Supervised Learning Final Project Description

Problem Definition:

The idea of the project is to implement a human activity recognition(HAR) classifier. According to the training data we defined five activities, including sitting-down, standing-up, standing, walking, sitting. Accelerometer data is collected using 4 sensors on waist, left thigh, right ankle and upper arm.

Application Prospective:

The Human Activity Recognition can be applied to benefit people with special needs such as providing suggestion for exercise or energy consuming. As the Apple Healthkit released, more applications tracking human activities will be growing. In the future, we may be able to recognize more activity based on machine learning method and big data application.

Team Members:

Yudu Du, Yihan Xu, Shuxian Shen

Training & Test data set:

The data set is from Pontifical Catholic University of Rio de Janeiro. It's collected on 8 hours of activities of 4 healthy subjects using 4 sensors on body. We are planing to use 70% data as training data set and rest 30% data as testing data set.(Link)

The attribute of HAR data set is described as following:
user (text)
gender (text)
age (integer)
how_tall_in_meters (real)
weight (int)
body_mass_index (real)
x1 (type int, contains the read value of the axis 'x' of the 1st accelerometer, mounted on waist)
y1 (type int, contains the read value of the axis 'y' of the 1st accelerometer, mounted on waist)
z1 (type int, contains the read value of the axis 'z' of the 1st accelerometer, mounted on waist)
x2 (type int, contains the read value of the axis 'x' of the 2nd accelerometer, mounted on the left thigh)
y2 (type int, contains the read value of the axis 'y' of the 2nd accelerometer, mounted on the left thigh)
z2 (type int, contains the read value of the axis 'z' of the 2nd accelerometer, mounted on the left thigh)
x3 (type int, contains the read value of the axis 'x' of the 3rd accelerometer, mounted on the right ankle)
y3 (type int, contains the read value of the axis 'y' of the 3rd accelerometer, mounted on the right ankle)

z3 (type int, contains the read value of the axis 'z' of the 3rd accelerometer, mounted on the right ankle)
x4 (type int, contains the read value of the axis 'x' of the 4th accelerometer, mounted on the right upper-arm)
y4 (type int, contains the read value of the axis 'y' of the 4th accelerometer, mounted on the right upper-arm)
z4 (type int, contains the read value of the axis 'z' of the 4th accelerometer, mounted on the right upper-arm)

Methods:

We have not decided which specific algorithm to use yet. We are thinking to build a classifier using iterative linear regression algorithm or use decision tree to do this project. The input data is [x1, y1, z1, x2, y2, z2, x3, y3, z3, x4, y4, z4], which are the accelerations of three orthometric directions got from four accelerometers at different sites. The label is class, which include sitting, sitting-down, standing, standing-up, and walking. We plan to use 70% of the data set to do the training and the rest 30% to do the testing. We have a large quantity of data, which is good for training. But we still need to choose model carefully to avoid overtraining caused by improper model complexity. Across validation can be applied to check if we are overtraining. We may applying boosting to deal with the problem depending on the performance, decision tree is also an promising option we are consider right now.

Evaluation:

Since we will use 30% data as testing data set, we collect classifier class result and compare with the real result, and calculate the accuracy of the classifier and compare it other classifier to get the performance of our classifier.