A Comparative Study of Classification Models for Human Activity Recognition Using Smartphone Data

Name:	**Sushil Vemuri, Anuprita Kulkarni**
Registration No./Roll No.:	19317, 19170
Institute/University Name:	IISER Bhopal
Program/Stream:	Physics
Problem Release date:	January 19, 2022
Date of Submission:	April 24, 2022

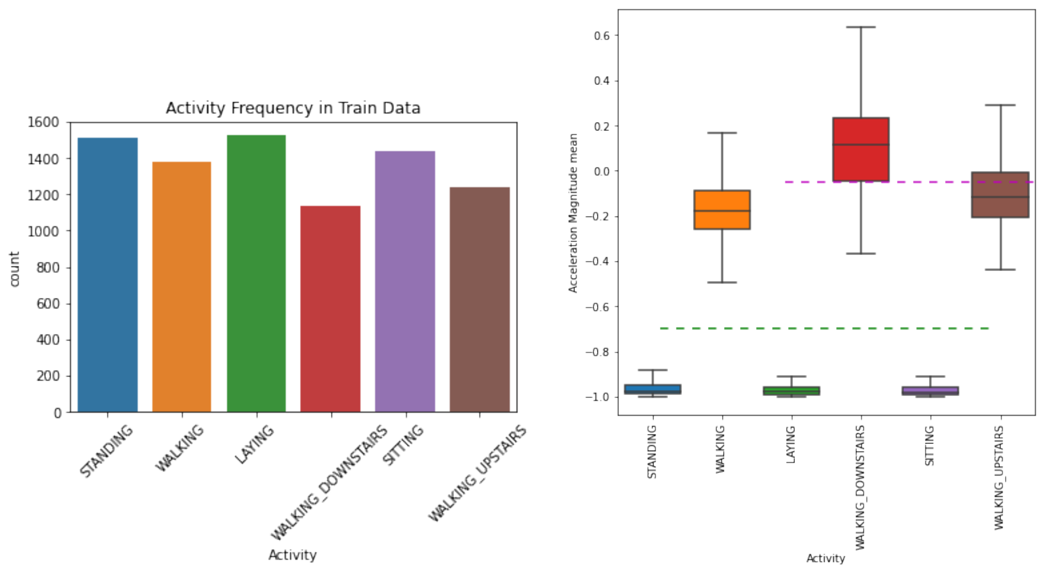
1 Introduction

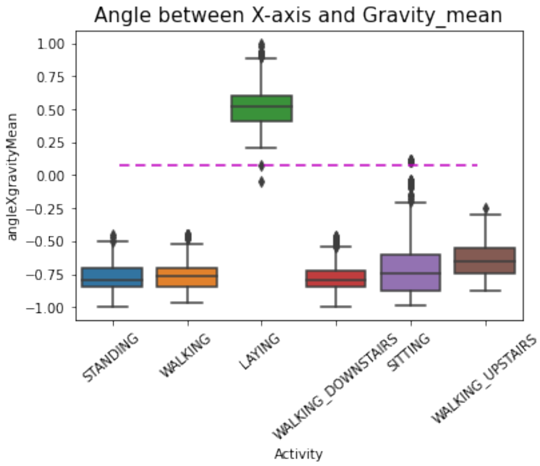
Smartphones have become an integral part of daily life, thanks to the vast assortment of functionality offered through advanced technology. Sensors serve to be a significant aspect thereof, and facilitate the collection of vast amounts of information about the user's daily life and activities. Human Activity Recognition is a problem that focuses on the identification and classification of movements using these sensors for making relevant inferences. This project approaches the problem of HAR from a comparative perspective - the aim is to apply various Machine Learning models (on the UCI dataset) and study and compare the performance of each with respect to the others, and obtain insights regarding the same.

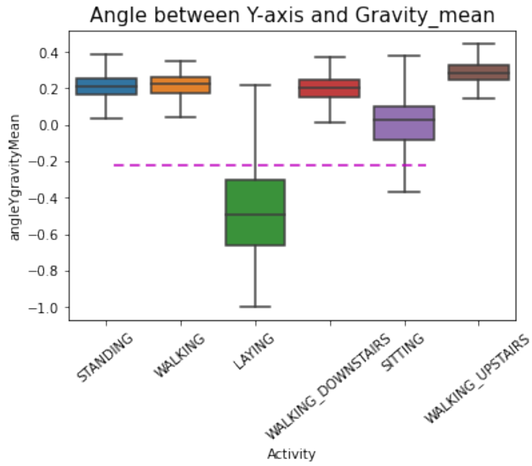
2 Methods

Exploratory Data Analysis

The first step, naturally, is understanding the data. The dataset provided comprises the recordings made by a waist-mounted smartphone which monitored the daily activities of 30 participants. The 6 activities performed are STANDING (1), WALKING (2), LAYING (3), WALKING DOWNSTAIRS (4), SITTING (5), and WALKING UPSTAIRS (6). Following are the attempts at data visualization:







From EDA we observe that the dataset is quite balanced w.r.t activity frequency. Additionally, the feature measuring mean acceleration can easily separate static activities (sitting, standing, laying) from the dynamic ones (walking, walking upstairs/downstairs). The features measuring angle between body axes and gravity vector are helpful in separating the laying activity from others. Thus, we have identified some potentially useful features. Furthermore, it has been observed that accelerometer and gyroscope are the 2 major sensors contributing to the readings, with accelerometer data being slightly more prevalent than gyroscope.

Data Preprocessing

The dataset was checked for null/NaN/duplicate values, and none were found. One-hot encoding is employed to deal with categorical data (such as test subject). The dataset has been already formulated such that all valued are scaled to lie between [-1, 1], so no additional scaling is required.

Feature Engineering

The goal of feature extraction in classifying applications is to reduce the feature space of the original data while still representing the data and class label relationship. The dimensionality of the dataset is very high, with 562 features. Regardless of this, preliminary tests of Machine Learning models produced results of decent accuracy even on raw data. The methodology that was followed in this experiment is given below:

1. Apply selected models on raw data and compare performances based on accuracy and other measures to identify the model that is performing best.

2. Now individually apply various feature engineering techniques such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Pearson Correlation, and ANOVA test, and observe the performance of the model that was selected in step 1.

3. Select the approaches that positively affected the results (in terms of metrics such as accuracy) and finalize a combination of these methods as the feature engineering pipeline to be applied on the test data.

Models Used

Multiple models were implemented on the raw data, such as k-Nearest Neighbors (kNN), Support Vector Machine (SVM), Decision Tree, Random Forest, AdaBoost, and Logistic Regression. The performances of all have been documented. Hyperparameter tuning was performed on each using scikit-learn's GridSearchCV, which applies a grid search to an array of hyperparameters and then cross-validates the model using k-fold cross-validation. Subsequently, confusion matrices were made as well, for all the models implemented on the raw data. The complete notebook can be found in the following GitHub repository:
`https://github.com/TheBlackCat22/Human-Activity-Recognition-using-Smartphone-Data`

3 Observations and Evaluation

The scores from the classification reports of the implementation of models on raw data have been documented in Table 1.

Classifier	Accuracy	Precision	Recall	F-measure
kNN	0.9805	0.98	0.98	0.98
SVM	0.9923	0.99	0.99	0.99
Decision Tree	0.9348	0.93	0.93	0.93
Random Forest	0.9834	0.98	0.98	0.98
AdaBoost	0.9644	0.97	0.97	0.97
Logistic Regression	0.9866	0.99	0.99	0.99

Figure 1: Performances of models on raw data. Note that the precision, recall and f-measure scores have been macro-averaged.

We observe that Support Vector Machine (with parameters 'C': 33.333400000000005, 'degree': 1, 'gamma': 'scale', 'kernel': 'rbf') performed best on our training set, with an accuracy score of 0.9923139158576052. We therefore proceed with the application of dimensionality reduction and feature selection techniques on the data, with SVM (rbf kernel) as the preferred model for evaluation. We now apply two dimensionality reduction methods (PCA, LDA) and three feature selection methods (Pearson Correlation, ANOVA, Recursive Feature Elimination).

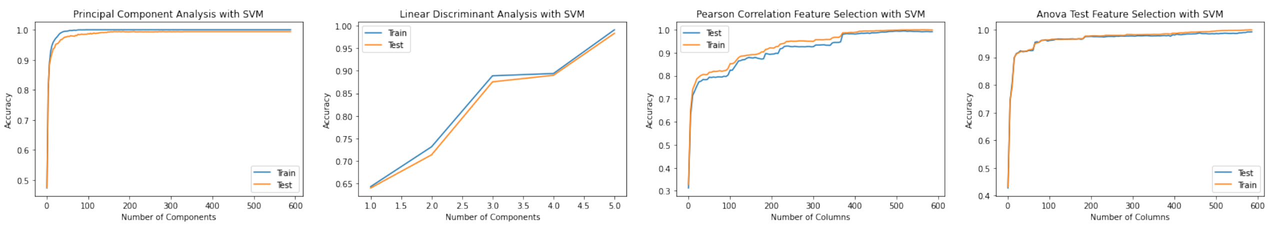


Figure 2: Performance of Feature Engineering Techniques on Train-Test Split

	PCA	LDA	Pearson	ANOVA
Accuracy	0.9943	0.983	0.9943	0.9919
n_components	163	5	516	586

Figure 3: Performance Scores with accuracy and number of components

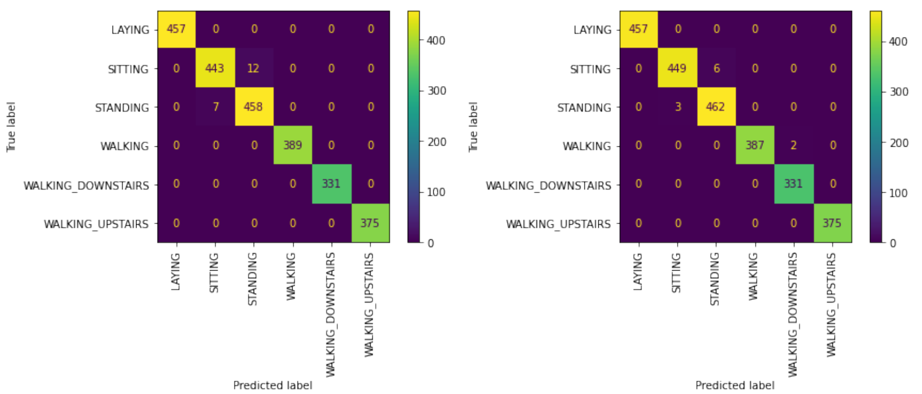


Figure 4: SVM Confusion matrix before (left) and after (right) feature engineering

It is evident that PCA and Pearson Correlation actually helped enhance the accuracy of the model. The performance upgrade is also evident via the confusion matrices before and after feature engineering. An increase in accuracy from 0.9923139158576052 to 0.9955501618122977 is observed, along with precision, recall and f-measure each scoring 1 in the classification report.

4 Analysis

Principal component analysis is an unsupervised dimensionality reduction technique that determines the most accurate data representation in a lower-dimensional space while preserving the largest variance in the data. On the other hand, Linear discriminant analysis performs dimensionality reduction while preserving the largest variance in the class information. LDA aims to maximize the distance between the projected class means while keeping small variance within each class.
Overfitting can occur when an algorithm puts way too much emphasis on maximizing inter-class discrimination in a highly dimensional dataset. Therefore, LDA may have good performance on a low-dimensional dataset, but it is likely to overfit in the case of high dimensionality such as ours. This can explain why LDA performed poorly as compared to PCA. Furthermore, LDA is constrained by the maximum number of columns it can consider (no. of classes - 1).

One of the main issues we faced with this dataset is the very large feature space, and the high variance that comes along with it. Support Vector Machine therefore turns out to be an effective algorithm for this problem, thanks to the regularization penalty which helps combat higher variances.

One challenge faced by all the models we implemented was distinguishing between SITTING and STANDING activities. This can be because both are static activities in which the subject is stationary, which is why the redings from the sensors may not be very well-separable. A potential solution can be to incorporate additional features in the dataset which can differentiate between the two activities more effectively.

5 Conclusion

The peak performance was exhibited by SVM with rbf kernel. On applying feature selection based on Pearson Correlation, the columns were reduced from the initial 590 (562 + one hot encoded subjects and labels), to 516. On further applying PCA on the resultant set, we further reduced the features from 516 to a final total of 157 principal components.
Following this dimensionality reduction and feature selection, there was an increase in the accuracy of the model, indicating that the original dataset had a lot of features with high mutual correlation.
Through this project, we have therefore explored a variety of classification techniques and studied the effects of feature engineering methods. Human activity recognition using smartphone can have a plethora of applications such as fitness trackers, sleep detection, etc.

References

1. Overfitting in linear feature extraction for classification of high-dimensional image data `https://www.sciencedirect.com/science/article/abs/pii/S0031320315004355`

2. Human Activity Recognition using Smartphone Sensors `http://cs229.stanford.edu/proj2016/report/LingMoore-HumanActivityRecognitionUsingSmartphoneSensors-report.pdf`

3. What Does Your Smartphone Know About You? (EDA) `https://www.kaggle.com/code/morrisb/what-does-your-smartphone-know-about-you`

4. Everything you need to know about Linear Discriminant Analysis `https://www.digitalvidya.com/blog/linear-discriminant-analysis/`