Combining PCA and LDA for Facial Recognition

Brendan O'Connor
Department of Computer Science and Electrical Engineering,
University of Missouri-Kansas City,
Kansas City, MO, USA
bmo7x9@mail.umkc.edu

I Brendan O'Connor am the sole author of this paper all external resources have been referenced accordingly.

Abstract

In this experiment we seek to improve the classification of PCA and LDA by combining them. Our hope is to achieve further performance improvements for applications for which performance of facial recognition is a determining factor. We utilized averaging, minimum, and maximization rules for combination in order to achieve improved results. We also attempted to fuse multiple instances of data via averaging which lead to further improvements of the respective classifiers. Our results indicate that the best approach for combining both instances and classifiers is by averaging. Such approaches helped offset the deficiencies of the individual classifiers and instances.

Keywords—Machine Learning, Supervised Learning, PCA, LDA, Facial Recognition, Classification, Computer Vision, Biometrics

I. Introduction

In the current state of computing we have mastered systems for capturing and processing data. However, much of the software systems we currently have are brittle and don't perform well unless under well defined circumstances laid out by the programmers. Additionally, our devices are starting to become the basis for storing personal information and performing high security sensitive task such as financial transactions.

Where these two problems meet is expanding the capabilities of our computers to accurately identify the proper users of the device under a varying degree of circumstance. The result has been the field of biometrics, or using physical traits of the user to authenticate their identity. In this paper we seek to use facial image data as a means to authenticate the identity of various subjects.

This is primarily a classification problem where each unique subject is a class. We attempt to assign a class (identity) to each image that the system is presented with. It is important to note that a critical aspect of this system is being able to establish the class of each sample with a high degree of confidence as this system needs to perform its job in performance critical applications (see above).

For our experiment we attempt to combine existing and well research classification methods (PCA and LDA) in an attempt to achieve further performance improvements on facial recognition and classification.

II. Methods

A. PCA

The first classification scheme that we used was principal component analysis. This approach while unsupervised has resulted in fairly accurate predictions, *Figure 2*. From a high level though, which will be discussed later, a benefit of using PCA is its ability to contribute information when in combination with other classifiers which is the approach we took. In this regard, the individual performance of PCA is irrelevant.

To perform PCA we initially mean normalized our train data[1]. This was done as a means of centering the data around the axis. From their we constructed a covariance matrix of the data. Taking the eigenvector of this matrix results in eigenvalues that when sorted descending provide us with which vectors contribute the most information for creating a new basis for evaluating our data[2]. We use this newly create space to differentiate between each data point while reducing the dimensionality of our data. We utilize this space by projecting first our our mean normalized training set and then later our mean normalized test set. This classifier was used in combination with LDA to produce a high accuracy classification system.

B. LDA

Linear Discriminant Analysis does not differ substantially from the PCA approach discussed above. Building on the approach of PCA, LDA utilizes information of the specific classes of the data to create a supervised approach to classification. This approach is strong in clustering classes together while also separating distances across classes, *Figure 2*. This approach also represents a strong scheme alone for classification of data. Used in conjunction with PCA and combination rules, as we will see in the results section, seeks only to enhance the performance of the classification scheme, see results.

To perform LDA we separated the images by class and took the mean for each class[3]. These means were used to mean normalize each class and then create a covariance matrix for the scatter within each class. After computing the covariance matrix for each class we combine them all via matrix addition to get the within scatter matrix of all the classes[4]. Next we calculate the scatter between the classes. To do this we calculate the mean of the entire data set by averaging each class mean[5]. We take this value and for each class subtract out the respective mean and create a covariance matrix with the resulting matrix [6]. In the same way we performed with the within scatter matrix, we add all of the between scatter matrices together to get the the between scatter matrix for the entire data set [7]. Once we have both of these scatter matrices we multiply the inverse of the within class scatter matrix by the between class scatter matrix (each for the entire data set) [8]. These are the two matrices we want to optimize in conjunction with the goals of LDA as stated above. In the same way the we took the eigenvector for PCA we do the same with LDA. After sorting the components descending and taking the optimal subset of components (for the purposes of this experiment we took ~90% of the proportion of variance)[9].

From there we normalized our test set by subtracting out the mean of the entire data set and projecting it into this newly created space[10]. For the testing data we similarly mean normalized on the mean of the test data and projected it into the newly created space. We use this space in combination with the nearest training data instance in this space to classify our data. This resulted in a strong classification system as seen in *Figure 2*. However, we do observe this classification scheme could and should be improved which motivates our next section on combining PCA and LDA.

C. Classifier Combination

In attempt to improve on already proven classification schemes we attempt to combine the classifiers in order to achieve the highest results possible and to increase the reliability of our classification. For this experiment we focused solely on score level fusion or combining the classifiers after they have assigned a score to each data instance as to whether it is the same person or an imposter. The three different schemes we used were averaging, minimum, and maximum. After we get the scores for PCA and LDA for the same data measurement we take the minimum, maximum, or average of the two values depending on the scheme. The results are discussed below.

D. Multi-instance Fusion

In another attempt to improve the accuracy of the system we attempted to use combine multiple data instances from the same subject and then scored it. For our experiment we simply averaged five of the data instances to fuse them. The results for this are discussed below.

III. RESULTS AND DISCUSSION

As mentioned above and seen in *Figure 2* both LDA and PCA represent strong classifiers in their own right, LDA more so than PCA. However, since we want to ideally reach a perfect classification scheme we attempt to used combination and fusion to improve the accuracy of the classifier.

In observing the varying accuracies for each combination rule we see that the results are a bit mixed, *Figure 1*. The minimum rule, to begin with, performed the worst *Figure 3*. Although it is interesting to note that it did see some performance improvement when the threshold is sufficiently high. However, when the threshold for acceptance is decreased we see some volatility in the accuracy that we don't see in the other schemes. We venture this is due to scores taken from the PCA scheme that on average results in less stable classification. Because of this in the regions where LDA wasn't chosen our resulting ROC curve resembles the PCA curve more than the LDA curve.

Moving on we see that the max combination rule results in a nearly whole sale reproduction of the LDA curve, *Figure 2 & 4*. This is likely due to LDA having higher scores/accuracy than PCA. Because of this it seems the combination rule heavily favored the selection of the LDA score over the PCA score. This is likely what resulted in having such similar scores with LDA. While the LDA results are strong for the purposes of this experiment we were attempting to improve on them so from this perspective the max and the min rules don't present strong improvements over previous approaches.

If we move on to observe the average rule, however, we see our best results *Figure 5*. It seems that this approach was best at achieving at the highest threshold level the performance improvements from PCA while also offsetting some of the performance decreases discussed when using the minimum rule. We note that this curve is nearly ideal. It is also important to note that our system didn't cover a large data set and as such further research should be done to see how these approach scale on larger data sets. It is possible that averaging proves to be a less effective scheme when not used on small data.

Finally we a look at our multi-instance fusion in *Figure 6*. The results were a noticeable improvement on the original schemes. Although we did see the most improvement from the PCA classification, largely because it had the most gains to realize. In comparison with the averaging rule it performed almost as well when looking at the multi-instance fusion for LDA. It could be worth exploring using multi-instance fusion in combination with an average combination rule to see what results could be achieved. On a smaller data set similar to the one that was used in this scenario in likely would not yield much improvements as the results are already close to ideal. However, if we were to expand the scope of this experiment to large sets of data there is a potential that combining classifiers and multiple instances could achieve strong results at a larger scale.

IV. CONCLUSION

In closing we see that we achieved what we sought out to at the onset of this experiment, which was to see if we could improved the classification performance of well known and tested schemes by combining them into a single scheme.

What was discovered in the process is that the different combination rules are far from equal in their performance. It seems that averaging produced the best results on the data this experiment was performed on. This is likely because it is able to combine information from both systems simultaneously rather than using a rule to pick an either or approach of which will work better.

In the either or approach of max and min there was clear regions where you could tell which classifier was used, whereas in the averaging scheme it seemed each classifier offset the other. The other insight achieved was that multi-instance fusion did seem a means of improving each respective classifier but work still remains to determine if these results would benefit the classifiers when combined.

References

[1] Ethem Alpaydin, Introduction to Machine Learning, 3rd edition, MIT press, 2014, section 6.3, page 121.

[2] Ethem Alpaydin, Introduction to Machine Learning, 3rd edition, MIT press, 2014, section 6.3, page 121.

[3] Ethem Alpaydin, Introduction to Machine Learning, 3rd edition, MIT press, 2014, section 6.8, page 140.

[4] Ethem Alpaydin, Introduction to Machine Learning, 3rd edition, MIT press, 2014, section 6.8, page 141.

[5] Ethem Alpaydin, Introduction to Machine Learning, 3rd edition, MIT press, 2014, section 6.8, page 141.

[6] Ethem Alpaydin, Introduction to Machine Learning, 3rd edition, MIT press, 2014, section 6.8, page 141.

[7] Ethem Alpaydin, Introduction to Machine Learning, 3rd edition, MIT press, 2014, section 6.8, page 143.

[8] Ethem Alpaydin, Introduction to Machine Learning, 3rd edition, MIT press, 2014, section 6.8, page 145.

[9] Ethem Alpaydin, Introduction to Machine Learning, 3rd edition, MIT press, 2014, section 6.8, page 145.

[10] Ethem Alpaydin, Introduction to Machine Learning, 3rd edition, MIT press, 2014, section 6.3, page 124.

APPENDIX

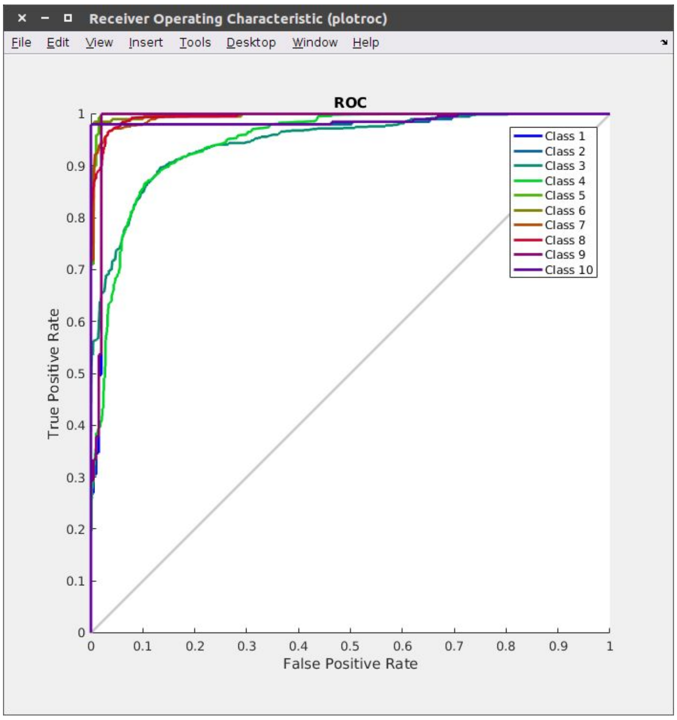

Figure 1. Graph of LDA, PCA, Average rule, Minimum rule, and Maximum rule, respectively

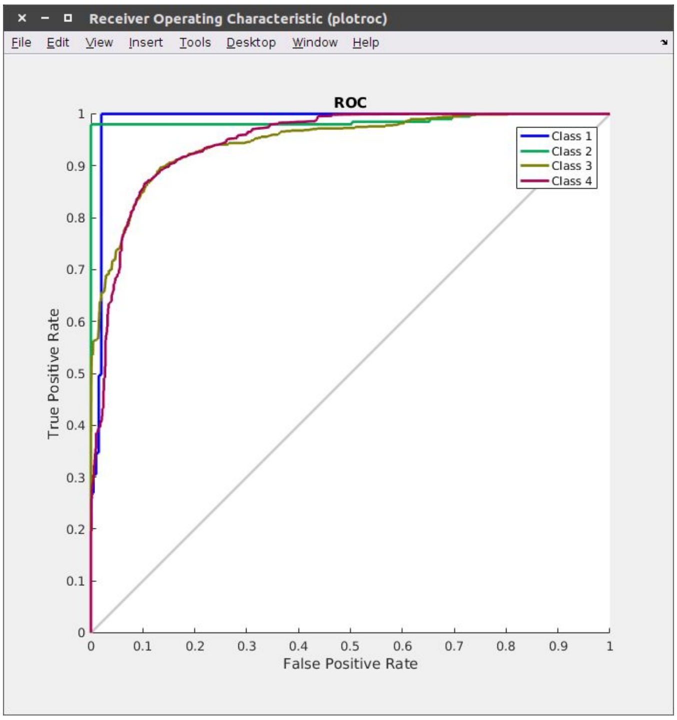

Figure 2. LDA and PCA curves, respectively

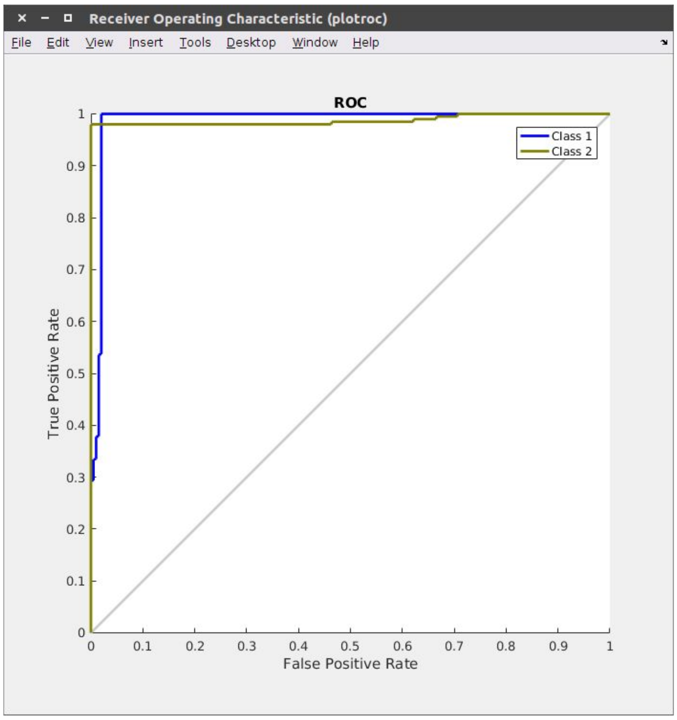

Figure 4. Maximum Combination Rule

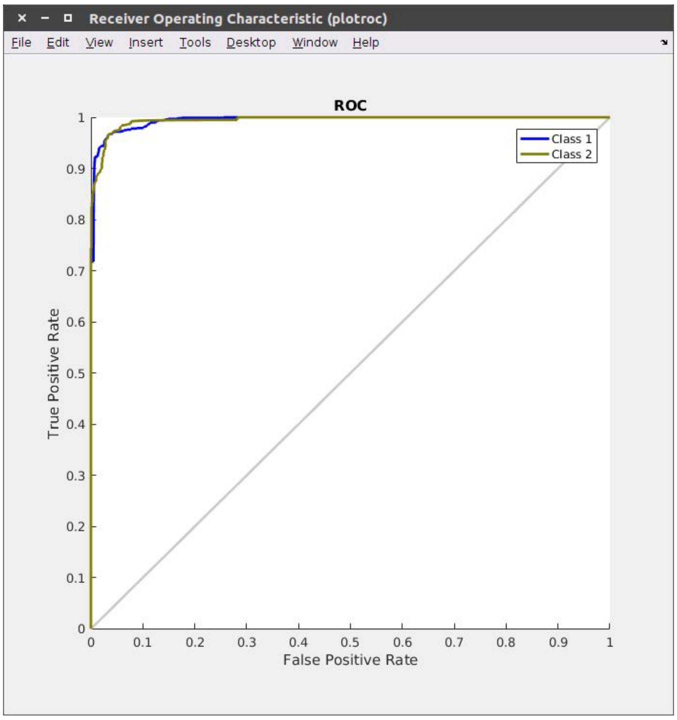

Figure 3. Minimum Combination Rule

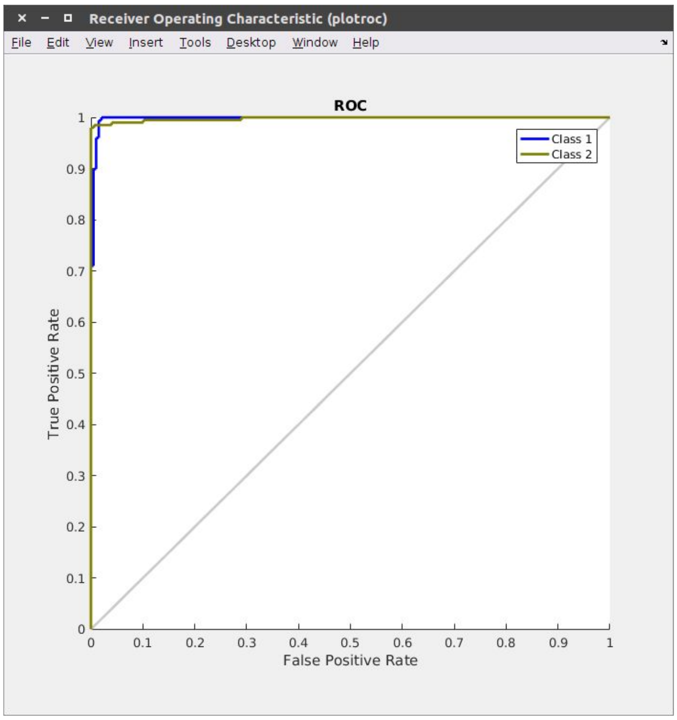

Figure 5. Average Combination Rule

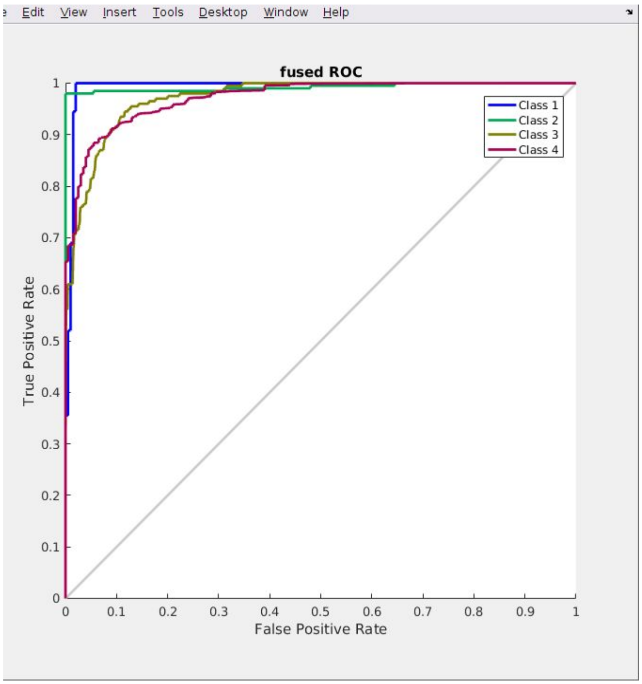

Figure 6. Multiple-Instance Fusion for LDA and PCA, respectively