in most of the principal components since it's a homogenous vertex coordinate. However, in the fifth principal component, the contributions of all features are low since the fifth principal component retains 0.026% of the total variance.

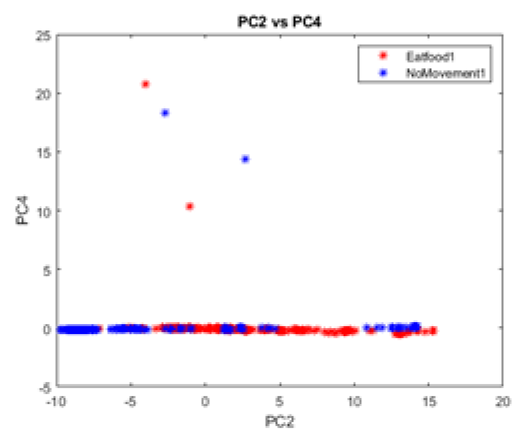

PC1 vs PC2
Eatfood1
NoMovement1
PC1
PC2


PC1 vs PC3
Eatfood1
NoMovement1
PC1
PC3


PC1 vs PC4
Eatfood1
NoMovement1
PC1
PC4


PC1 vs PC5
Eatfood1
NoMovement1
PC1
PC5


PC2 vs PC3
Eatfood1
NoMovement1
PC2
PC3


PC2 vs PC4
Eatfood1
NoMovement1
PC2
PC4

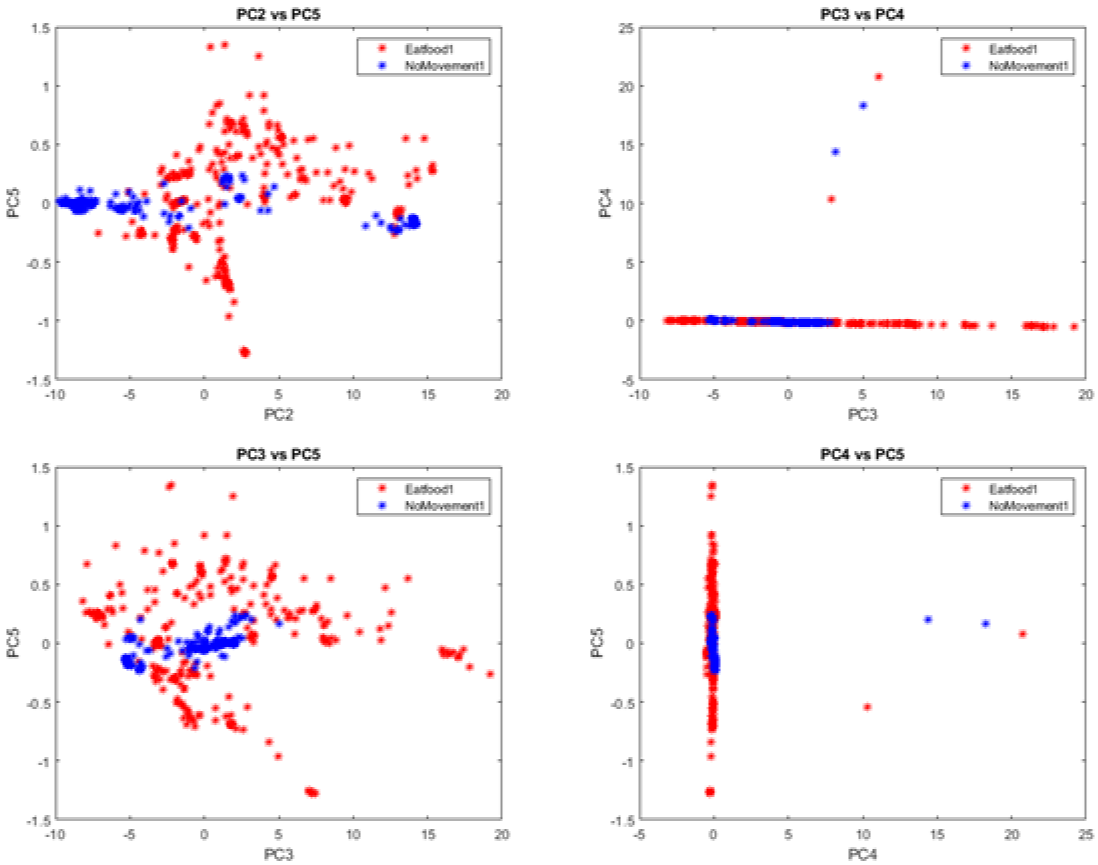


Fig 13. Plots representing the projection of the original feature matrix onto all possible combinations of the first five principal components.

Subtask 4: **Results of PCA**

When it comes to distinguishing the 2 activities, scatter plots with PC1 as one of the dimensions gives better representation than the other scatter plots with combinations not involving PC1 as one of the dimensions since PC1 retains 79.07% of the total variance. There are also few anomalies/outliers in all of the scatter plots.

Subtask 5: **Argue whether doing PCA was helpful or not**

If we observe the scatterplots before applying PCA, we can notice that the signals of orientation in the original feature space are well separable. This is in accordance to the orientation signals having higher spread than the other signals. Therefore, we expected that the eigenvectors of the covariance matrix resulting as an output of executing PCA to have dominant orientation components. However, the signals for accelerometer in the x and y directions were not good features to aid in distinguishing the two activities. The scatterplots obtained after applying PCA were in accordance to what we expected. Hence, PCA was helpful in efficiently distinguishing the eating and non-eating activity.

5. References:

[1]Sushmalekha Shankar Birur [2018], *"Hand Gesture Recognition using Unsupervised Learning"*
[2]Smithsonian Air and Space Musuem, "Roll, Pitch and Yaw", https://howthingsfly.si.edu/flight-dynamics/roll-pitch-and-yaw
[3]Eider C. P. Silva, Esteban W. G. Clua, Anselmo A. Montenegro, 2015 14th Brazilian Symposium on Computer Games and Digital Entertainment, *"Sensor Data Fusion for full arm* tracking using Myo armband and Leap Motion"
[4]Principal Component Analysis, https://en.wikipedia.org/wiki/Principal_component_analysis
[5]MathWorks Documentation for PCA, https://www.mathworks.com/help/stats/pca.html