**HumanActivityRecognition**

**Project Description**

This project is to build a model that predicts the human activities such as Walking, Walking\_Upstairs, Walking\_Downstairs, Sitting, Standing or Laying.

This dataset is collected from 30 persons(referred as subjects in this dataset), performing different activities with a smartphone to their waists. The data is recorded with the help of sensors (accelerometer and Gyroscope) in that smartphone. This experiment was video recorded to label the data manually.

**How data was recorded**

By using the sensors(Gyroscope and accelerometer) in a smartphone, they have captured '3-axial linear acceleration'(tAcc-XYZ) from accelerometer and '3-axial angular velocity' (tGyro-XYZ) from Gyroscope with several variations.

prefix 't' in those metrics denotes time.

suffix 'XYZ' represents 3-axial signals in X , Y, and Z directions.

**Feature names:**

1. These sensor signals are preprocessed by applying noise filters and then sampled in fixed-width windows(sliding windows) of 2.56 seconds each with 50% overlap. ie., each window has 128 readings.

2. From Each window, a feature vector was obtianed by calculating variables from the time and frequency domain.

In our dataset, each datapoint represents a window with different readings

3. The accelertion signal was saperated into Body and Gravity acceleration signals(tBodyAcc-XYZ and tGravityAcc-XYZ) using some low pass filter with corner frequecy of 0.3Hz.

4. After that, the body linear acceleration and angular velocity were derived in time to obtian jerk signals (tBodyAccJerk-XYZ and tBodyGyroJerk-XYZ).

5. The magnitude of these 3-dimensional signals were calculated using the Euclidian norm. This magnitudes are represented as features with names like tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag and tBodyGyroJerkMag.

6. Finally, We've got frequency domain signals from some of the available signals by applying a FFT (Fast Fourier Transform). These signals obtained were labeled with prefix 'f' just like original signals with prefix 't'. These signals are labeled as fBodyAcc-XYZ, fBodyGyroMag etc.,.

7. These are the signals that we got so far.

* tBodyAcc-XYZ
* tGravityAcc-XYZ
* tBodyAccJerk-XYZ
* tBodyGyro-XYZ
* tBodyGyroJerk-XYZ
* tBodyAccMag
* tGravityAccMag
* tBodyAccJerkMag
* tBodyGyroMag
* tBodyGyroJerkMag
* fBodyAcc-XYZ
* fBodyAccJerk-XYZ
* fBodyGyro-XYZ
* fBodyAccMag
* fBodyAccJerkMag
* fBodyGyroMag
* fBodyGyroJerkMag

8. We can esitmate some set of variables from the above signals. ie., We will estimate the following properties on each and every signal that we recoreded so far.

**mean():** Mean value

**std():** Standard deviation

**mad():** Median absolute deviation

**max():** Largest value in array

**min():** Smallest value in array

**sma():** Signal magnitude area

**energy():** Energy measure. Sum of the squares divided by the number of values. **iqr():** Interquartile range

**entropy():** Signal entropy

**arCoeff():** Autorregresion coefficients with Burg order equal to 4

**correlation():** correlation coefficient between two signals

**maxInds():** index of the frequency component with largest magnitude

**meanFreq():** Weighted average of the frequency components to obtain a mean frequency

**skewness():** skewness of the frequency domain signal

**kurtosis():** kurtosis of the frequency domain signal

**bandsEnergy():** Energy of a frequency interval within the 64 bins of the FFT of each window.

**angle():** Angle between to vectors.

9. We can obtain some other vectors by taking the average of signals in a single window sample. These are used on the angle() variable' `

* gravityMean
* tBodyAccMean
* tBodyAccJerkMean
* tBodyGyroMean
* tBodyGyroJerkMean

**Y\_Labels(Encoded):** In the dataset, Y\_labels are represented as numbers from 1 to 6 as their identifiers.

WALKING as 1

WALKING\_UPSTAIRS as 2

WALKING\_DOWNSTAIRS as 3

SITTING as 4

STANDING as 5

LAYING as 6

**Train and test data were saperated**

The readings from 70% of the volunteers were taken as trianing data and remaining 30% subjects recordings were taken for test data

Data All the data is present in 'UCI\_HAR\_dataset/' folder in present working directory.

* Feature names are present in 'UCI\_HAR\_dataset/features.txt'
* Train Data
* 'UCI\_HAR\_dataset/train/X\_train.txt'
* 'UCI\_HAR\_dataset/train/subject\_train.txt'
* 'UCI\_HAR\_dataset/train/y\_train.txt'
* Test Data
* 'UCI\_HAR\_dataset/test/X\_test.txt'
* 'UCI\_HAR\_dataset/test/subject\_test.txt'
* 'UCI\_HAR\_dataset/test/y\_test.txt'

**Data Size :**

27 MB

**Quick overview of the dataset :**

Accelerometer and Gyroscope readings are taken from 30 volunteers(referred as subjects) while performing the following 6 Activities.

1. Walking

2. WalkingUpstairs

3. WalkingDownstairs

4. Standing

5. Sitting

6. Lying.

Readings are divided into a window of 2.56 seconds with 50% overlapping. Accelerometer readings are divided into gravity acceleration and body acceleration readings, which has x,y and z components each. Gyroscope readings are the measure of angular velocities which has x,y and z components. Jerk signals are calculated for BodyAcceleration readings. Fourier Transforms are made on the above time readings to obtain frequency readings. Now, on all the base signal readings., mean, max, mad, sma, arcoefficient, engerybands,entropy etc., are calculated for each window. We get a feature vector of 561 features and these features are given in the dataset. Each window of readings is a datapoint of 561 features.

**Problem Framework :** 30 subjects(volunteers) data is randomly split to 70%(21) test and 30%(7) train data. Each datapoint corresponds one of the 6 Activities.

**Problem Statement :** Given a new datapoint we have to predict the Activity