PATTERN RECOGNITION
SEMESTER PROJECT REPORT 1

E. Pruthvik reddy S20170020203
E. Pavan Kumar S20170020205
Shubham raj gupta S20170020238

Daily Sports and Activities

Problem

To train a classifier to predict which activities users are engaging in based on sensor data collected from devices attached to all four limbs and the torso. This will be accomplished by extracting the features from the sensor data and training machine learning classifiers.

Dataset Information

The dataset comprises motion sensor data of 19 daily and sports activities each performed by 8 subjects in their own style for duration of 5 minutes. Five sensors are placed on the torso, arms, and legs.

8 subjects all participated in the same 19 activities. Each of the 5 devices (4 limbs and 1 torso) have 9 sensors (x,y,z accelerometers, x,y,z gyroscopes, and x,y,z magnetometers). The data is collected in 5 second segments with a frequency of 25 Hz for a total of 5 minutes for each activity for each user.

Data Format: CSV

Attribute Information

- sitting (*A1*),
- standing (*A2*),
- lying on back and on the right side (*A3 and A4*),
- ascending and descending stairs *(A5 and A6*),
- standing in an elevator still (*A7*)
- and moving around in an elevator (*A8*),
- walking in a parking lot (*A9*),
- walking on a treadmill at a speed of 4 km/h (in flat and 15 deg inclined positions) *(A1 0 and A11),*
- running on a treadmill at a speed of 8 km/h (*A12*),
- exercising on a stepper (*A13*),
- exercising on a cross trainer (*A14*),
- cycling on an exercise bike in horizontal and vertical positions (*A15 and A16*),
- rowing (*A17*),
- jumping (*A18*),
- and playing basketball (*A19*).

Feature Extraction

The features are the mean, variance, skewness, and kurtosis of each row's distribution and the maximum five peaks of the discrete Fourier transform of a segment with the corresponding frequencies.

We'll normalize each feature to values between [0,1], then flatten each 5 second segment into a single row with 1140 features. Such a large number of features will introduce the Curse of Dimensionality and reduce the performance of most classifiers. So we'll reduce the dimensions by applying Principal Component Analysis (PCA).

Classification algorithms to be compared

I. Logistic regression
II. Decision trees
III. Support Vector Machine (SVM)