Human Activity Recognition

Akshay P
Machine Learning Intern
Terra AI solutions
Hyderabad,India

Abstract

In this project,various machine learning models and Deep Neural Networks are used in an attempt to predict Human activity based on HAR UCI dataset which contains different activities recorded with different subjects.

Introduction

The goal of this project was to develop a model to predict human activity precisely.
Project has a great significance in present and future days since it has wide variety applications such in Robotics,surveillance system etc .. I have developed 5 machine learning models, which takes various parameters recorded as inputs and classify the activities.

Dataset & Features

Data set is collected from recordings of 30 human subjects captured via smartphones enabled with embedded inertial sensors. The data set has 10,299 rows and 561 columns.

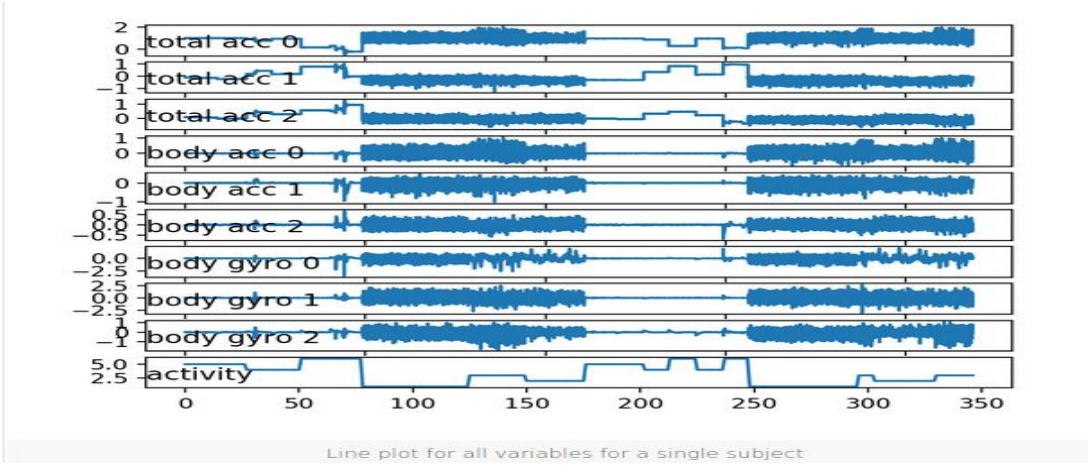
The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer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

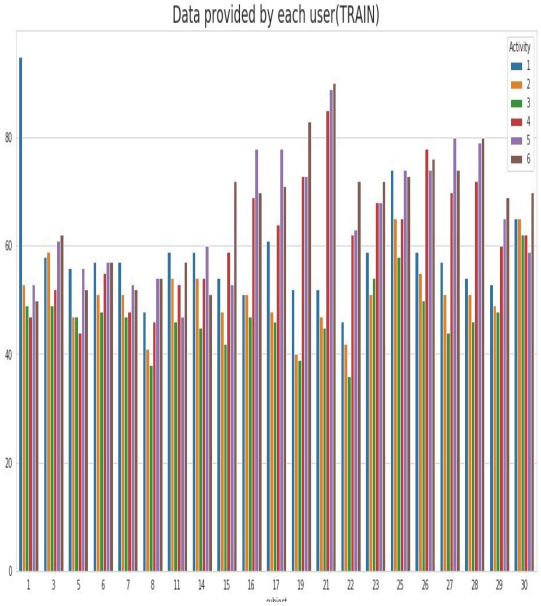
Data visualisation

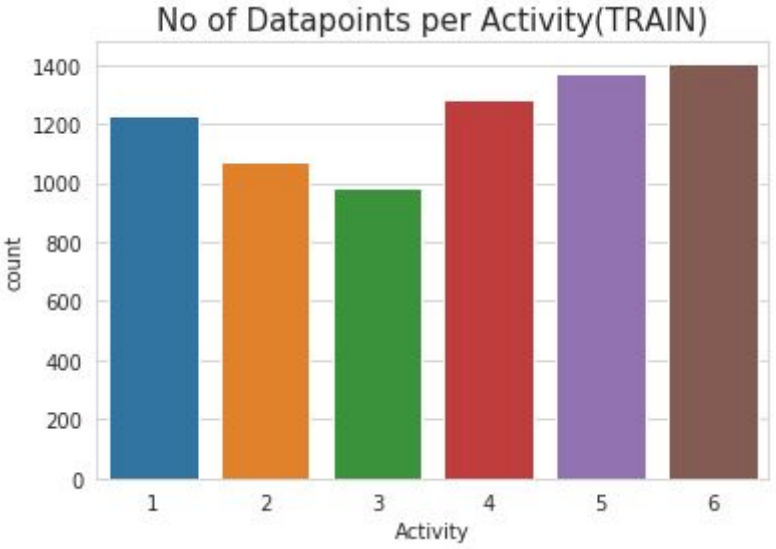
Train dataset is clean and contains no duplicates and Null values.Frequency of activities done by each of the subjects is visualised using a bar graph which shows that certain subjects performed more activities than others.

From plot for a single subject we can observe that this subject has performed the same general sequence of activities twice and some activities are performed more than two times. This suggests that for a given subject we should not make assumptions about what activities may have been performed.



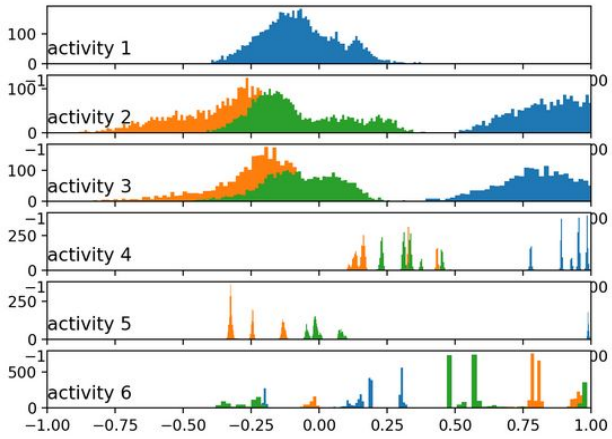
Line plot for all variables for a single subject

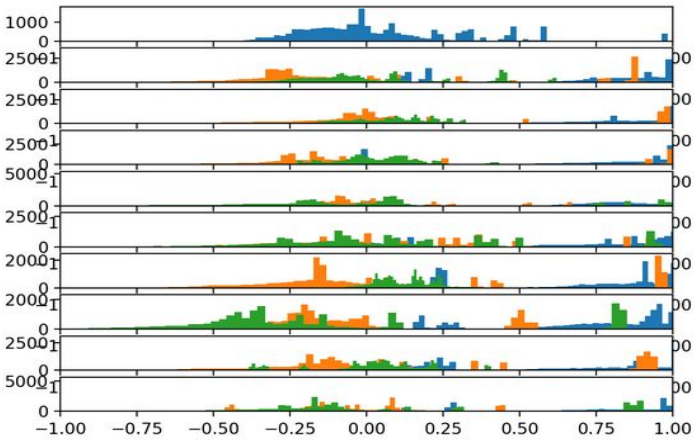




Each of the three axes on a given plot have a different color, x, y, and z are blue, orange, and green respectively.We can see that the distribution for a given axis appear Gaussian with large separate groups of data.

From histogram of total acceleration by activity we can see that each activity has a different data distribution, with a difference between the large movement (first three activities) with the stationary activities (last three activities). Data distributions for the first three activities looks Gaussian.





Models

1.Logistic regression:

Logistic regression is one of the basic model in machine learning which is mainly used for classification purpose.
y = e^(b0 + b1*x) / (1 + e^(b0 + b1*x)) is the logistic regression equation.
Logistic regression with parameters C=1 and penalty=l2 was fitted into dataset.
Accuracy of 0.9387139704425579 was obtained on performing training.

2.Support vector classifier:

Support vector machines is one of the highly complicated and most powerful machine learning Algorithm.
It classifies data by increasing dimensionality and classify datapoints at higher dimension.
Classification with parameters C=1 and kernel=’linear’ was fitted to dataset.
Accuracy of 9400129901439578 was obtained.

3.Decision Tree:

Tree based learning algorithms are considered to be one of the best and mostly used supervised learning methods.They have better accuracy ,stability and easy to interpret.
Here data is classified based on information entropy.

Classification with parameters max_depth=5 and criterion=entropy was fitted into dataset
Accuracy of 0.8711755157535481 was obtained

4.Random Forest:

The random forest is a classification algorithm consisting of many decisions trees. It uses bagging and feature randomness when building each individual tree to try to create an forest of trees whose prediction is more accurate than that of any individual tree.
Classification with parameters max_depth=11 and no.of trees=110 was fitted into dataset.
Accuracy of 0.9211003667272484 was obtained

5.Neural networks:

It is one of the main tool used in machine learning which was inspired by the activity of neurons in brain.ANN performs forward as well as backward propagation multiple times to reduce loss function
Our model uses relu as activation function in hidden layers and softmax at the final layer.
Three hidden layers were used with batch size=32 and epochs=25 to train the model.
Accuracy of 0.9426340801086218 was obtained.

Conclusion

- Variation between static and dynamic datapoints was evident
- In most cases dynamic activity curve is Gaussian
- On performing EDA variation between same activity datapoints was evident
- ANN was found to be best model with accuracy of 0.9426340801086218

References

- Wikipedia activity recognition
- Towardsdatascience.com
- [6] J. Yang, J. Lee, and J. Choi, “Activity Recognition Based on RFID Object Usage for Smart Mobile Devices,” J. Comput. Sci. Technol., vol. 26, no. 2, pp. 239–246, Mar. 2011.