DSC 483 Mini-Project

Human Activity Recognition Using Smartphones Dataset

Chunlei Zhou

January 2020

1. Introduction

Human activity recognition (HAR) has wide applications in medical research and human survey system [1]. With the development of real-time sensors, the opportunity to measure and analyze the personalized health-care data has increased.

In this project, we try to recognize human activities based on data from the built-in sensors in waist-mounted smartphones. More specifically, our objective is to build predictive models that accurately classifies six activities of daily living (ADL): Walking, Walking Upstairs, Walking Downstairs, Sitting, Standing, and Laying [2].

Exploratory analysis and predictive modeling techniques are used to identify distinctive ADL. Tow classification models, multinomial logistic regression and multi-layer perceptron classifier are built and compared in Section 4. Dimensionally reduction technique is used to reduce the feature set. Classification model is built using the dimensionally reduced dataset in Section 5. Upon initial experiment, multinomial logistic regression and multi-layer perceptron classifier can achieve overall predictive accuracy of 93.64% and 94.31% respectively.

2. Data Collection and Preparation

2.1 Source of Dataset

The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using the embedded accelerometer and gyroscope, 3-axial linear acceleration and 3-axial angular velocity were captured at a constant rate of 50Hz. The data have been labeled manually based on video-recorded experiments [2].

Both the accelerometer and gyroscope signals were time domain signals (prefix 't' to denote time). Then they were pre-processed by applying noise filters and were sampled in fixed width sliding windows of 2.56 seconds and 50% overlap (128 readings/window). The acceleration signal was then separated into body acceleration and gravity (*tBodyAcc-XYZ and tGravityAcc-XYZ*) using a Butterworth low-pass filter [2].

Subsequently, the body linear acceleration and angular velocity were derived in time to obtain Jerk signals (*tBodyAccJerk-XYZ and tBodyGyroJerk-XYZ*). Also, the magnitude of these threedimensional signals was calculated using the Euclidean norm (*tBodyAccMag*, *tGravityAccMag*, *tBodyAccJerkMag*, *tBodyGyroMag*, *tBodyGyroJerkMag*) [2].

Finally, a Fast Fourier Transform (FFT) was applied to some of these signals producing *fBodyAcc-XYZ, fBodyAccJerk-XYZ, fBodyGyro-XYZ, fBodyAccJerkMag, fBodyGyroMag, fBodyGyroJerkMag*. (prefix 'f' indicates frequency domain signals) [2].

These signals were used to estimate variables of the feature vector for each pattern:
'-XYZ' is used to denote 3-axial signals in the X, Y and Z directions. Moreover, statistical summaries of all features were calculated as additional features to provide a more detailed and complete description of ADL [2].

2.2Data Overview

The final dataset contains 10,299 records obtained from 30 different experiment subjects. Each record is a 561-feature vector. The features are normalized and bounded within [-1, 1].

2.3 Train-Test Split

The data is split into training and testing set roughly based on 80%:20%. To avoid data leakage, the same subject does not appear in both training and testing set. Thus, the subjects 2, 5, 7, 8, 9, 10, 11 were put in the testing set, and other subjects were put in the training set.

3. Exploratory Analysis

To better describe the data set, the exploratory analysis was conducted. Only the mean values of the selected core features (*tBodyAcc-XYZ, tGravityAcc-XYZ, tBodyAccJerk-XYZ, tBodyGyro-XYZ, tBodyGyroJerk-XYZ, tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag, tBodyGyroJerkMag*) were used for this analysis.

Figure 1 is the correlation heatmap of the core features. Colors, with blue representing negative relationship and red representing positive relationship, indicate the correlation (ranging from -1 to +1) between each pair of two features. The darker the color, the higher the correlation.

According to Figure 1, it is obvious that there are high correlations among *tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag, tBodyGyroJerkMag*. Also, some relationships exist between the above mentioned five features and *tBodyAcc-XYZ*. Furthermore, there are correlations between the three dimensions (XYZ) of the same sensor signal (e.g. *tGravityAcc*).

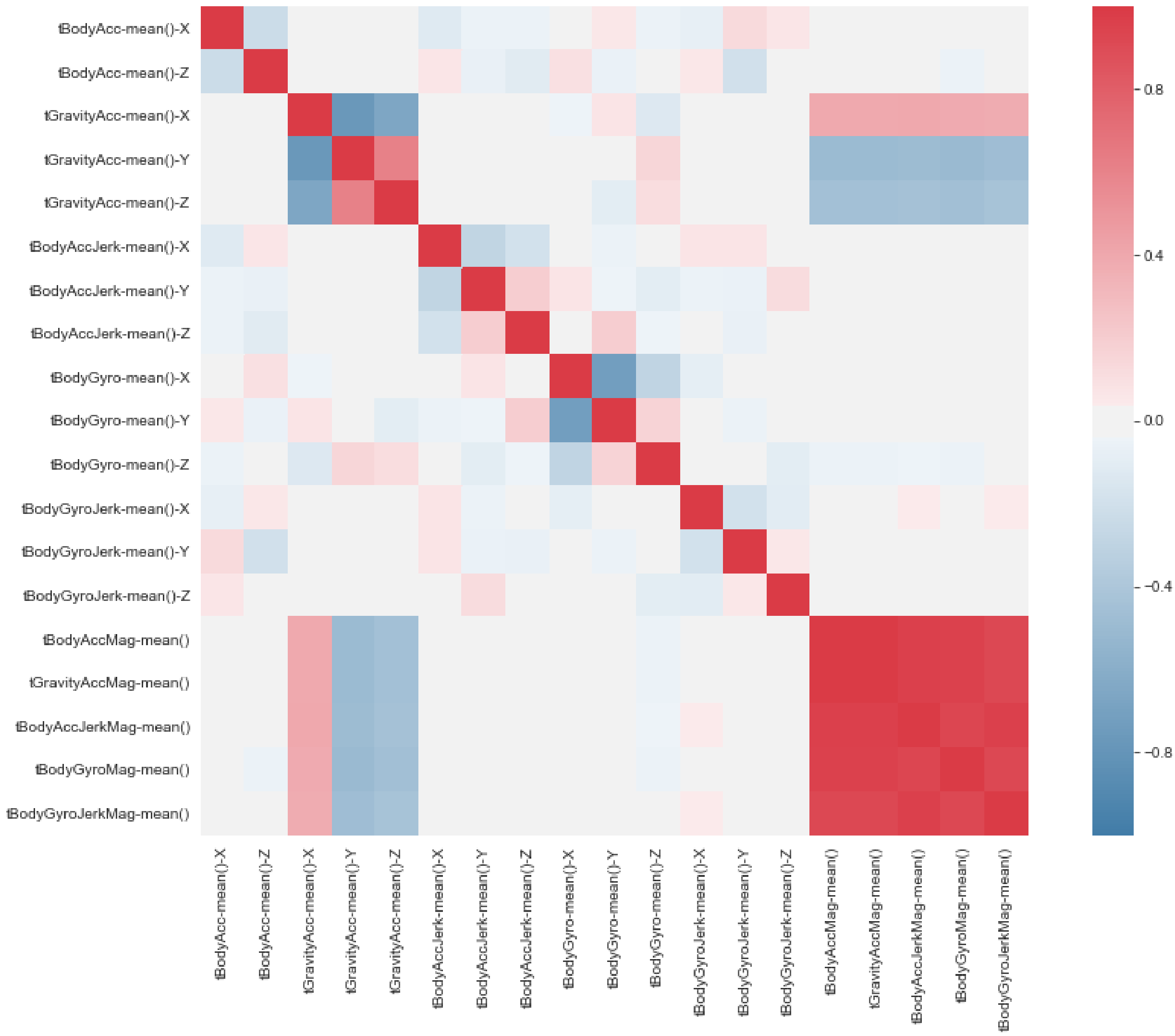

Figure 1. Correlation Heatmap

4. Methods

In this project, two models, multinomial logistic regression and multi-layer perceptron, are chosen and compared.

4.1 Models

4.1.1 Multinomial Logistic Regression

Multinomial logistic regression is a classification method that generalizes logistic regression and multiclass problems [3]. Newton's method was used to estimate the model parameters. Loss function of the model was cross-entropy loss.

4.1.2 Multi-Layer Perceptron

A multilayer perceptron (MLP) is a class of feedforward artificial neural network. It utilizes a supervised learning technique called backpropagation for training and can distinguish non-linear separable data [4]. Hidden layer size of the model is 830 (number of samples:number of neurons was roughly 10:1). To achieve convergence and to avoid overfitting, the number of iterations was chosen to be 1000. The loss function for this classification model was Cross-Entropy.

Cross-Entropy Loss

$$\text{Cross Entropy Loss} = \mathrm{L}(\theta) = -\sum_{i}^{m} y^{(i)} \log\left(h_\theta\left(x^{(i)}\right)\right) + \left(1 - y^{(i)}\right) \log\left(1 - h_\theta\left(x^{(i)}\right)\right)$$

4.2 Feature Selection

The data set has 561 features. To reduce the complexity of the predictive model, only a subset of features was selected.

The purpose was to select the least number of features without sacrificing the model accuracy (predictive accuracy should be 80% or 90%). Thus, features were ranked from the most important one to the least important one.

Two methods were used to calculate the feature importance. One is F-Score, and the other one is P-Value. The higher the F-Score, the more important the feature. The lower the P-Value, the more important the feature.

4.2.1 Dimensionality Reduction

In addition to the simple feature selection, the principal component analysis (PCA) was also applied to reduce the feature dimensionality. The target model performance was 90%.

PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The first component is of largest possible variance and the last component is of least possible variance [5].

Using PCA may create a model with same predictive accuracy but less feature dimensions.

5. Results

5.1 Multinomial Logistic Regression

	precision	recall	f1-score	support
LAYING	1.00	1.00	1.00	371
SITTING	0.88	0.89	0.88	341
STANDING	0.90	0.88	0.89	353
WALKING	0.90	0.99	0.94	384
WALKING_DOWNSTAIRS	0.98	0.98	0.98	305
WALKING_UPSTAIRS	0.97	0.88	0.93	337
accuracy			0.94	2091
macro avg	0.94	0.94	0.94	2091
weighted avg	0.94	0.94	0.94	2091

Figure 2. Classification Result of Multinomial Logistic Regression

Predicted	LAYING	SITTING	STANDING	WALKING	WALKING_DOWNSTAIRS	WALKING_UPSTAIRS
Actual						
LAYING	1	65	113	0	0	9
SITTING	0	8	51	15	13	84
STANDING	0	0	0	45	112	53
WALKING	95	83	12	0	0	0
WALKING_DOWNSTAIRS	89	3	0	62	0	0
WALKING_UPSTAIRS	7	0	0	113	36	9

Figure 3. Confusion Matrix of Multinomial Logistic Regression (Without Normalization)

Predicted	LAYING	SITTING	STANDING	WALKING	WALKING_DOWNSTAIRS	WALKING_UPSTAIRS
Actual						
LAYING	0.005319	0.380117	0.538095	0.000000	0.000000	0.054545
SITTING	0.000000	0.046784	0.242857	0.078947	0.084416	0.509091
STANDING	0.000000	0.000000	0.000000	0.236842	0.727273	0.321212
WALKING	0.505319	0.485380	0.057143	0.000000	0.000000	0.000000
WALKING_DOWNSTAIRS	0.473404	0.017544	0.000000	0.326316	0.000000	0.000000
WALKING_UPSTAIRS	0.037234	0.000000	0.000000	0.594737	0.233766	0.054545

Figure 4. *Confusion Matrix of Multinomial Logistic Regression (Normalized)*

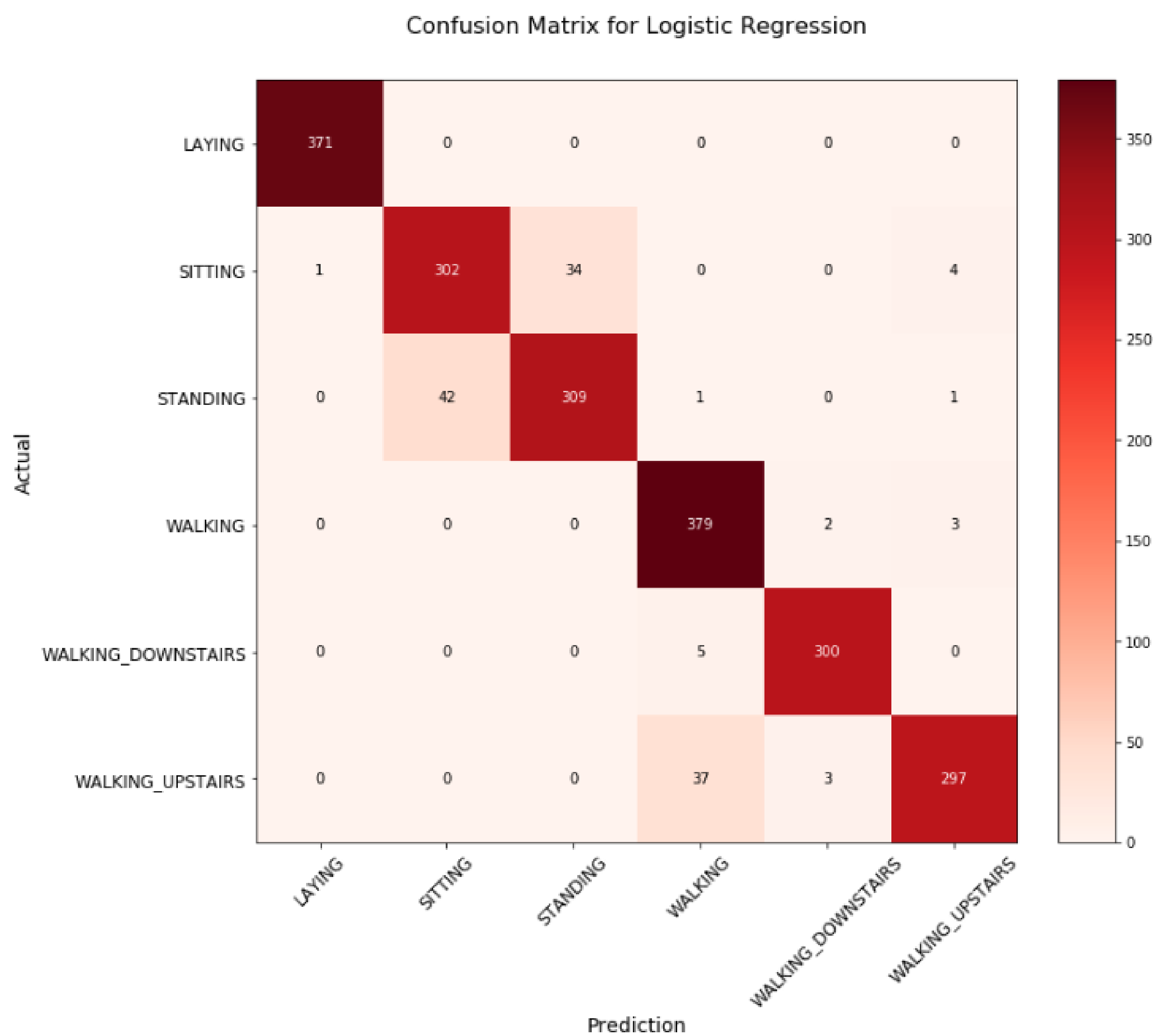

Figure 5. Plot of Confusion Matrix for Multinomial Logistic Regression

Based on Figure 2 – Figure 5 above, multinomial logistic regression performs well on predicting ADL using testing data. The overall accuracy of the model is 93.64% (94% approximately). Regarding to LAYING, the prediction is 100% correct. The prediction of STTING is least accurate with a correctness percentage of 88% only.

The results indicate that the data is linear separable, and thus, the model was chosen for further analysis (dimensionally reduction).

5.1.1 Multinomial Logistic Regression with Reduced Number of Features

As mentioned in section 4.2, a subset of features was selected based on the feature importance, and two scores, F-Score and P-Value, were used respectively to calculate the feature importance.

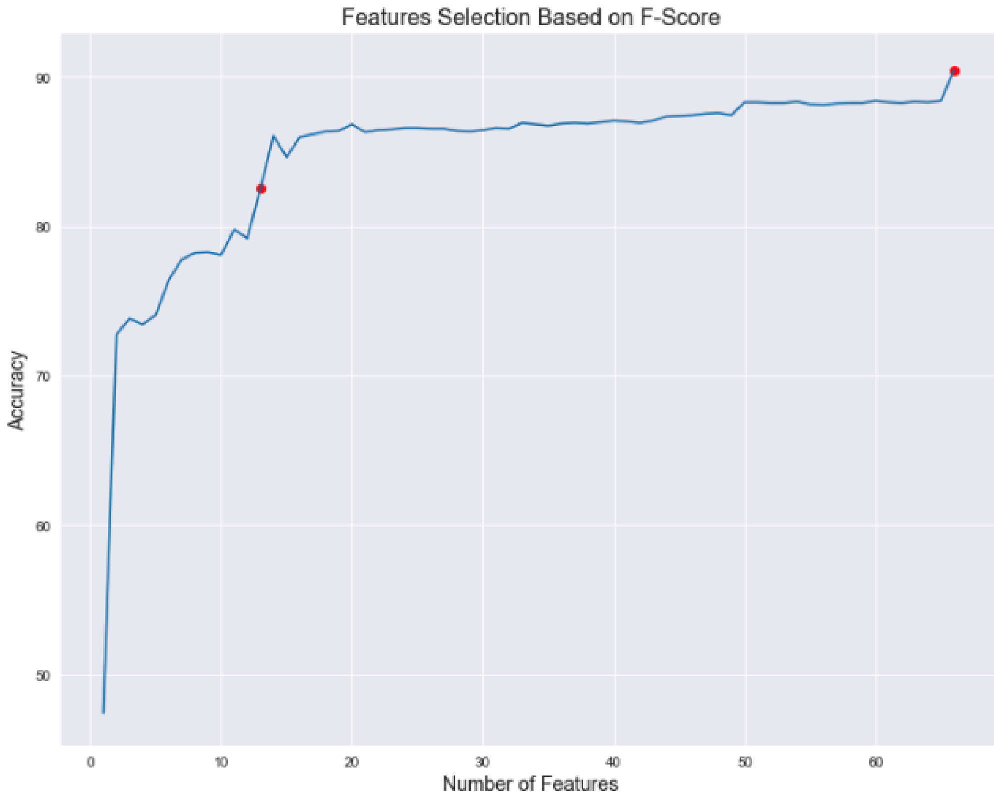

Figure 6. Change of Test Accuracy with the Number of Features Selected Based on F-Score

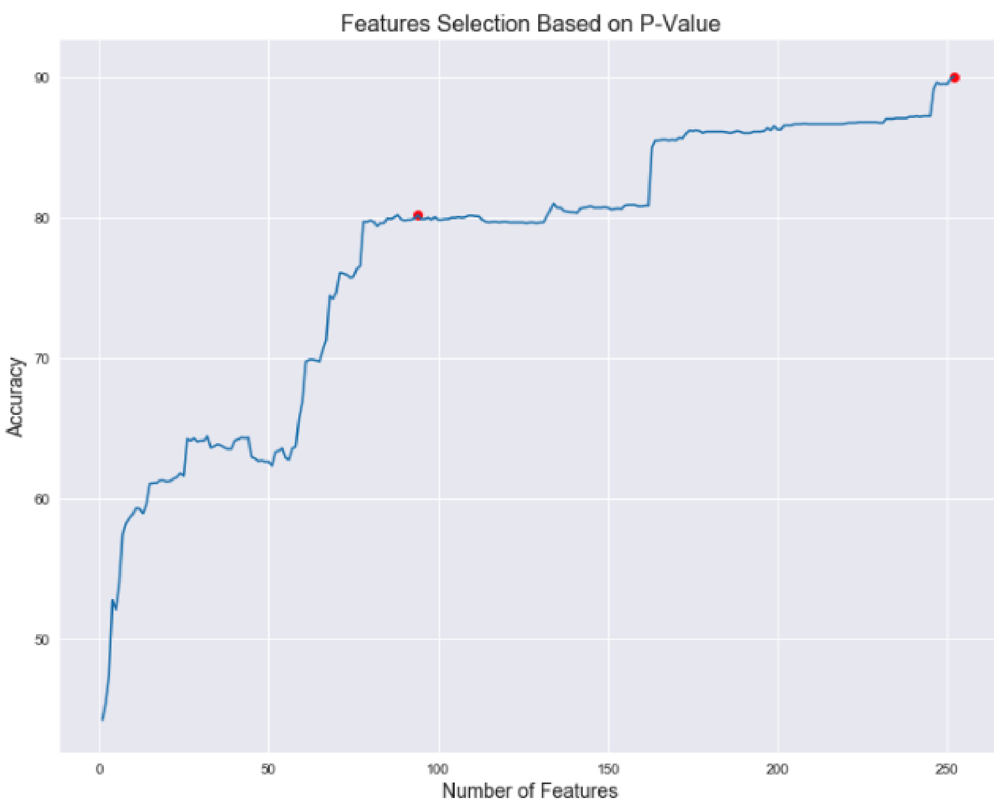

Figure 7. Change of Test Accuracy with the Number of Features Selected Based on P-Value

Using F-Score as the measure of feature importance, to maintain 80% model accuracy, 13 original features should be selected. Using the same measurement, to maintain 90% model accuracy, 66 original features should be selected. (Figure 6)

Using P-Value as the measure of feature importance, to maintain 80% model accuracy, 88 original features should be selected. Using the same measurement, to maintain 90% model accuracy, 252 original features should be selected. (Figure 7)

F-Score is a better measurement for feature importance.

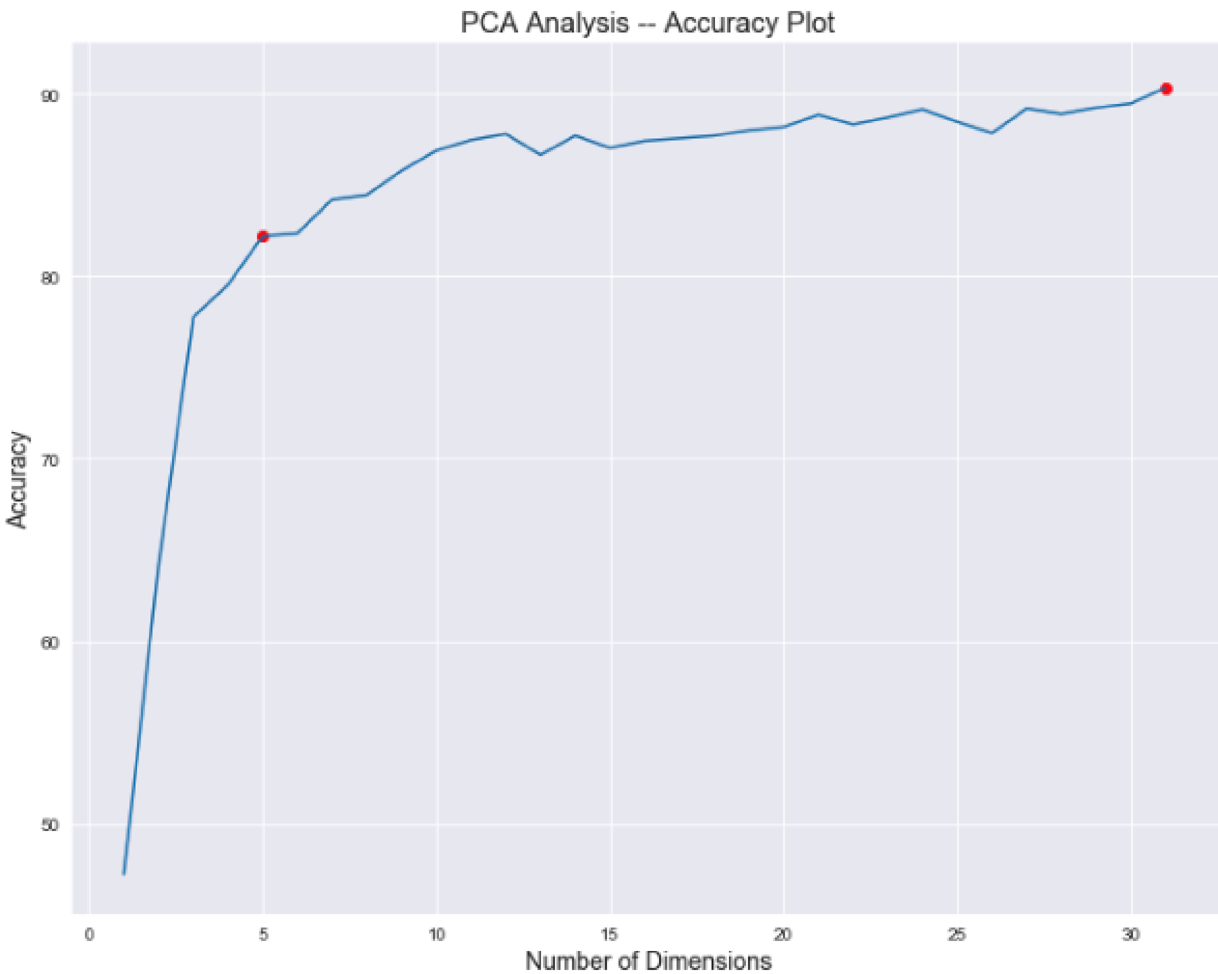

Figure 8. Change of Test Accuracy on multinomial logistic regression with PCA

Based on Figure 8, to achieve the 90% model accuracy, the first 31 dimensions should be selected.

Comparing PCA with simple feature selection, with the same model performance, the number of feature dimension used to train the model, dropping from 66 original features to 31 PCA, was 35 less.

5.2 Multi-Layer Perceptron

	precision	recall	f1-score	support
LAYING	1.00	1.00	1.00	371
SITTING	0.90	0.91	0.91	341
STANDING	0.92	0.90	0.91	353
WALKING	0.92	0.96	0.94	384
WALKING_DOWNSTAIRS	0.94	0.99	0.97	305
WALKING_UPSTAIRS	0.98	0.90	0.94	337
accuracy			0.94	2091
macro avg	0.94	0.94	0.94	2091
weighted avg	0.94	0.94	0.94	2091

Figure 9. Classification Result of MLP

Predicted	LAYING	SITTING	STANDING	WALKING	WALKING_DOWNSTAIRS	WALKING_UPSTAIRS
Actual						
LAYING	1	81	100	1	0	5
SITTING	0	4	56	10	13	88
STANDING	0	0	0	33	115	62
WALKING	95	81	14	0	0	0
WALKING_DOWNSTAIRS	89	3	0	62	0	0
WALKING_UPSTAIRS	7	0	0	110	39	9

Figure 10. Confusion Matrix of MLP (Without Normalization)

Predicted	LAYING	SITTING	STANDING	WALKING	WALKING_DOWNSTAIRS	WALKING_UPSTAIRS
Actual						
LAYING	0.005319	0.473684	0.476190	0.005263	0.000000	0.030303
SITTING	0.000000	0.023392	0.266667	0.052632	0.084416	0.533333
STANDING	0.000000	0.000000	0.000000	0.173684	0.746753	0.375758
WALKING	0.505319	0.473684	0.066667	0.000000	0.000000	0.000000
WALKING_DOWNSTAIRS	0.473404	0.017544	0.000000	0.326316	0.000000	0.000000
WALKING_UPSTAIRS	0.037234	0.000000	0.000000	0.578947	0.253247	0.054545

Figure 11. Confusion Matrix of MLP (Normalized)

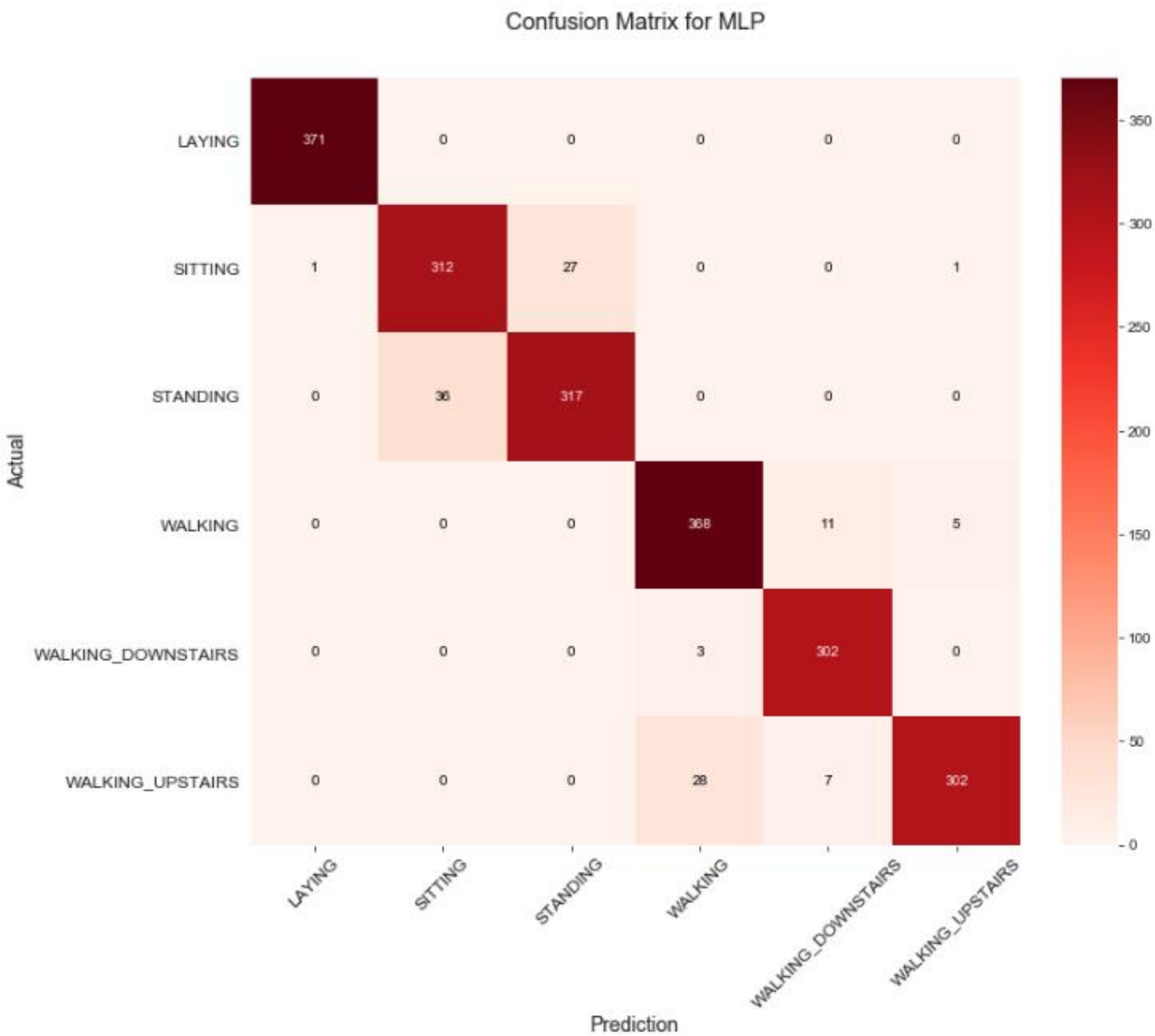

Figure 12. Plot of Confusion Matrix for MLP

Figure 9 – Figure 12 show a descent model performance of Multi-Layer Perceptron on predicting ADL using testing set. The overall accuracy of the model is 94.31% (94% approximately). Similar to multinomial logistic regression, the MLP has the highest predictive accuracy on LAYING, 100%, and the lowest predictive accuracy on STTING, 90% (higher than that of the multinomial logistic regression model).

6. Conclusion

Using the HAR data, we can train classification models that achieve an overall accuracy of about 90% in the prediction of six distinct ADLs. Among all ADLs, LAYING is the one with the highest prediction accuracy. To achieve 90% model accuracy, we can either use the 66 most importance features ranked by F-Score or use the first 31 PCA.

Reference

[1] Rasekh, Amin, Chen, Lu, & Yan. (2014, January 30). Human Activity Recognition using Smartphone. Retrieved January 27, 2020, from https://arxiv.org/abs/1401.8212

[2] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. A Public Domain Dataset for Human Activity Recognition Using Smartphones. 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013. Retrieved January 27, 2020, from https://archive.ics.uci.edu/ml/datasets/Human Activity Recognition Using Smartphones

[3] Logistic regression. (2019, December 31). Retrieved January 27, 2020, from https://en.wikipedia.org/wiki/Logistic_regression

[4] Multilayer perceptron. (2019, October 26). Retrieved January 27, 2020, from https://en.wikipedia.org/wiki/Multilayer_perceptron

[5] Principal component analysis. (2020, January 23). Retrieved January 27, 2020, from https://en.wikipedia.org/wiki/Principal_component_analysis