

# Machine Learning Engineer Nanodegree

## Capstone Proposal

Vivekananda Ponnaiyan November 17th, 2017

### Proposal

#### Domain Background

Human activity recognition aims to recognize the actions and goals of a human being from a series of observations about their actions and/or environmental conditions. Since the 1980's this has been a very big research field due to its strength in providing personalized support for many different applications and its connection to many different fields of study such as medicine, human-computer interaction and sociology.

I am passionate about health and fitness and one of the largest challenges in getting healthy is maintaining a habit of exercise. One key motivator for exercise is the ability to track one's progress and to see an increase in performance with respect to time. However tracking one's exercises can become a chore and many people don't have the motivation to log all their exercises. Hence a system which can automatically log activities accurately would go a long way in allowing people to recognize the increase (or decrease) in performance and has other applications such as coaching etc.

#### Problem Statement

In this project we aim to train a classifier to be able to distinguish between 6 different activities. Each activity will be modeled as a class label which will be learned using the training set. After that we will use the test set to prove that our model works on unseen data.

#### Datasets and Inputs

The dataset used is the Human Activity Recognition Using Smartphones Data Set from UC Irvine.

The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The experiments have

been video-recorded to label the data manually. The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data.

The sensor signals (accelerometer and gyroscope) were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window). The sensor acceleration signal, which has gravitational and body motion components, was separated using a Butterworth low-pass filter into body acceleration and gravity. The gravitational force is assumed to have only low frequency components, therefore a filter with 0.3 Hz cutoff frequency was used. From each window, a vector of features was obtained by calculating variables from the time and frequency domain.

### **Solution Statement**

The solution will be a computer program which can take in sensor data and then classify it as one of the 6 classes using the VERY NEWEST research in LSTMs. The Classifier will output a label for each activity of the test set.

### **Benchmark Model**

We will use a support vector machine as a benchmark for our classifier. And the evaluation criteria will be the accuracy of the classifier on samples of data containing 1 previously labelled activity.

### **Evaluation Metrics**

We will use the accuracy as our main evaluation metric. Since the data is balanced this should be a good evaluation metric. We will also look at Recall and Precision so as to better validate our classifier.

### **Project Design**

The project will involve the following steps: 1. Analyze the data to make sure that it is a well balanced data set. 2. Pre-process the data to fix all missing values either using 0 or mean as appropriate. 3. Split the data into training and test data. 4. Use cross validation to train a SVM model to gain a benchmark model. 5. Plot the SVM model performance using accuracy, recall and precision. 6. Design and implement a model using LSTM based Recurrent Networks which can leverage the temporal relationships between data. 7. Plot the LSTM model performance using accuracy, recall and precision. 8. Compare both LSTM and SVM.