Internet Of Things

Activity Classification

Mili Goyal (IMT2017513)

Lakshay Agarwal (IMT2017026)

Introduction

This report presents a project which aims at detecting human activities like standing, sitting, walking, jogging, running and climbing up/down the stairs using decision fusion algorithms, signal processing , and various techniques of machine learning. We propose to build a machine learning model which helps us to classify these activities based on the collected data.

Dataset

The data was collected using an application called ‘SensorRecord’ using cell phones which allows a user to use a variety of different sensors. The data we have collected uses two sensors : accelerometer and gyroscope.

<u>Accelerometer sensor</u>: The accelerometer sensor can be used to measure the acceleration exerted upon the sensor. Usually the acceleration is given in two or three axis-vector components that make up the sum/net acceleration.

<u>Gyroscope sensor</u> : It is a device that can measure and maintain the orientation and angular velocity of an object. These are more advanced than accelerometers. These can measure the tilt and lateral orientation of the object whereas accelerometer can only measure the linear motion.

Collection of a good dataset is one of the most important parts for applying various machine learning concepts. Therefore to gather a good dataset we use this app and collect accelerometer and gyroscope data using different cellular phones by a number of different people varying over different age groups so that our data is whole some and covers various different dynamics. This process of data collection allows us to have a lot of variance in our dataset which ultimately will help us in achieving better classification results from our machine learning model.

The different attributes present in our dataset our -

- Timestamp
- Milliseconds
- Acceleration-x
- Acceleration-y
- Acceleration-z
- Gyro-x
- Gyro-y
- Gyro-z
- Label

Here the (acceleration-x, acceleration-y, acceleration-z), (gyro-x, gyro-y, gyro-z) respectively cover the acceleration and angular velocity along the x, y, z axis.

The labels for different activity types are -

Activity Type	Label
Walking	1
Running	2
Jumping	3
climbing	4
sitting	5
standing	6

Data Analysis

Data analysis is an approach to analyse the collected data and detect the main characteristics of data using some visualization techniques.

When the phone is held in the upright position with the screen facing to the user, the Z axis points to the outside of the screen, the X axis is horizontal and points to the right, the Y axis is vertical and points up

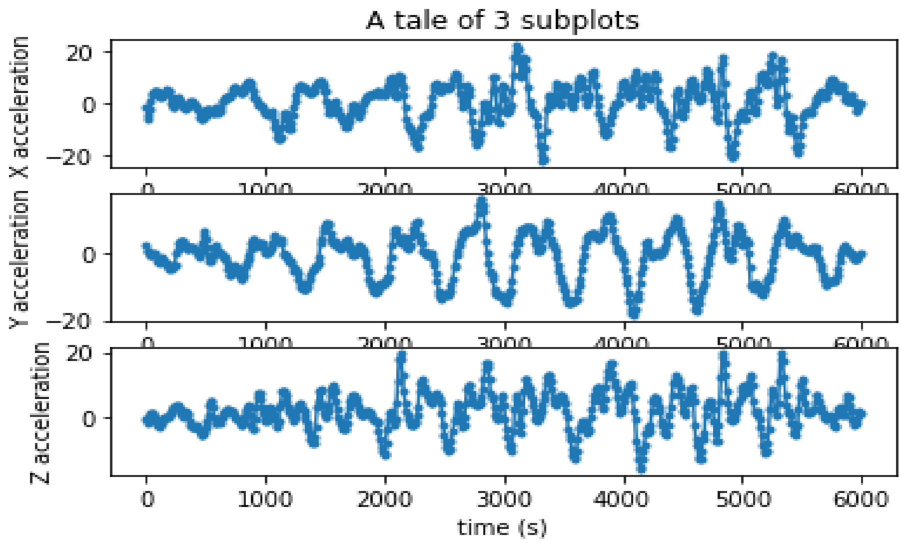
From our analysis of the sensor (Accelerometer and Gyroscope) data collected from the android mobile phones, we can say that :

The sensor data depends a lot on the orientation. A proper orientation is a must to get the proper data and valid results. So, we were very careful with holding the mobile phones while collecting the data.

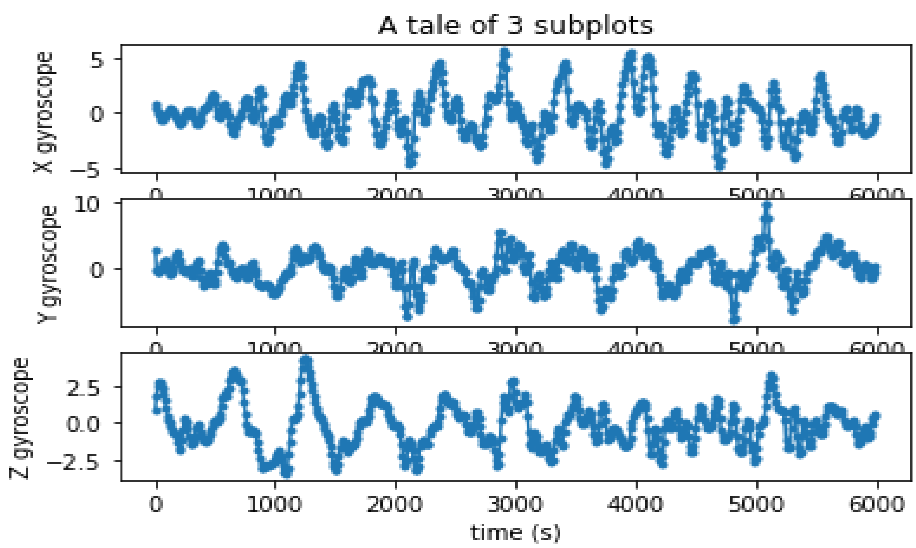
The most important error source of an accelerometer is the bias. The bias is the offset of its output signal from the true value, in m/s2.

Here are some of the different plots which will help us analyse our accelerometer and gyroscope data along different axis of the phone.

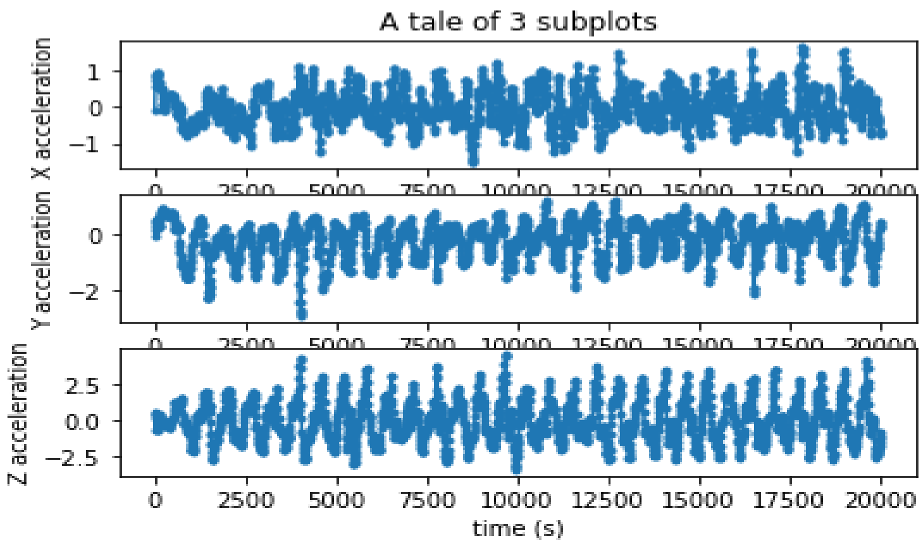
Accelerometer values while running -



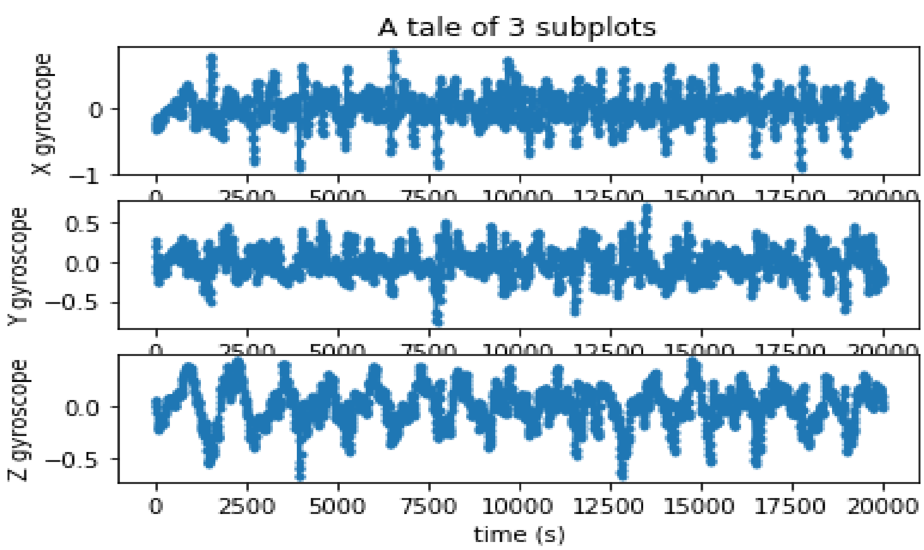
Gyroscope values while running -



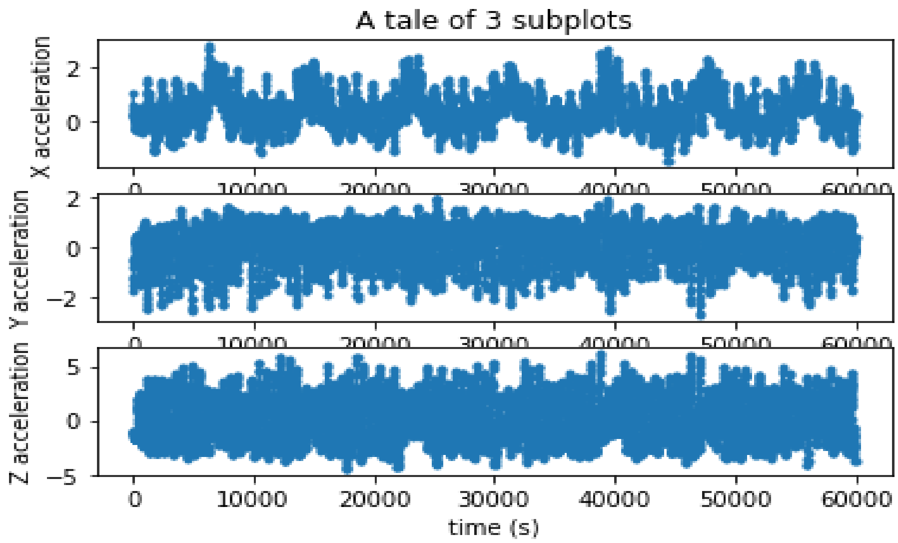
Accelerometer values while walking -



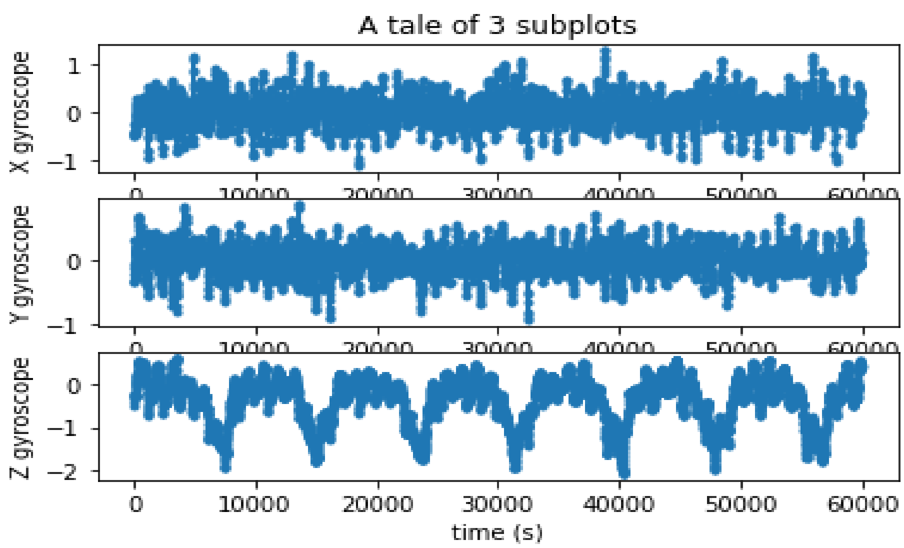
Gyroscope values while walking -



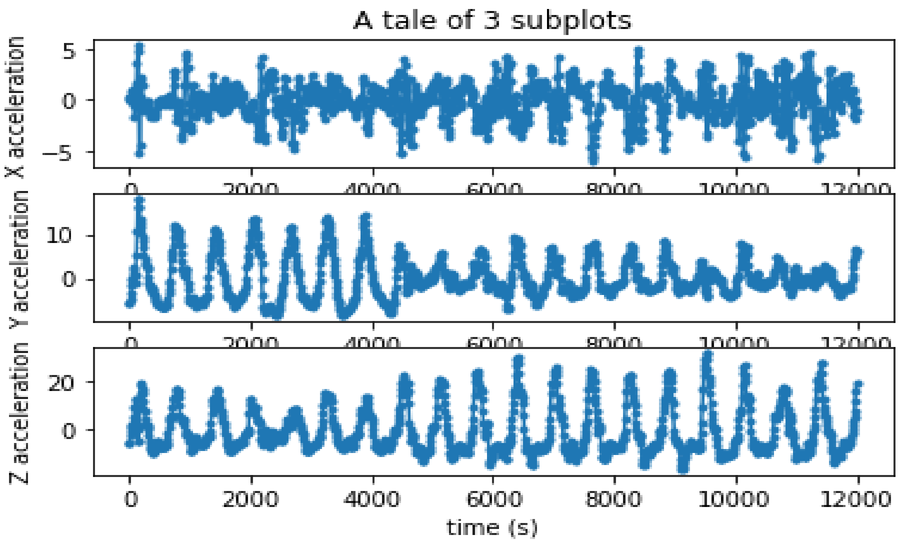
Accelerometer values while climbing -



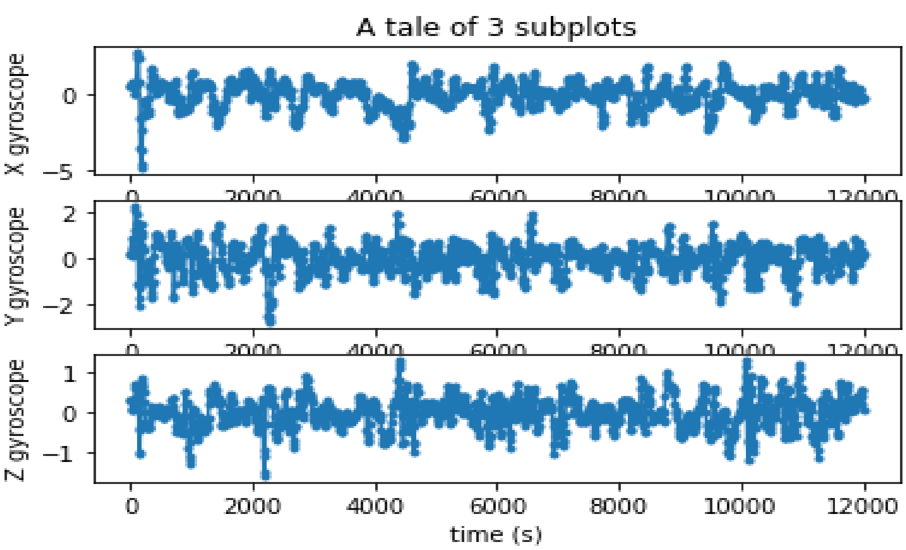
Gyroscope values while climbing -



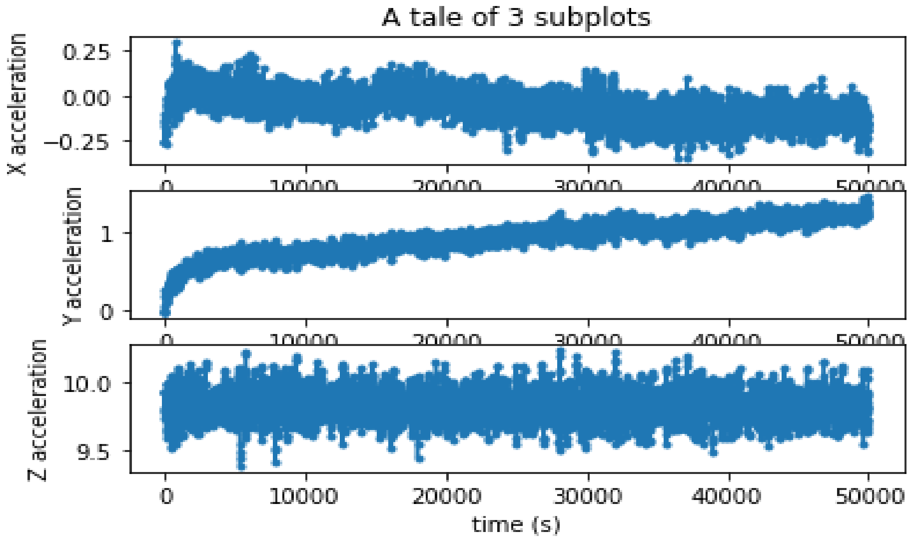
Accelerometer values while jumping -



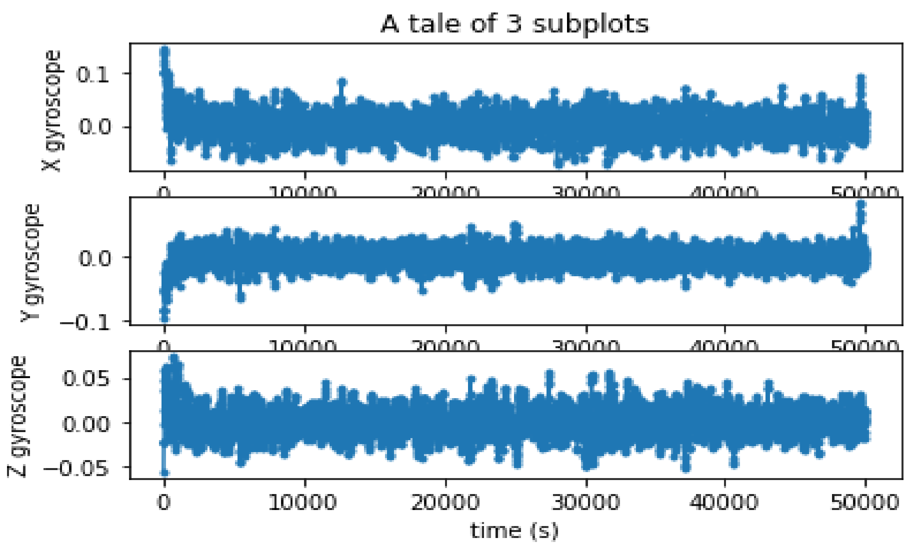
Gyroscope values while jumping -



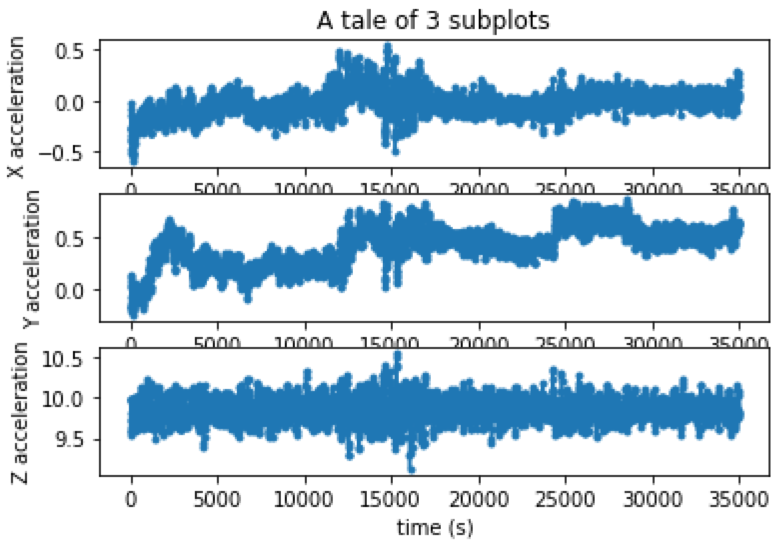
Accelerometer values while sitting -



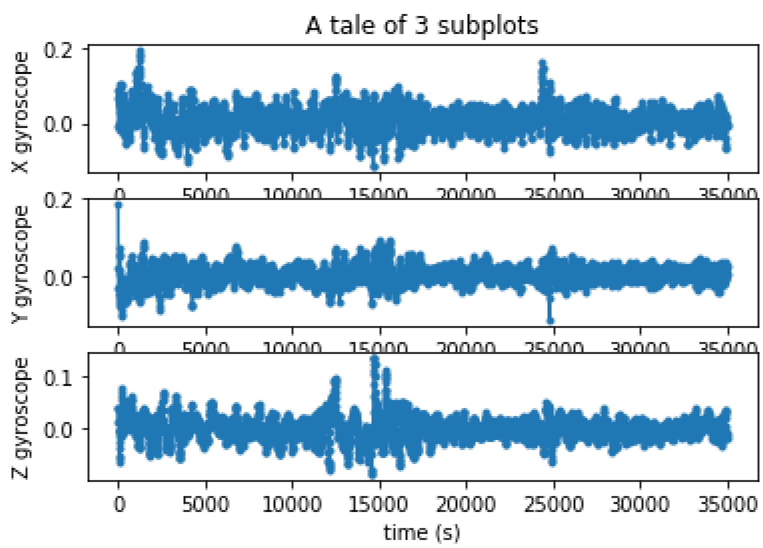
Gyroscope values while sitting -



Accelerometer values while standing -



Gyroscope values while standing -



Feature Engineering

For training our proposed machine learning model we require data points which present certain features in them. These features are the essential part of the model which will ultimately help us classify the different activities with respect to their features correctly.

To extract different features from the time series data, the time series data was divided into windows of 1 second (containing 100 sampled points each) which was regarded as subject.

The different features that we extracted are -

- Mean : This is to collect the mean of all the values within a window taken. (As we have many timestamp data present in a single window, mean serves as a good feature)
- Max : We take the max of the values among a given window of data. This gives us an upper bound for that window.
- Min : Min gives us the min value for the given data window. It serves as a lower bound and thus provides necessary information for modelling.
- Mode : This is the data value that appears most often in the data window.
- Median : Provides the medium of the data in the data window.
- Root Mean squared (RMS) value
- Standard deviation : This also serves as a good feature as it helps us in tracking the variations of data in the specified window of data.

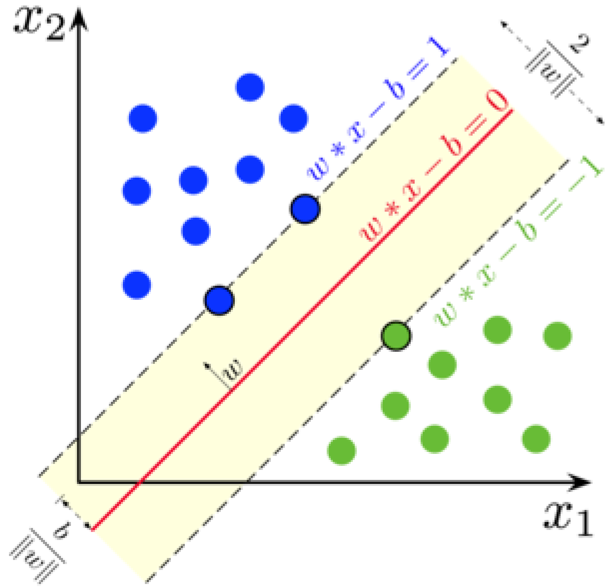
All these values were computed for each of the coordinate values of the accelerometer and gyroscope.

Model Building

For the model building part, we use the SVM classifier to classify the given data into the subsequent category of sitting, walking, running, jumping, standing and climbing.

Support Vector Machine (SVM) Classifier : The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space(N — the number of features) that distinctly classifies the data points.

To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence.



We form a feature vector of all the features obtained and feed them into the training classifier. The result we get is as follows :

We get around 80% classification accuracy with our features and the model.

Our accuracy can even be improved if we have even more data. For the purpose of this assignment, we ourselves collected the data and that was limited.