Title : Exercise classifier

Contributions:
Connor Gibbons : Data Collection, Preprocessing, and Model Training
Nicholas Kenney : Data Collection, Visualization and Analysis, and Model Training
Wellan Sok: Data Collection, Feature Extraction, Algorithm Design

Mission:
The goal of this project was to use machine learning to differentiate between different exercises by analyzing patterns in data we recorded. This project focuses on the user being able to accurately record activities throughout their day using their mobile devices, allowing users to gain insight into their daily activities.

Applications:
A classifier capable of determining what workout is being performed by the user would be most useful if implemented within a fitness tracker. For example, the Apple Health app contains information about the duration of various workouts you have done, how many repetitions, and how many calories you burned during the workout – which is heavily dependent on which workout you are doing. Having a classifier such as this would make all of those metrics easier/possible to calculate.

Data Collection, Model Training, and Testing / Analysis:
Data collection for this project involved using a sensor logging application on our smartphones that we placed in our left pockets to reduce variability. Within the application we specifically used the accelerometer and gyroscope sensors. After collecting multiple instances of ~1 minute workouts, we then had to combine the data across the multiple sensors in order to make it usable by our classifier. To do this, we implemented a python script that appended gyroscope data to each respective line in the accelerometer data csv, and saved that as one file. This way, one window of data would include both accelerometer and gyroscope information. All of our data was recorded at 100Hz, and training/classification was done with a window size of 150, or 1.5 seconds. We chose this window because we wanted a window size to include at least 1 repetition of a typical exercise. Among the various features we extracted, we found that the most impactful ones were entropy magnitude, which finds the magnitude over a window, sorts it into buckets, and calculates the entropy of those values. Another useful metric was rms_y, which is the root mean square of the y-axis accelerometer data. Critically, using the gyroscope to calculate pitch, specifically the sum of the pitch values across each window, allowed us to increase our accuracy by 5%, a significant increase, which justified using multiple sensors.

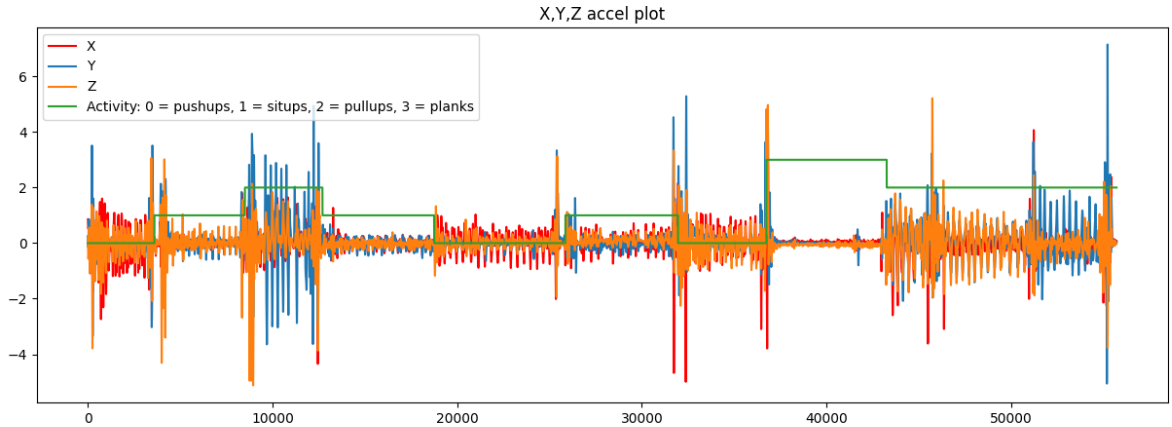
Results

Confusion Matrix and Accuracy with Weight Lifting exercises added:
[actual left to right: pushups, situps, pullups, planks, curls, skullcrushers]

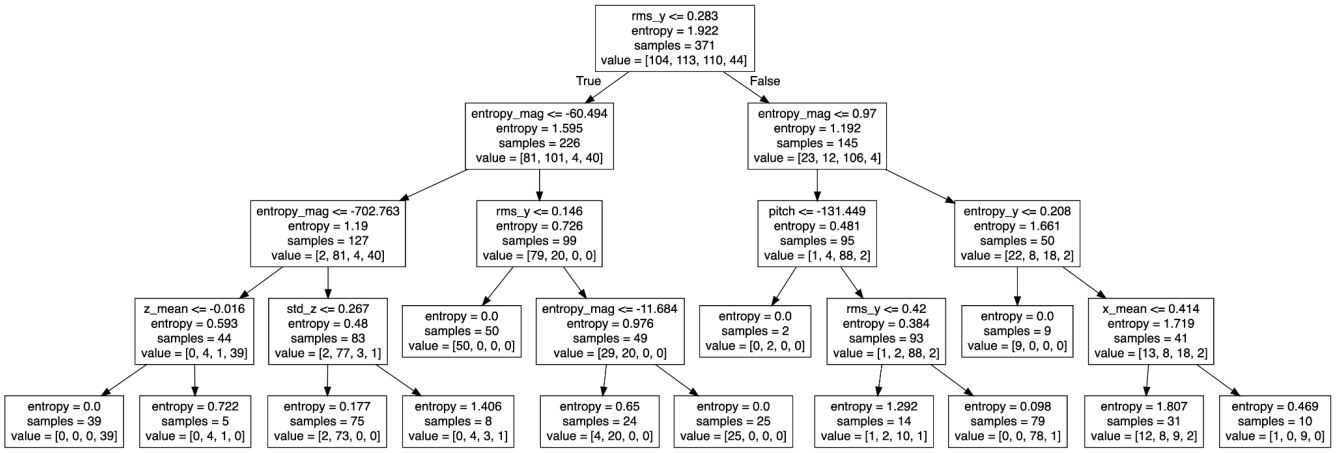
```
[[ 7  0  4  0  0  0]
 [ 1  4  2  0  0  0]
 [ 0  2 11  0  0  0]
 [ 0  0  2  4  0  0]
 [ 0  2  1  0  0  0]
 [ 0  1  3  0  0  0]]
```

```
Accuracy:0.5521141649048624
Precision:0.6916621810882816
Recall:0.4936892492039551
```

Plot of Combined Accelerometer Data with Labels used by Classifier:



Decision Tree made by Final Classifier:



Learnings from the Project:

We learned that more data is oftentimes harder to produce but the value it brings when training a model/classifier is tremendous. We also learned that when collecting our own data we should think of what we can record to make it easier in the future if we do decide to use another sensor, so we wouldn't have to go back and record again. We also learned that by trying to add additional exercises to classify it greatly reduced the accuracy of our classifier. As shown above with curls and skullcrushers, two types of weight lifting exercises, the classifier could not make any sense of them.

How to Improve the Project Further:

We could improve upon this project by fine tuning our algorithm to allow for minimal error when detecting activities and counting repetitions. We also can find more participants to collect a wider variety of data to make sure we can generalize appropriately among a wider audience. We also could expand upon what activities can be detected and counted accurately. It would be interesting to add Magnetometer data. This would allow us to more accurately determine a person's orientation, which would be critical for differentiating better many exercises.