Universitat de Barcelona

Computer Vision

Lab 6: Face Detection and Recognition

Authors: Emer Rodriguez Formisano and Jorge Alexander

Date: 20th November 2017

Face Recognition

Gender Recognition

Dataset

The ARFace dataset contains image data of the faces of 85 different subjects, each either male or female, taken over several photo sessions. ARFace.person contains a unique person identifier for each of the people in the dataset. ARFace.internal contains the image data for the persons. It has dimensions of 1188 x 2210. 2210 columns, one for each photo taken, and 1188 rows which contain the image data, a reshaped 36 by 33 image matrix (36*33 = 1188). [1]

Fig.1; A face sample from ARFace.internal

Objectives

The aim of this work was to analyse how different techniques, PCA and LDA, can be used for gender recognition from the analysis of face image data.

PCA, principal component analysis, works by calculating the principal components of the dataset. These are orthogonal and uncorrelated sets of values which can be used to represent a high percentage of the variance of the dataset but with less dimensions, thus reducing the dimensions needed to achieve the same performance in classification tasks.

When applied to the ARFace dataset, the implementation used generates Eigenfaces, which is an interesting way to represent the principal components of the PCA of the dataset visually. It shows which features of a face can be used to explain the most variance.

Fig.2, First 30 eigenfaces generated by PCA

It can be observed that the distinct markings of features such as the eigenfaces eyes or mouth, indicate that they are good features for discriminating between faces.

Implementation

To validate the classification accuracy, the true and false positives and negatives are calculated, using fold validation. Fold validation is used to test the model in the training phase to prevent problems such as overfitting. In the implementation used, the fold validation loops through the number of cross-validation folds to be used, looking up the next fold of subjects using a random permutation. These are stored in a column vector “n”. Then, the index of the subjects is looked up and stored in vector “index” in order to extract the training data for that group of subjects, and test a classifier (in this case K-nearest-neighbours) to return the performance scores. [2]

The performance scores that were used to test the data were the following:

- *Sens* - sensitivity, the true positive rate, gives the percentage of positive labels that were correctly identified as such. *TP/(TP+FN)*
- *Spec* - specificity, the true negative rate, gives the percentage of negative labels that were correctly identified as such. *TN/(TN+FP)*
- *Prec* - precision, is the fraction of positive values that were identified correctly. *TP/(TP+FP)*
- *FAR* - the fall out rate, gives the fraction of negative scores. *FP/(FP+TN)*
- *Recall* - the recall is the same as the sensitivity (see *Sens*) *TP/(TP+FN)*
- *Acc* - accuracy, the sum of the true positives and negatives over the total sample size, gives the percentage of correctly classified values. *(TP+TN)/(TP+TN+FP+FN)*
- *Error* - error, the percentage of incorrectly classified positives and negatives over the total sample size. *(FP+FN)/(TP+TN+FP+FN)*
- *ConfusionMatrix* - the confusion matrix is a table, that allows a summary of the performance. In this case, it is a 2x2 grid, that has the mean of the true positive and false negatives as the first row, and the mean of the false positives and true negatives in the second row. *[TP FN; FP TN]*

[3]

Results

The results show three different component analysers:

1. PCA dim=5; the principle component analysis using the top 5 components based on their contribution to the dimensions variance.
2. PCA var=95%; the principal component analysis using the components which contribute to 95% of the variance. The number of dimensions needed was 123.
3. LDA; the linear discriminant analysis which aims to find a linear combination between the dimensions.

Table.1, benchmarks of different techniques on the same ARFace dataset.

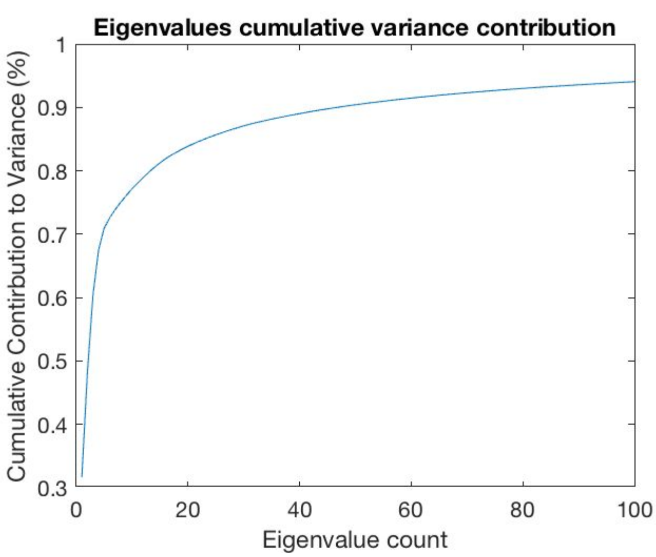
Metric	**PCA dim=5**	**PCA var=95%**	**LDA**
Sens	69.58	33.96	99.43

Spec	88.95	94.62	99.92
Prec	81.56	83.06	99.89
FAR	11.05	5.38	0.08
Recall	69.58	33.96	99.43
Acc	80.96	68.08	99.71
Error	0.19	0.32	0.00
ConfusionMatrix	59.7	26.1 13.5	108.7	30.9	60.1 6.3	110.7	87.9	0.5 0.1	119.5
Dimensions	5	123	2
K nearest neighbours	2	2	2

From the table, it is clear to see that the LDA has the highest sensibility, specificity, precision, recall, accuracy and lowest error and fall out rate. Therefore, from these results it is observed that the LDA approach outperforms the PCA approaches.

However, it is interesting to note the results of PCA95, with an explained variance of 95%, against the PCA with 5 dimensions. The sensibility, recall and accuracy are worst than the standard PCA approach, but it achieves better specificity, precision and fall out rate. As dimensions were increased, it appears to have affected the percentage of positive labels labelled correctly which in turn affected the sensibility, recall and accuracy.

Graph.1, Eigenvalue contribution to variance.



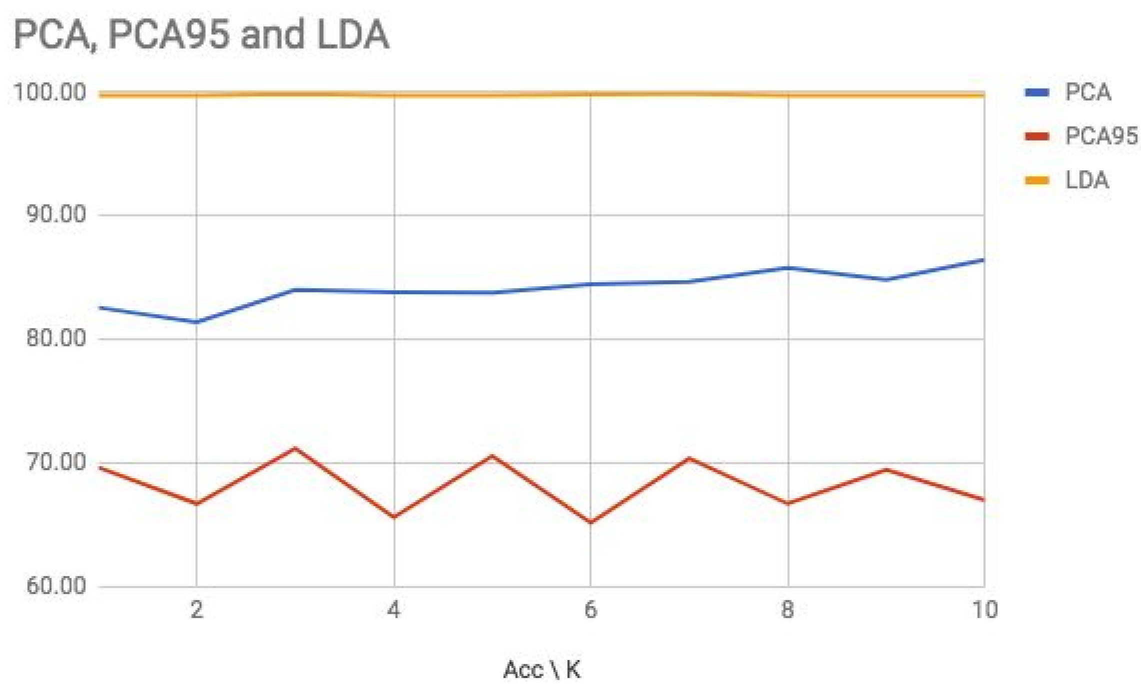
In fact, if the eigenvalues contribution to variance is seen plotted as in Graph.1, it can be seen that beyond around 10 eigenvalues, the variance explained does not increase so rapidly. This would suggest that the gains in specificity, precision and fall out rate above 10 eigenvalues stop to justify the decrease in sensibility, recall and accuracy.

Table.2, The accuracies of the different techniques for increasing K.

Acc \ K	1	2	3	4	5	6	7	8	9	10
PCA	82.60	81.39	84.04	83.85	83.80	84.47	84.66	85.82	84.86	86.44
PCA95	69.66	66.73	71.20	65.63	70.58	65.19	70.38	66.73	69.47	67.02
LDA	99.76	99.76	99.86	99.76	99.76	99.81	99.86	99.76	99.76	99.76

It is observed that the LDA approach retains an almost constant accuracy, slightly increasing at K=2, K=6 and K=7. For standard PCA it gradually increases as k is increased. PCA95 is observed to fluctuate on even numbers for K. A clearer illustration of this can be observed in Graph.2.

Graph.2, Visualisation of increasing k-value with the different techniques



Conclusions

It was observed that the LDA performed significantly better than the PCA and PCA95. The code sample below shows that PCA and PCA95 perform without the labels of the images, whereas the LDA algorithm uses the labels of the images to model the differences between the labels. This allows the LDA to achieve a higher accuracy than the PCA methods.

```
CodeSample.1, PCA, PCA95 and LDA
mat_features_pca = feature_extraction('PCA', images);
mat_features_pca95 = feature_extraction('PCA95', images);
mat_features_lda = feature_extraction('LDA', images, labels);
```

The comparison of PCA and PCA95, demonstrated that an increase in the number of principle components and eigenvalues used, does not lead to a better performance score. In fact PCA95 was observed to have a much lower accuracy (among other metrics) than PCA, even though it was using components that accounted for 95% of the variance. This is because the increase in components, lead to a lower percentage of positive labels labelled correctly, which in turn affected the sensibility, recall and accuracy negatively. This can be understood by the fact that there are a few main components, that describe the fundamental structure of the data. Once these are captured adding more components simply adds noise and although more variance is represented, more variance does not necessarily mean better data.

When performing this analysis on subject recognition, it has to be made sure that in each fold iteration there is at least one image of each of the people in the training set, because otherwise there will be no relevant images to classify the subject in that iteration.[4]

Questions

[1]
• Which is the information contained in ARFace.person?
• Why the size of the eld internal, size(ARFace.internal), is 1188 x 2210?

[2]
• What are the variables 'n' and 'index' of the function fold_validation.m?

[3]
• The provided code computes some evaluation measures of the results obtained with 'PCA' (dim=5), 'PCA95' (95% variance explained) and 'LDA'. Which is the meaning of these measures?

[4]
• Knowing that the AR-Face database has several instances (photos) of the same subject, how do you have to distribute the samples in the fold-cross validation strategy for the subject recognition problem?