Goal of the assignment

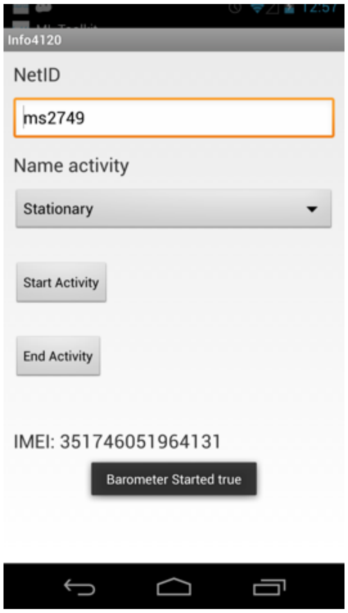
The goal of this assignment is to develop an automatic activity recognition system with the smartphone sensor data. More specifically, the goal is to classify different physical activities like walking, walking up and down stairs, going up and down elevators, and running with smartphone's built-in accelerometer and barometer sensors.

Software

We will provide an Android app that can gather gyroscope, accelerometer and barometer sensor data.

Data Collection

- Each student should use a unique NetID and collect data using the application provided.
- Using the application is straightforward. After entering NetID, select the activity that you are doing next, and press "start activity" [This step will show "Barometer started false" if don't have barometer sensor]. When you finish the activity press "end activity". Also, during activity put the phone in your pocket. You don't need to worry about being precise about start and end. We will discard first 2 seconds and last 2 seconds of the activity you are doing so that you have enough time to start and end an activity and put the phone in your pocket.
- We will collect two sets of data with respect where we hold the device-in your pocket/in the hand from each of following activity
  - Collect 2 min of stationary data
  - Walk-on-flat-surface for approximately 2 min
  - Walking-upstairs
  - Walking-downstairs
  - Take elevator-up
  - Take elevator-down
  - Run on the flat surface for 2 mins



NetID

ms2749

Name activity

Stationary

IMEI: 351746051964131

Barometer Started true

(Try to take stairs (up and down) and elevator (up and down) multiple times so that you get 2 mins of data approximately for each. You don't need to record walking in between these tasks).

- Once the data collection is over, please get the files from the internal storage of your android phone (look for a folder called INFO4120), zip it and share it back to us (a google drive folder link will be provided on the course piazza page).
- Lastly, please download the A2_MatlabCode.zip file and unzip it.
  Link: https://www.dropbox.com/s/4oyr7fbg6afymjo/A2_MatlabCode.zip?dl=0
- Please download your and everyone else's data from the following link. If you can't find your folder here then you may have uploaded on the Google Drive folder late. In that case, you need to download your data separately from the Google Drive folder.
  https://www.dropbox.com/s/s0usllo5n4upup0/CS590U_A2_Data.zip?dl=0
- Place your data and everyone else's data in the A2_MatlabCode/allData folder. You should already be able to see some folders there which was collected previously.

Data Analysis and Interpretation

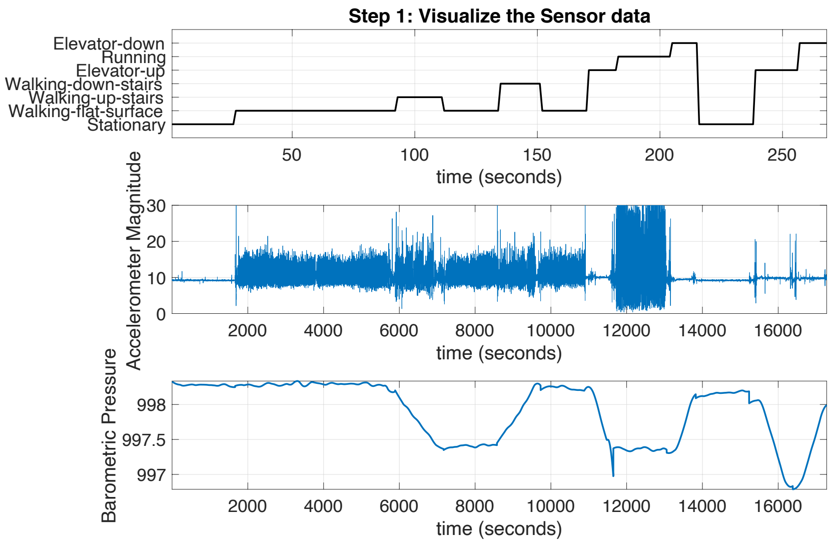


Figure 1: A similar visualization will be created by the Matlab script that was provided. Here we are visualizing barometric pressure and accelerometer magnitude time series for different activity classes.

- **Step 1 (10 Points):** Open main_student.m file and put your net id in line number 7.

For example.

```
my_netid = 'aaleksanyan'; %<---------- input your net id
```

Now run the script to visualize the sensor data for different activities. Figure 1 shows a plot which will be automatically generated by the Matlab script. Do you see any interesting trend in the magnitude of the 3D accelerometer data and the raw barometer data? Explain what you see. Can you think of any specific feature that might help us to differentiate among different activities? You will write a paragraph on this.

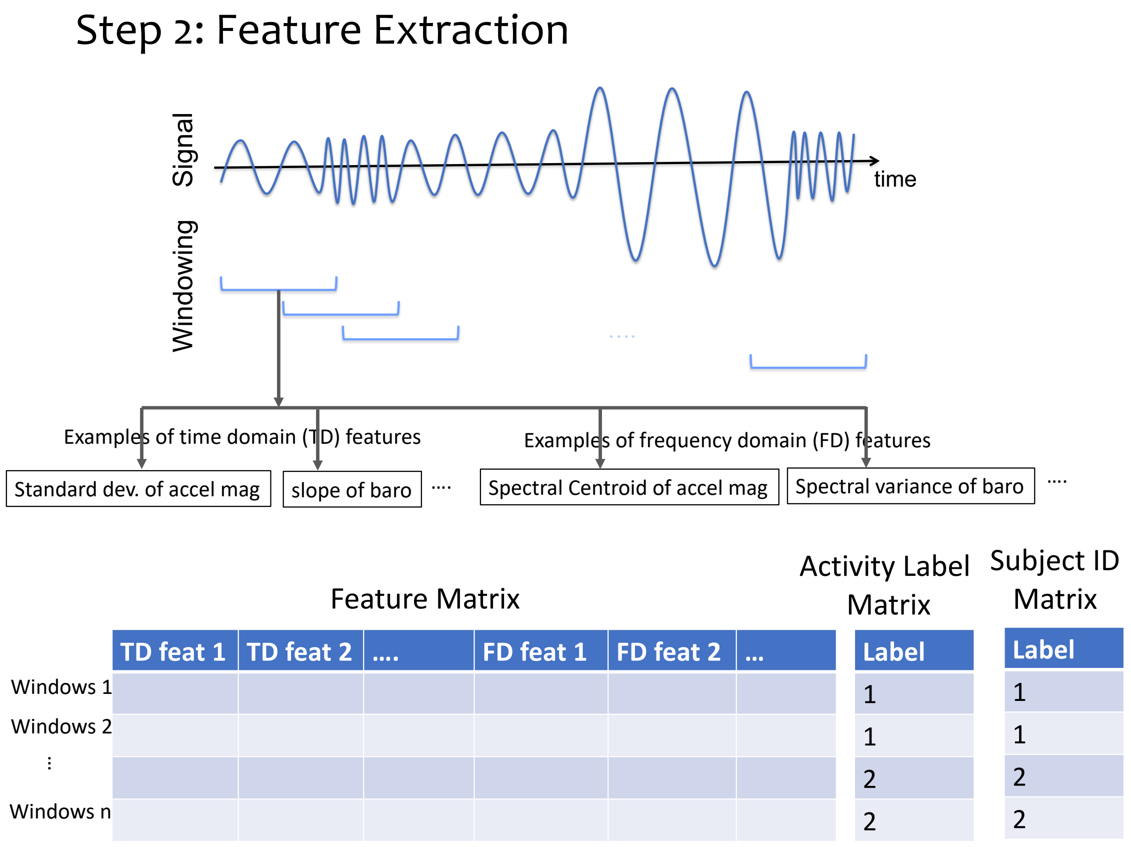
- **Step 2 (15+15 Points):** You will extract features from the accelerometer magnitude and barometric pressure time series data. The windowing will be done by a sliding window where the length of the window will be 10 seconds and window shift will be 2 seconds. Here you have to extract two sets of features- a) time domain (TD) features, b) frequency domain (FD) features. You will need to implement 10 time domain functions and 5 frequency domain functions to extract a total of 20 time domain features (10 accel mag TD features and 10 barometer TD features) and 10 frequency domain features (5 accel mag FD features and 5 barometer FD features).

10 Time Domain Functions	median Standard Deviation Skewness Mean Crossing Rate Slope (fit a line and estimate the m from y=mx+c) Interquartile range 25th Percentile Number of peaks Mean peak values Mean peak distance
5 Frequency Domain Functions	Spectral Centroid Spectral Spread Spectral Rolloff 75%

	Filter bank (divide entire spectra into 2 equal frequency ranges and estimate the signal energy in each of the freq ranges)

You will implement these features in the extractFeatures.m file and put all the features in the featureVector matrix which will be returned to the main_student.m file in line number 72. All the features extracted from all the subjects' directory will be appended together in the featureMatrix matrix in line number 79 in main_student.m

The figure below outlines the complete feature extraction pipeline.



Step 2: Feature Extraction

Examples of time domain (TD) features

Standard dev. of accel mag | slope of baro | ....

Examples of frequency domain (FD) features

Spectral Centroid of accel mag | Spectral variance of baro | ....

Feature Matrix

	TD feat 1	TD feat 2	....	FD feat 1	FD feat 2	...
Windows 1						
Windows 2						
⋮						
Windows n						

Activity Label Matrix

Label
1
1
2
2

Subject ID Matrix

Label
1
1
2
2

- **Step 3 (20 Points):** Now train a K-nearest neighbor classifier with everyone else's data and test on your data. Please make sure that your data remains unseen by the trained classifier. Compute the confusion matrix and then compute the precision, recall and F1 score for each activity separately. Please share all the performance metrics in your report in a tabular form. Now repeat this step with a Random Forest classifier, compare the performance of these two classifiers and comment on it.

In order to achieve this, you will need to operate on featureMatrix variable. In featureMatrix variable the last column in subject id where your own netid will get the value of zero. The second to last column of featureMatrix contains activity labels.

- **Step 4 (20 Points):** Now train and test your classifier with only time domain features and compute precision, recall and accuracy for each activity separately. Similarly, train and test your classifier with only frequency domain features and compute all the same performance metrics (Confusion matrix, precision, recall and F1 score). Pick a classifier for this step. Use the same train-test partition as in step 3. Which domain (i.e., frequency domain vs time domain) is helping you the most in terms of building your activity recognition system? Comment on it.
- **Step 5 (20 Points):** So far, we have tested our classifier only on your own data. In order to get a more robust, reliable and unbiased performance estimate, you will implement **Leave One Subject Out Crossvalidation**. If you have in total N subjects in your dataset, you will run N iterations where in each iteration you will train a classifier with N-1 subject's data and test it on the remaining subject's data. In this manner, in each iteration we will test the trained classifier on a different subject's data. We will accumulate the confusion matrix from each of the iteration and then at the end estimate the recall, precision and f1 score with the accumulated confusion matrix. For this task, you are free to select any classifier you like.

Let's imagine with we 3 subjects in our dataset: A, B, C		
Iteration 1	Iteration 2	Iteration 3
Train on A and B Test on C	Train on B and C Test on A	Train on A and C Test on B

Deliverables:

- Zip your report, and code and upload on Moodle.