Human Activity Recognition with Smartphones

1 Who's In this group?

- Alireza Rezazadeh, UIN: 654852032
- Mindaugas Varaska, UIN: 671096337
- Jason Yeoh, UIN: 670134303
- Faraz Shakir, UIN: 656857563
- Rohit Saikrishnan Ramesh, UIN:667804521
- Thomas Leuenberger, UIN: 659058340

2 What Data Set?

The proposed data set for this project is "Human Activity Recognition with Smartphones" from kaggle.com. This data set was created by University of California–Irvine, and utilizes smart phones to collect data about peoples' daily activities.

The goal of this data collection is to categorize activities into one of six different categories - walking, ascending stairs, descending stairs, sitting, standing, and lying. The data set is made up of information obtained from smartphone sensors. This project would seek to use the data obtained from the smart phones and machine learning techniques from class to classify this senor data into one of these six categories.

Data was gathered from 30 participants with ages between 19-48. Each person spent time performing one of the six classification activities while the smartphone sensors recorded data about the participant's movement and activities. The participants were video recorded while they performed these tasks, thus allowing for human labeling.

The dataset contains some digitized analog signals from the accelerometer and gyroscope. The creators of this dataset did pre-processing on these signals to avoid dealing with noisy data. That is, they applied noise filters to these signals. Additionally, a butterworth (a filter as flat as possible in the pass band) filter was used to separate gravitational force from the acceleration data. During this filtering the sensor signals were sampled and each record in the data set contains a window of these digitized analog signals with both time and frequency domain information. This means the Fourier transform was performed on these windows, and this frequency information provides yet another classification tool.

3 Machine Learning tasks

The goal of this project is multifold, first the project sets out to build a classifier using supervised learning. The goal of the classifier would be to assign a given record into one of the six possible activity categories. Second, the project sets out to use clustering to identify clusters in the data. Before clustering, the data will be flattened to 2-D to allow for some level of visualization. The clusters would then be analyzed for any new insights into the data.

4 Techniques

Overall, this project sets out to implement multivariate supervised learning techniques. As previously mentioned, the data set contains labeled records, and each record contains the following useful information:

- Triaxial acceleration from the accelerometer (total acceleration) and the estimated body acceleration.
- Triaxial Angular velocity from the gyroscope.
- A 561-feature vector with time and frequency domain variables.
- Its activity label.
- An identifier of the subject who carried out the experiment.

This project would set out to take a multi variant supervised learning approach to classifying this data into one of the six categories provided above. The techniques described in chapter 5 of Alpaydin would be used to build the classifier. Note that the dataset has been partitioned into separate training and testing sets, with a split of 70:30 on all samples. In order to maximize the accuracy rate on the test set, the project will utilize k-fold cross validation on the training set.

Clustering would involve the use of KNN cluster, using k=6. This portion of the project will be an application of chapter 7 of Alpaydin. Additionally, the data will be flattened using PCA/ tSNE.