**Title: Predicting user activity with Samsung phone accelerometer data**

**Introduction:**

Given the ubiquity of “smartphones” or mobile computing device that are frequently carried continually on the person, there is great interest in finding application of these devices in monitoring health and lifestyle habits of users. One possible use of these devices is to track physical exercise by using the device’s built in accelerometer and gyroscope to detect the type of activity that the user is engaged in.

For this purpose, data were collected from 30 users carrying a Samsung Galaxy S II on a waistband and performing various activities. Raw data was pre-processed by applying noise filters and sampling rates. The goal of this data analysis is to, based on the input data from the smartphone device, predict the activity type that the user is engaged in. [1]

**Methods:**

*Data Collection*

For our analysis, data was used from the sample described above. A data file containing 7,352 observations of 563 variables was used. This data set contains 561 variables describing the data from the gyroscope and accelerometer on the smartphone device, along with a variable identifying the user and a variable identifying the activity being performed at the time of the observation.

The data were separated into two sets, a training set (containing data from subjects 1-26) and a testing set (containing data from subjects 27-30).

*Exploratory Analysis*

Exploratory Analysis was performed by examining plots, tables and summary statistics on the observed data. A number of transformations were applied to the raw data to make data more useful for analysis, including data typing and scaling (i.e. Activity type was changed to a class of “factor”). Missing values were identified, data quality was assessed, and preliminary analysis performed to determine the terms for use in the prediction model relating various details about the observation to the activity being performed.

*Statistical Modeling*

To create the model that predicts the activity type based on the sensory input from the smartphone, a standard classification tree method was used. [2] The model was created using all available variables in the test data set to build the classification tree. Model fitment was determined through standard means such as Residual Mean Deviance and Misclassification Error Rate.

*Reproducibility*

All analysis performed for this assignment are reproduced using standard R code. To reproduce this analysis exactly, the source file provided by UCI’s machine learning repository must also be acquired.

**Results:**

The data used in this analysis contains information about 7,352 observations of the activities of 30 subjects. In each of these observations, the subject performed one of the following activities: walking, walking upstairs, walking downstairs, sitting, standing, or laying. While the activity was performed, data were collected from the inertia sensors on the smartphone.

**References:**

|  |  |
| --- | --- |
| [1] | D. A. A. G. L. O. Jorge L. Reyes-Ortiz, "Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine," 12 2012. [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones. [Accessed 05 12 2013]. |
| [2] | H. C. B. a. R. A. Irizarry, "Lecture 4: Tree-based Methods," 02 2010. [Online]. Available: http://www.cbcb.umd.edu/~hcorrada/PracticalML/pdf/lectures/trees.pdf. [Accessed 05 12 2013]. |
| [3] | J. Leek, "Sample Analysis Assignment," Coursera.com, 2013. |