**Using Machine Learning to Achieve Proper Weight Lifting Form**



**Tamara Monge**

May 28, 2017

Springboard Data Science Career Track

Capstone #2 Proposal

1. **What is the problem you want to solve?**

Classify a bicep curl movement sequence into one of 5 form-classes (one class is correct form, the other four are various incorrect forms). If the sequence is classified as incorrect, identify how it is incorrect and recommend changes so the performer may achieve correct form.

1. **Who is your client and why do they care about this problem?**

Two types of clients will be interested in this study. The first client is exercisers – they will want to know if they are using proper form while performing lifts. If the algorithm finds they are using improper form, the exerciser will want to know how to correct it. The second client is personal trainers – they will be interested in the algorithm eventually replacing one of their functions, the “form-coach,” and thereby freeing them to put their time and effort into their other function: designing workouts to achieve the exerciser’s goals.

Later, after data has been collected for a variety of lifts, the algorithm could be developed into a mobile app with 4-6 external electrodes that are placed on the body according to the exercise being performed.

1. **What data are you going to use for this?**

The dataset for this project comes from PUC-Rio’s Human Activity Recognition Group website: <http://groupware.les.inf.puc-rio.br/har>. The data is freely available for download. It is in CSV format.

The dataset comprises 839 movement sequences (bicep curl repetitions), each consisting of anywhere from 1 to 73 timestamp recordings. Each recording has 159 features. The salient features are: user, timestamp, form-class, and position/motion information for four accelerometers. The accelerometers are positioned on the working arm’s bicep, the working arm’s forearm, the belt (navel), and the dumbbell. The position/motion data reported for each sensor is: roll, pitch, and yaw, as well as the raw gyroscope (in x, y, and z), accelerometer, (in x, y, z, and total), and magnetometer (in x, y, and z) readings. The total number of recordings is 39,242 (~48 MB).

1. **In brief, outline your approach to solving this problem.**

My approach will consist of: (1) Data acquisition, (2) Data cleaning, (3) Exploratory Data Analysis, (4) Time Series analysis – making the time series stationary, (5) Supervised Learning classification, (6) Model results and (7) Story writing and recommendations. The supervised learning problem will classify the form into one of five possible classes – Correct and 4 common mistakes.

1. **What are your deliverables?**

My deliverables for this project will be: a summary report, a PowerPoint slide deck, and the code created to perform the analysis.