Machine Learning

2228 - CSE 6363 - SEC 002

HW2 - REPORT

Names of Group members

1. Pratik Antoni Patekar (1001937948)
2. Amrita Singh (1001937490)
3. Ruthvik Kumar Myadam (1002026231)
4. Prathibha Lakkidi (1001962876)
5. Harshini Kandimalla (1001960046)

Results :

PCA

Quantitative Analysis :

Svd_solver	n_components	Accuracy	Precision	Recall	f1 score
randomized	10	0.50	0.42	0.46	0.41
randomized	150	0.86	0.87	0.75	0.80
randomized	300	0.81	0.78	0.74	0.75

Svd_solver	n_components	Accuracy	Precision	Recall	f1 score
full	10	0.51	0.42	0.47	0.41
full	150	0.85	0.83	0.75	0.78
full	300	0.81	0.79	0.72	0.74

Svd_solver	n_components	Accuracy	Precision	Recall	f1 score
arpack	10	0.49	0.41	0.45	0.40
arpack	150	0.84	0.81	0.78	0.79
arpack	300	0.82	0.78	0.74	0.75

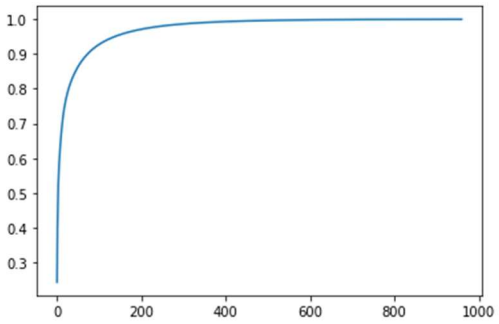
ICA :

Quantitative Analysis :

Algorithm	n_components	Accuracy	Precision	Recall	f1 score
parallel	10	0.43	0.38	0.43	0.36
parallel	150	0.81	0.71	0.76	0.73
parallel	300	0.80	0.70	0.74	0.72

Algorithm	n_components	Accuracy	Precision	Recall	f1 score
deflation	10	0.43	0.38	0.42	0.36
deflation	150	0.81	0.72	0.76	0.74
deflation	300	0.80	0.70	0.74	0.72

Discussion:

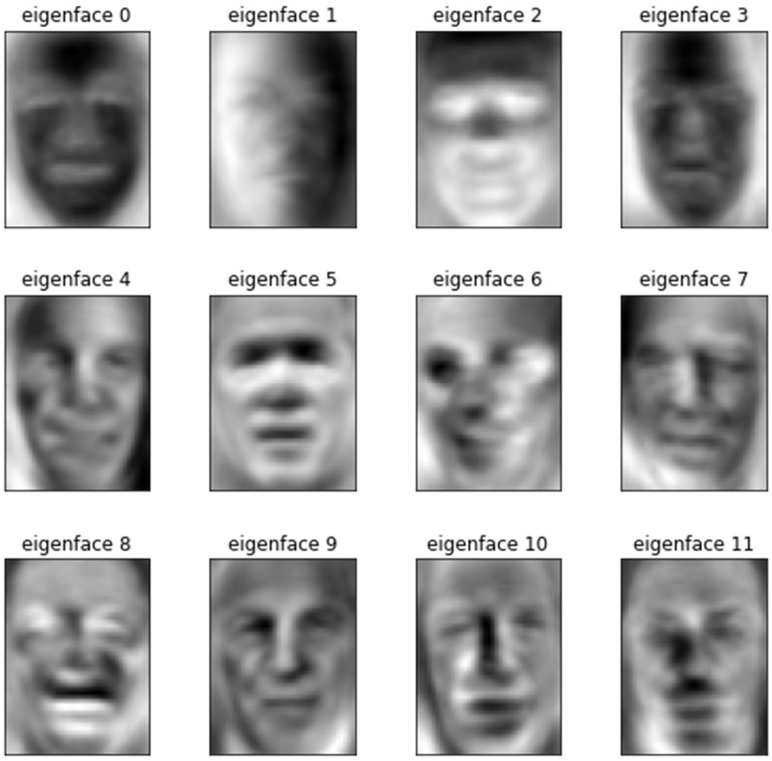


PCA: In PCA we have selected different values of two parameters n_components and svd_solver. For every solver we can see that the quantitative score of each model is higher for n_components = 150 as compared with that for 10 and 300.

This plot basically plots the cumulative variance ratio depending upon the n_components. Here the maximum variance change occurs at a value of 150 and for values greater than 150 the variance is almost constant i.e. 1. Due to this reason, the accuracy and other quantitative analysis scores are better for n_components = 150, lower for n_components = 10 and almost similar for n_components = 300.

ICA: In ICA, similar results can be seen. The difference between PCA and ICA is that, in PCA we decompose the number of components on the basis of variance and in ICA we decompose on the basis of independence between those components. In ICA, the Accuracy, precision, recall and f1 score is better for n_components = 150 as compared to other values.

Below are the eigenfaces that are generated in PCA.



In this, the portion of the images with maximum variance are brighter and the portion with the minimum variance are darker. In ICA, the eigenfaces that are generated do not have brighter or darker portions depending on variance. Whereas here the portion with edges are darker and highlighted compared to other portions. Below is the snap of the eigenfaces that are generated by ICA.