Project Title: Human Activity Recognition using Vanilla Multiclass Logistic Regression and Neural Networks

Arockia Richard Raj Manohar, Simeon Babatunde

Problem Description

Human Activity Recognition is the problem of using sensor data such as accelerometer data to predict human movements and activities. Given the rise of wearable sensors and devices, Human Activity Recognition has become an active area of research over the past few years. It has a wide range of applications such as the fall detection algorithm used in the Apple watch. Applying machine learning to classify human activities has been proven to yield highly accurate results [1]. In this project, we plan to predict human activities from a set of accelerometer data using machine learning. Concretely, we will use multi-class logistic regression and neural networks to predict human activities. Through our project, we aim to answer the following questions:

1. How our multi-class regression algorithm compares against a relatively complex model such as Adaboost which has been shown to produce highly accurate results in activity recognition
2. How our neural network model performs against Adaboost
3. How the three models compare against each other and
4. Is there a considerable improvement in accuracy when using models from well known machine learning libraries such as scikit learn as opposed to implementing from scratch

Data Description

Implementing a multinomial logistic regression for human activity recognition (HAR) requires dataset that contains various attributes that define human activities like **sitting, standing** and **walking.** The dataset for this project would be collected from the UCI machine learning repository. Specifically, this project would be based on the HAR dataset curated by the Groupware Technologies. This dataset contains 165,635 rows of data points and 18 columns consisting of 17 attributes and a class label (sitting, standing or walking). Each line contains semicolon-separated entries of user; gender; age; how_tall_in_meters; weight; body_mass_index; x1; y1; z1; x2; y2; z2; x3; y3; z3; x4; y4; z4; class. The x_i y_i and z_i values represent the accelerometer readings from different body locations i.e. waist, left-thigh, ankle and upper-arm.

Machine Learning Algorithms

We first plan to implement two machine learning models from scratch, without using any machine learning libraries:

1. Multi-Class Logistic Regression
2. Neural Network

Once the two models are implemented, we plan on evaluating the models and comparing them to see if one model performs considerably better than the other.
We then plan to implement the above two algorithms using popular libraries such as scikit-learn to see if they improve the accuracy considerably.

Evaluation

We plan to evaluate each algorithm independently followed by a comparison of the results. This also includes the scikit learn based algorithm. Evaluation will cover the following metrics: accuracy, precision, recall, and F1 score. The results of the evaluation will be presented through tables and confusion matrices.

References

1. Ugulino, W.; Cardador, D.; Vega, K.; Velloso, E.; Milidiu, R.; Fuks, H. Wearable Computing: Accelerometers' Data Classification of Body Postures and Movements. Proceedings of 21st Brazilian Symposium on Artificial Intelligence. Advances in Artificial Intelligence - SBIA 2012. In: Lecture Notes in Computer Science. , pp. 52-61. Curitiba, PR: Springer Berlin / Heidelberg, 2012. ISBN 978-3-642-34458-9. DOI: 10.1007/978-3-642-34459-6_6.