Human Activity Recognition Using Smartphones

João Roque, Ricardo Margarido

April 21, 2017

1 Introduction

The need for intelligent assistance systems is growing, to better monitor several disorders and situations such as: being physically impaired, personal fitness, elderly population etc. . In this project assignment it was proposed to us that we built a classifier capable of distinguishing classes in two scenarios: Scenario A - discriminating if a given person is walking or not, and Scenario B - differentiation among six classes (walking, walking upstairs, walking downstairs, sitting, standing and laying). The data provided was gathered from 30 subjects performing activities of daily living , while carrying a smartphone with embedded inertial sensors.
In this first project milestone, we deliver the following aspects of the work:

- Data Preprocessing - Feature reduction with PCA and LDA, feature selection using the Kruskal-Wallis test
- Minimum Distance and Fisher LDA classifiers
- Results comparison between data with and without preprocessing
- Code

To perform this work we used Python 3.5 and Matlab.

2 Methods

2.1 PCA and LDA

Both the PCA and LDA are feature reduction methods that allow us to reduce our 561 features to just 2 or 3. The main difference between both this methods is that while the LDA tries to model the difference between classes within the data (using it's labels), the PCA treats the entire data set as a whole, and does not make use of it's labels.
To better analyze which transformation translates into a better class division, we performed a PCA to reduce the data set dimension to both 2D and 3D, and a LDA to 2D.

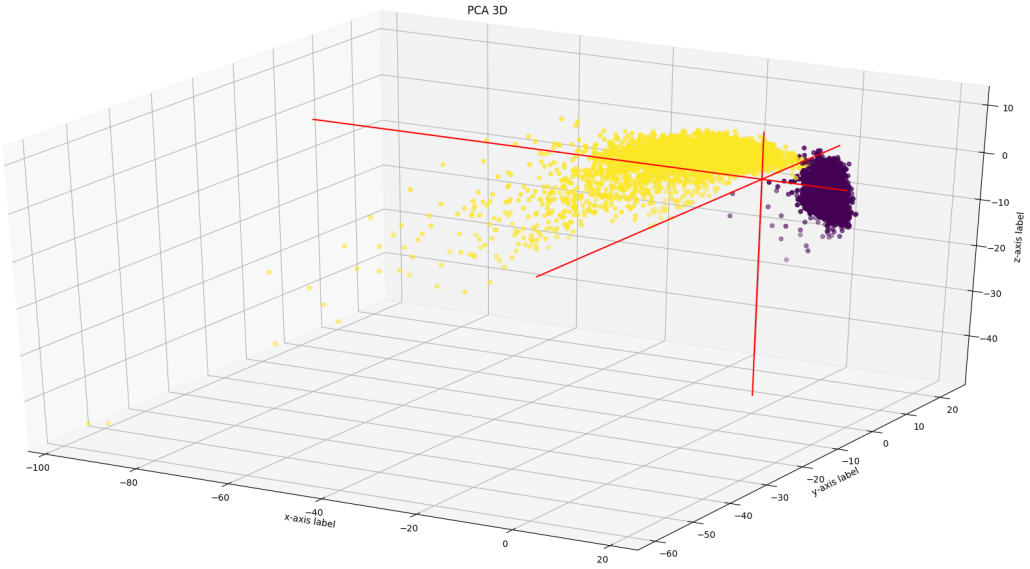

Figure 1: 3D PCA for binary problem

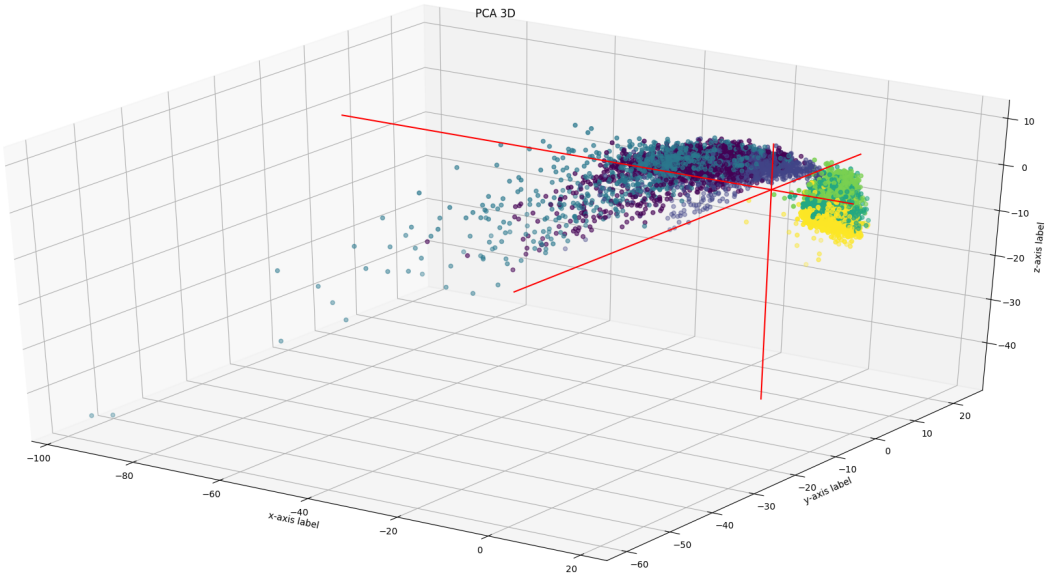

Figure 2: 3D PCA for multi class problem

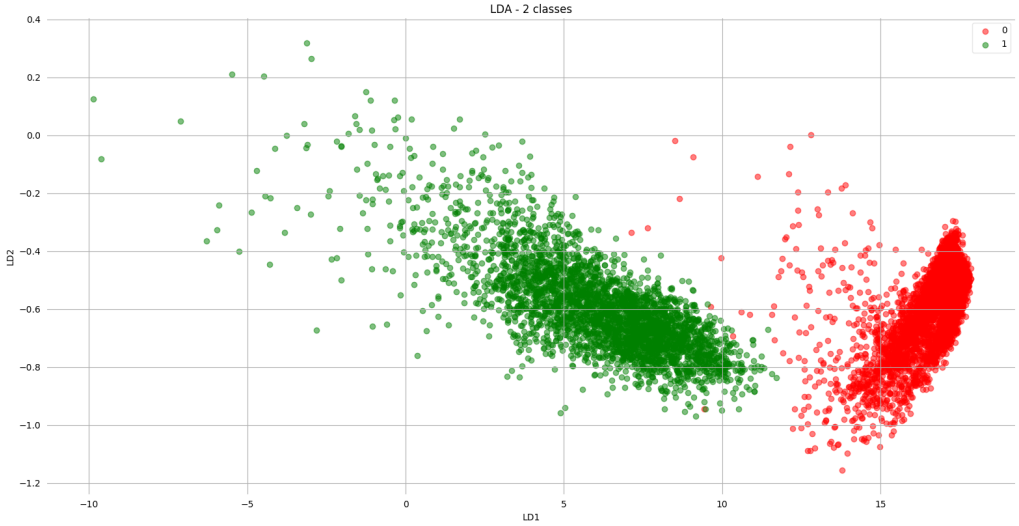

Figure 3: 2D LDA for binary problem

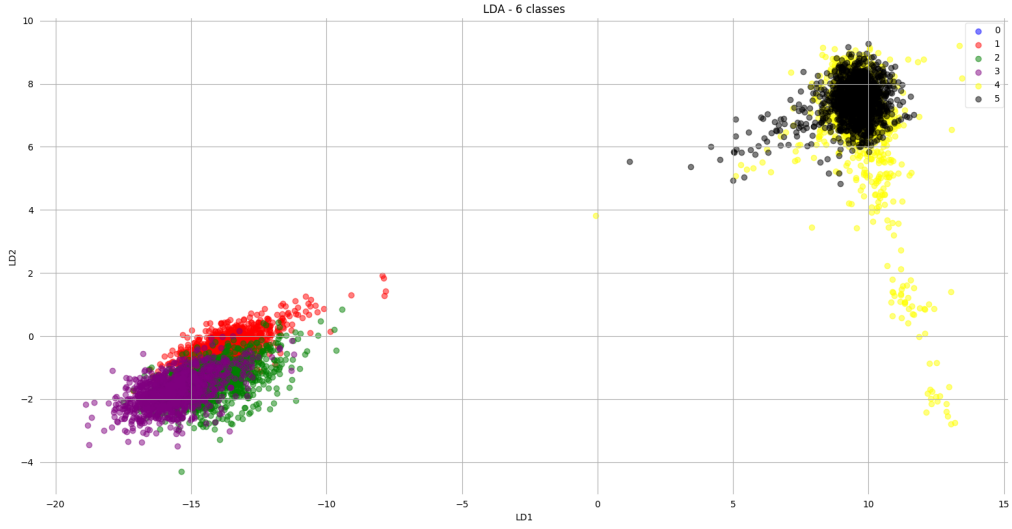

Figure 4: 2D LDA for multi class problem

To visualize the methods, we decided to demonstrate the 2D plots for the LDA and the 3D plots for the PCA, since the classification with the different dimensions is the most important part, and not the visualization itself. As it was expected, with both methods, the difference between the 'Walking' and 'not walking' classes is very clear. However, in the multi class scenario, we think that the LDA performed a better job in separating the important features.

2.2 Feature Reduction with Kruskal-Wallis and Scree Test

Our data set is composed by 561 features. We know for a fact, that some are much more statistically significant to differentiate between the classes in our scenarios than others. For that reason, we wanted to make sure to only use features that positively influence our classifier, and at the same time, reduce it's computing time. For that purpose, we used two different methods called Kruskal-Wallis test and Scree Test.
The Kruskal-Wallis method aims to test if a set of samples originate from the same distribution. In the process it uses each sample chi-squared value which allow us to choose the top 20 features that are more correlated, and we hope that contribute more to separate the classes. Since with python Kruskall-Wallis we were not being able to retrieve the chi-squared values, we performed this test with Matlab.
The Scree Test, is a method performed on the PCA that computes the eigenvalues, and also allows us to visually assess which features explain most of the variability in the data. From these methods, we were able to extract the most relevant features that we used on our classifiers.

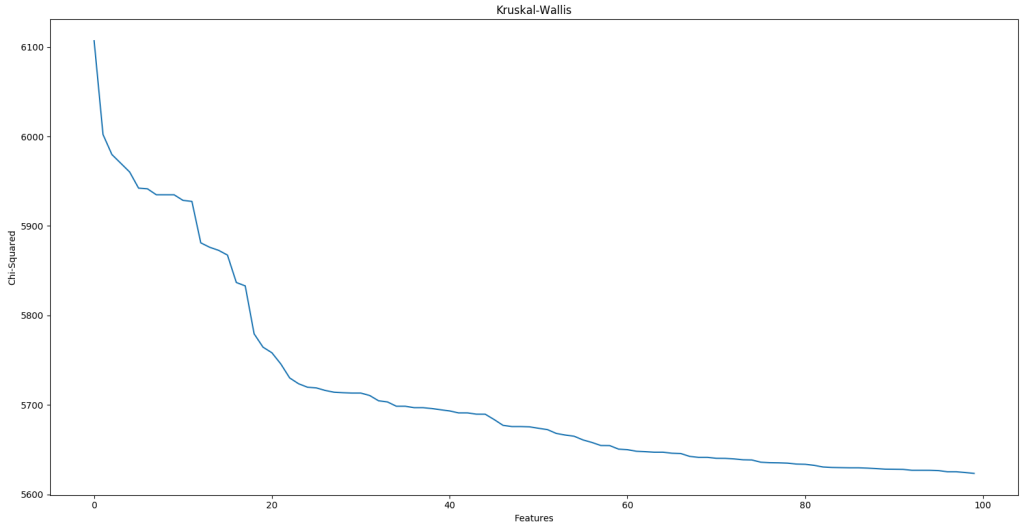

Figure 5: Plot of the Kruskall-Wallis chi-squared values for the most relevant features

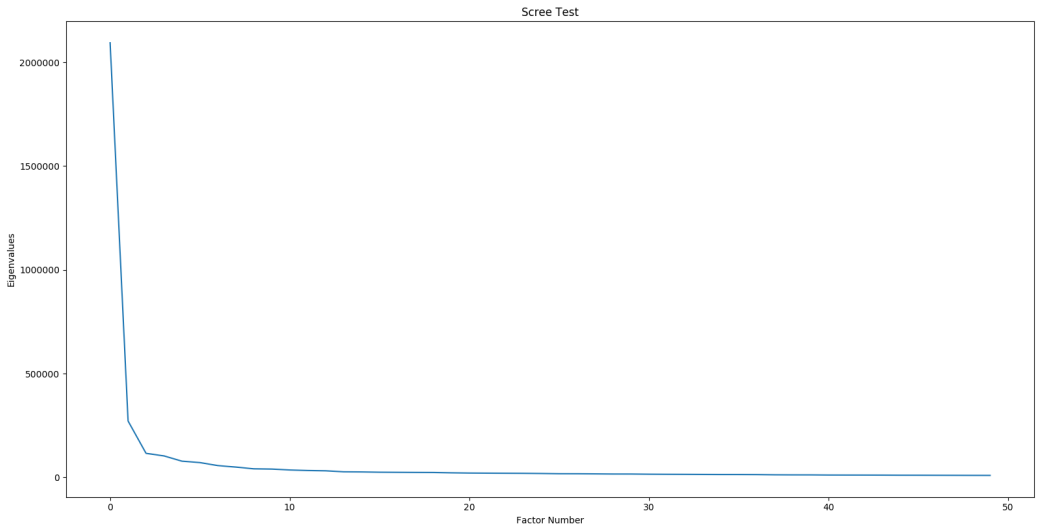

Figure 6: Scree test

2.3 Minimum Distance Classifier and Fisher LDA classifier

In this first assignment, we decided to only start classifying the binary problem, since it's more likely to yield better results. To start building our classifiers, we decided to apply for both the minimum distance and Fisher classifiers the following preprocessed datasets: base data set, PCA with Scree Test data set, LDA data set and Kruskal-Wallis data set.
The Minimum distance classifier simply computes the distance for each new sample to the mean of each class samples. The class chosen, is the closest to the new sample point.
The Fisher LDA classifier is also a very simple method that attempts to find a linear combination of features that characterizes or separates two or more classes of objects or events. Since our data seems very good to classify the binary problem, it is okay to use such simple classifiers for a first attempted, because we know beforehand that they will have positive results.

3 Results

Below, we can seen the confusion matrices for the binary problem, using the Minimum distance classifier created with the data sets discussed above, and the Fisher LDA classifier created with all the data set. These confusion matrices are the result of the predicted values for the given test data set. Through their observation, it is possible to see that as expected, their accuracy, sensibility and specificity values are very high. The best classifiers were the Minimum distance classifier built with preprocessed LDA data, and the Fisher LDA with all data. Since their results were already so good with such simple classifiers, we decided not to build the remaining Fisher LDA classifiers with the preprocessed data, as it would end up being redundant.

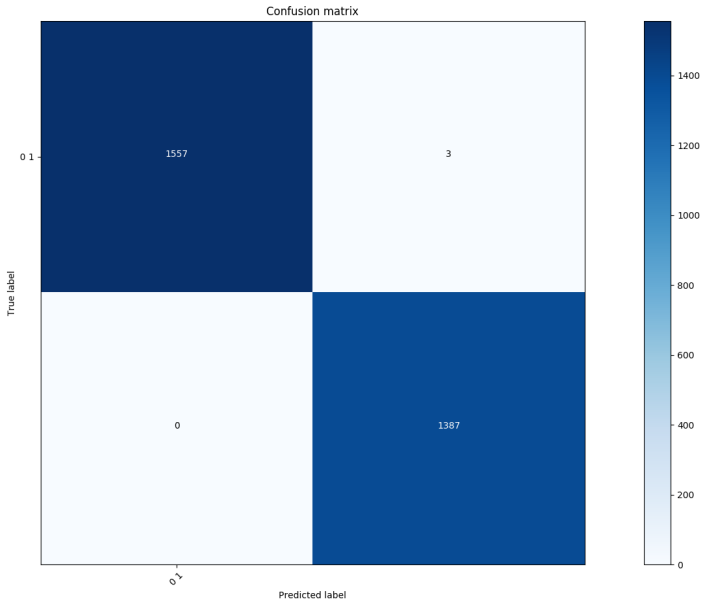

Figure 7: Confusion Matrix for all data with Minimum distance classifier.

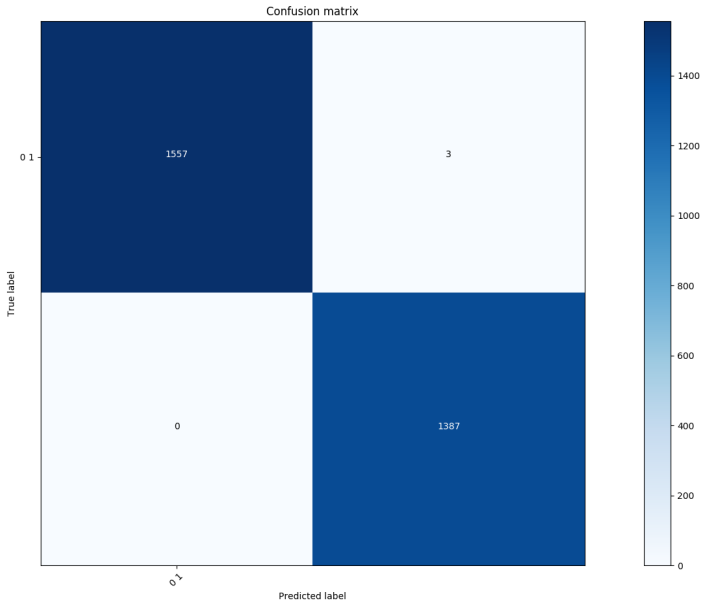

Figure 8: Confusion Matrix with PCA preprocessed data Minimum distance classifier.

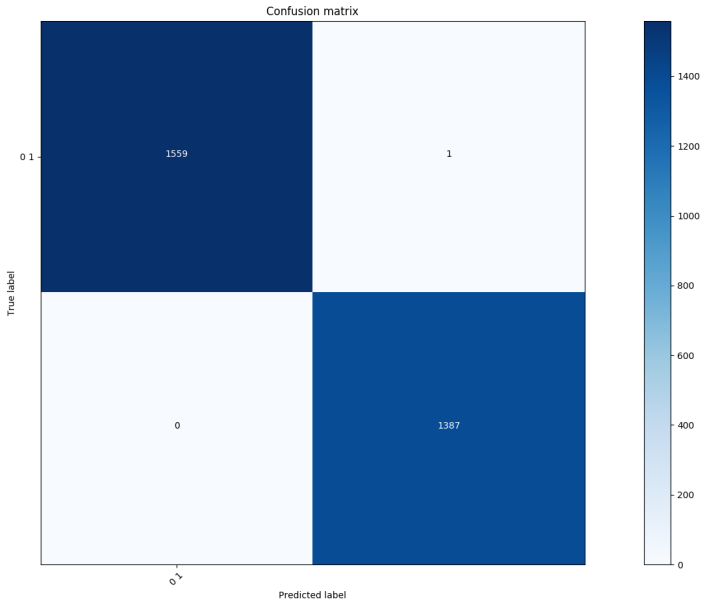

Figure 9: Confusion Matrix with LDA preprocessed data Minimum distance classifier.

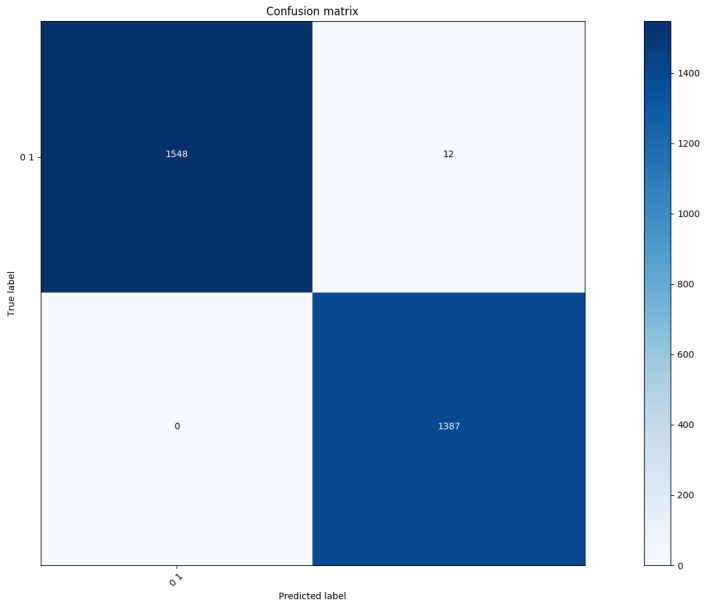

Figure 10: Confusion Matrix with Kruskall-Wallis preprocessed data Minimum distance classifier.

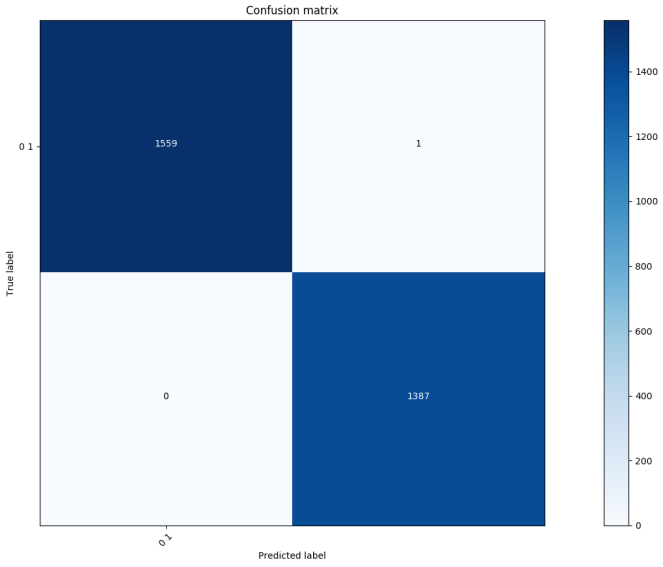

Figure 11: Confusion Matrix for all data with Fisher LDA classifier.

4 Discussion

As we had previously discussed, the binary problem was simple and solvable, because the given data was very easily divided in two separate clusters. Through multiple classifiers, we observed that the LDA dimension reduction and Fisher classifier yielded the best results, with all statistical parameters close to 100% . Tables with statistical parameters for each classifier were not produced since from the confusion matrices, it is very clear that all their results are similar and close to 100% correct classification. In the next reports, we will likely use LDA to approach the multi class scenario, and better classifiers such as neural networks and random forests, in hope to achieve the best results possible for this problem.