Understanding Human Activity While Using a Smartphone

Mahendra Pruitt

1. Introduction

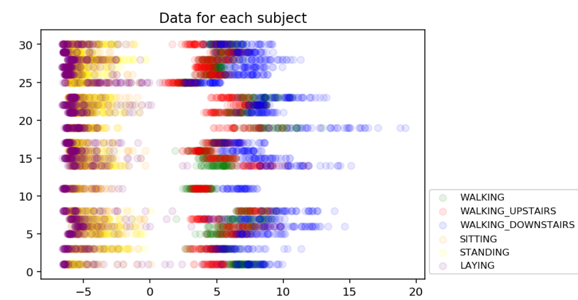
As smartphones are becoming a common device in today's society, they come with a plethora of features. One of those features includes accelerometers and gyroscopes used to detect positioning of the phone and individual to orient the screen properly. Because these devices are so small, we almost always have them in our hand wherever we go. This has caused us to have specific habits when we are doing a certain activity, such as walking or laying down. Therefore, the scope of this project was to see if certain characteristics could be identified based on the data collected from the smartphone's accelerometers and gyroscopes. The information from this data can have applications in programs to track your health or bad habits. For example, the data could be used to tell you what your habits are and how you can improve your day to day health. Specifically, this project focused on trying to build a model to classify six different types of activities

2. Description of the Data Set

This data set was derived from the UCI Machine Learning Repository. The experiment was carried out on 30 volunteers. Each person performed six different activities while wearing a Samsung Galaxy S II on the waist. The smartphone had embedded accelerometer and gyroscopes, so the researchers were able to capture 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The data was recorded manually by the researchers after videoing the experiments. By default, the data was pre-split with 70% being training data and 30% being test data. The data comes accompanied with inertia signals which is raw data that could be used. However, this project focused on the data that was already split, rather than the raw data. There is an updated version of this data set, however this project did not analyze that one. The only difference between the old version and the new version is that the new data set includes labels of transitions between activities, as well as the raw inertia signals instead of the preprocessed ones given in the older set.

3. Visualization of the Data

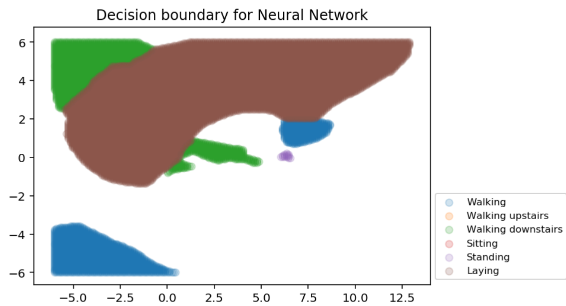
To visualize our data, we imported it into a numpy array and listed it. The y values here go up to 30, for each participant, while the x values represent the signals from the gyroscopes.



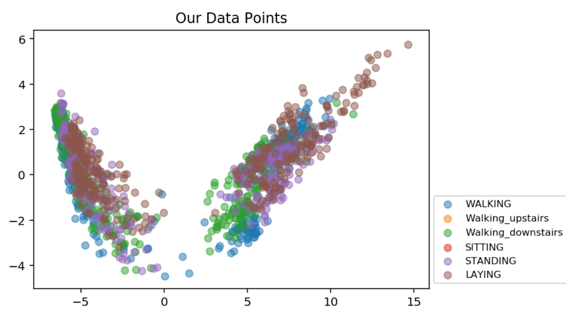
At first glance, we can see a distinct separation of the data where the subject is walking versus where they are standing, sitting, or laying still. In order to present the data in a 2D form, I had to project it from 561 dimensions to 2 dimensions by using PCA.

4. Models and Visualization

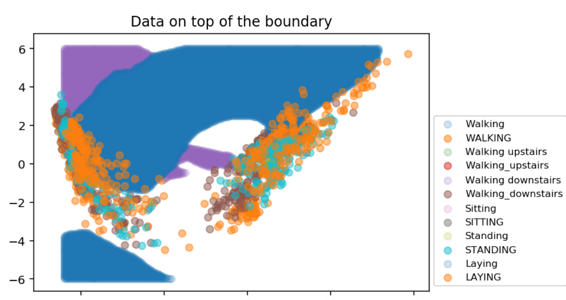
To classify my data, I used the built in multi-layer perceptron classifier and the k nearest neighbor classifier from the sklearn library. The decision boundary presented by this can be viewed below:



Here we view our data points separate from the boundary:

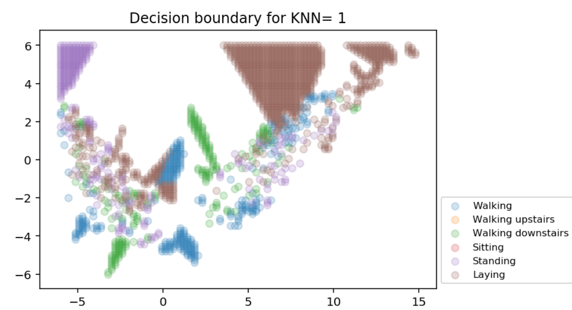


Then, we place these points on top of the boundary to view:

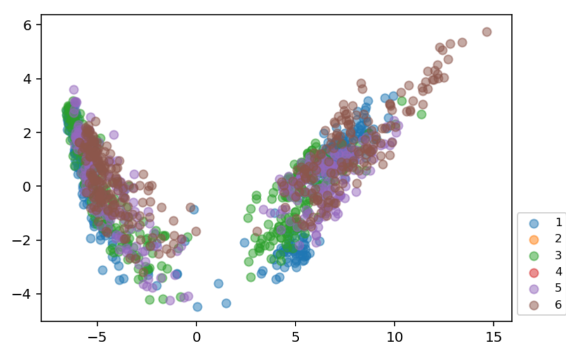


From this, we can see that there is a very high error rate. This is may likely be due to reducing the data down to two dimensional data. When we calculated the error rate produced by reducing the data, we got 83% error. When we calculated the error rate on data that was not projected down, we got around a 5% error rate, which is a vast difference.

When using K Nearest Neighbors, we tested using nearest neighbors from 1 to 10. The best results were when nearest neighbors were 1, as shown below:



And the data points are separate below:



While the data may look like it has high error rate on the boundary, we were able to get a 0% error rate with nearest neighbors = 1 when the data was not projected to be 2 dimensional. When the nearest neighbors were 10, we got a 15% error rate.

5. Results and Conclusion

While it is not visualized accurately in two dimensions, we were able to get proper classification of the data with low error rates. The data was hard to visualize and present due to the large amount of data and high dimensions. However, when we ran our classifiers on our data, we were able to get exceptional results.

6. References

Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. A Public Domain Dataset for Human Activity Recognition Using Smartphones. 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013.