PROJECT PROPOSAL- ACTIVITY RECOGNITION

Team
Chidambaram Crushev
Preeti Maurya
JunYong Lee
Stephanie

Motivation :

The data collected from activities recognized on smartphones can help in customizing the smartphone screen features accordingly. The dataset used in this project contains the records and features related to smartphone data which helps to predict or classify what activity an individual is currently doing based on the data. For instance, a user's activities involve whether he is standing, sitting, laying, walking, etc. Classification algorithms are one of the well known algorithms in supervised machine learning where the training dataset is labelled. Using classification algorithms on existing datasets, a model is trained to predict the classes new data can belong to. There are different classification algorithms such as, Perceptron, Adaline, Stochastic Gradient Descent, Decision Tree, SVM, Bayes classifier which can be applied to this dataset and recognize the activity of an individual.

Problem Description :

The dataset size is more than 60 MB and has 562 columns. These features data are collected from the smartphone data and help us to recognize the activity of an individual whether he is Standing, Sitting, Laying, Walking, Walking_downstairs and Walking_upstairs. For data collection, accelerometer and gyroscope readings were taken when the individuals were performing one of the above activities. These large features(561 columns) are extracted based on the accelerometer and gyroscope readings and they underwent various mathematical operations(Fourier transform). With this huge data, we can address the following questions:

1. What is the activity of an individual based on the smartphone data?
   (We can perform binary and multi-class classification for the data with machine learning algorithms and then make predictions)

2. What is the accuracy of different machine learning models? How to improve those accuracies?

3. Since the data is huge, how will you improve the running time of the algorithm?
   (We need more computing power to process this data and we will be using the NMSU supercomputer (Discovery cluster) to improve the running time.)