Statistical Methods in AI

ASSIGNMENT 2
FACE CLASSIFICATION

Soumyasis Gun
Roll no. 20171002

QUESTION 1.

a) What are Eigenfaces?

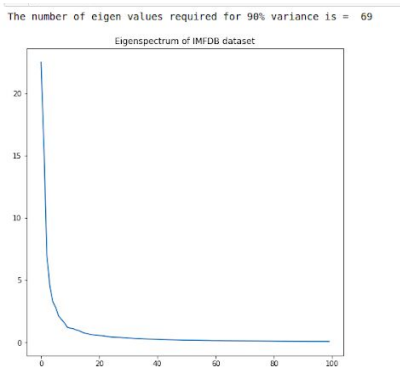
In a collection of human face images in a dataset, it is sometimes useful to measure the variances and use them to extract meaningful features. Since we know of dimensionality reduction techniques like PCA which utilise covariance to project data into the dimensions with maximum variance, we can similarly take a face dataset and project it into dimensions that represent useful information to problem setting (like maximising variance as in PCA and LDA)

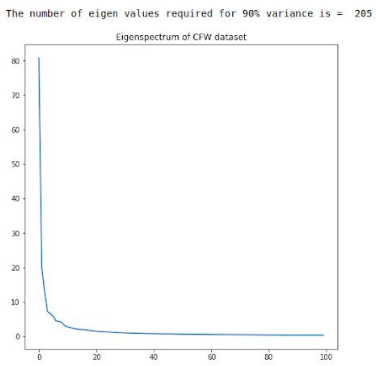
Eigenfaces are the eigenvectors corresponding to the dimensions of maximal variance in a human face dataset

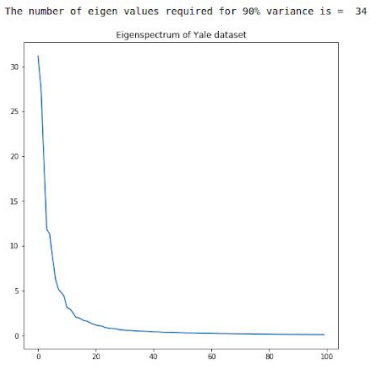
b) How many eigenvectors / faces are required to "satisfactorily" reconstruct a person in these three datasets? (Don't forget to make your argument based on eigenvalue spectrum) Show appropriate graphs, qualitative examples and make a convincing argument.

If we look at the plot of eigenvalue spectrum for a dataset, we can estimate the number of eigenvectors required to satisfactorily reconstruct a person. Let us consider 90% variance preservation.

The eigenvalue spectrum for different datasets are :







c) Reconstruct the image back for each case.

d) Which person/identity is difficult to represent compactly with fewer eigenvectors? Why is that? Explain with your empirical observations and intuitive answers

The persons per dataset who are difficult to represent will have maximum reconstruction error for their classes. For the given dataset, these are :

```
Dataset =  IMFDB
Using PCA, Class with highest reconstruction error =  ShilpaShetty ,id =  3
Dataset =  CFW
Using PCA, Class with highest reconstruction error =  ManmohanSingh ,id =  6
Dataset =  Yale
Using PCA, Class with highest reconstruction error =  7
```

Intuitively if we see in the dataset, we can see that the respective classes in each dataset have a lot of intra-class variance in terms of emotions, angle of face captured, closed and open eyes, lighting, so the results intuitively match to the empirical ones.

Also, given to extreme variations in the Cartoonised dataset CFW, it is the hardest dataset to reconstruct while aiming for preserving minimum number of eigenvalues, hence it needs the maximum number of eigenvalues out of all datasets.

QUESTION 2.

a) Use any classifier(MLP, Logistic regression, SVM, Decision Trees) and find the classification accuracy.

I have used an MLP classifier with hidden layer sizes of (1000,1000), optimised with Adam and used ReLU activations

b) Which method works well? Do a comparative study.

I have used 10 different combinations of features per dataset. They are : 'PCA, LDA, KLDA, KPCA, VGG, VGG+PCA, VGG+LDA, ResNet, ResNet+PCA, ResNet+LDA

Here both the kernel methods, KPCA and KLDA use a Radial Basis Function. The comparisons across the datasets and the features are :-

Comparison of different features on IMFDB Dataset

	Dimension	Accuracy	Error	Precision	Recall	F1 score
PCA	69.0	83.75	16.25	0.840132	0.859000	0.837246
LDA	7.0	80.00	20.00	0.805383	0.826786	0.789108
KLDA	7.0	15.00	85.00	0.021739	0.115385	0.036585
KPCA	69.0	78.75	21.25	0.790875	0.811265	0.785423
VGG	4096.0	88.75	11.25	0.885199	0.882077	0.883215
VGG+PCA	69.0	90.00	10.00	0.914103	0.899934	0.904112
VGG+LDA	7.0	61.25	38.75	0.715575	0.603993	0.589528
ResNet	2048.0	95.00	5.00	0.945933	0.940664	0.945486
ResNet+PCA	69.0	93.75	6.25	0.934829	0.939048	0.934250
ResNet+LDA	7.0	96.25	3.75	0.970238	0.961164	0.964451

Comparison of different features on CFW Dataset

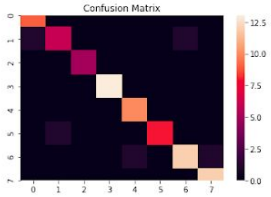
	Dimension	Accuracy	Error	Precision	Recall	F1 score
PCA	205.0	58.518519	41.481481	0.576911	0.603601	0.584794
LDA	7.0	36.296296	63.703704	0.383326	0.364368	0.362585
KLDA	7.0	15.555556	84.444444	0.063588	0.131061	0.085278
KPCA	205.0	52.592593	47.407407	0.582994	0.527699	0.508766
VGG	4096.0	60.000000	40.000000	0.577488	0.559267	0.558977
VGG+PCA	205.0	70.370370	29.629630	0.679819	0.657383	0.660999
VGG+LDA	7.0	60.740741	39.259259	0.565242	0.544065	0.545086
ResNet	2048.0	97.037037	2.962963	0.970504	0.977241	0.973366
ResNet+PCA	205.0	97.777778	2.222222	0.975815	0.983820	0.979284
ResNet+LDA	7.0	97.037037	2.962963	0.973280	0.972456	0.972103

Comparison of different features on Yale Dataset

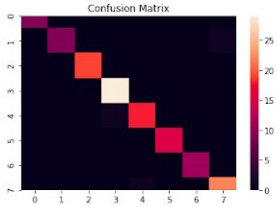
	Dimension	Accuracy	Error	Precision	Recall	F1 score
PCA	34.0	78.787879	21.212121	0.744048	0.773810	0.725510
LDA	14.0	90.909091	9.090909	0.922619	0.934524	0.915476
KLDA	14.0	6.060606	93.939394	0.004329	0.071429	0.008163
KPCA	34.0	90.909091	9.090909	0.833333	0.904762	0.852381
VGG	4096.0	57.575758	42.424242	0.467778	0.483333	0.448889
VGG+PCA	34.0	54.545455	45.454545	0.601111	0.466667	0.482222
VGG+LDA	14.0	54.545455	45.454545	0.444444	0.522222	0.465368
ResNet	2048.0	96.969697	3.030303	0.976190	0.982143	0.975510
ResNet+PCA	34.0	100.000000	0.000000	1.000000	1.000000	1.000000
ResNet+LDA	14.0	100.000000	0.000000	1.000000	1.000000	1.000000

The confusion matrices for the best models per dataset are shown in a form of heatmap below :-

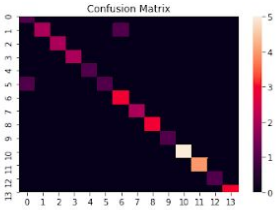
```
Best model for each dataset
Using  resnet  first
Projected using  pca

Dimension =  (320, 69)
Using  IMFDB  dataset
Validation Accuracy = 0.93750
Precision = 0.93700
Recall = 0.93905
F1 - Score = 0.93623
```




```
Using  resnet  first
Projected using  pca

Dimension =  (537, 205)
Using  CFW  dataset
Validation Accuracy = 0.97778
Precision = 0.97549
Recall = 0.98382
F1 - Score = 0.97929
```



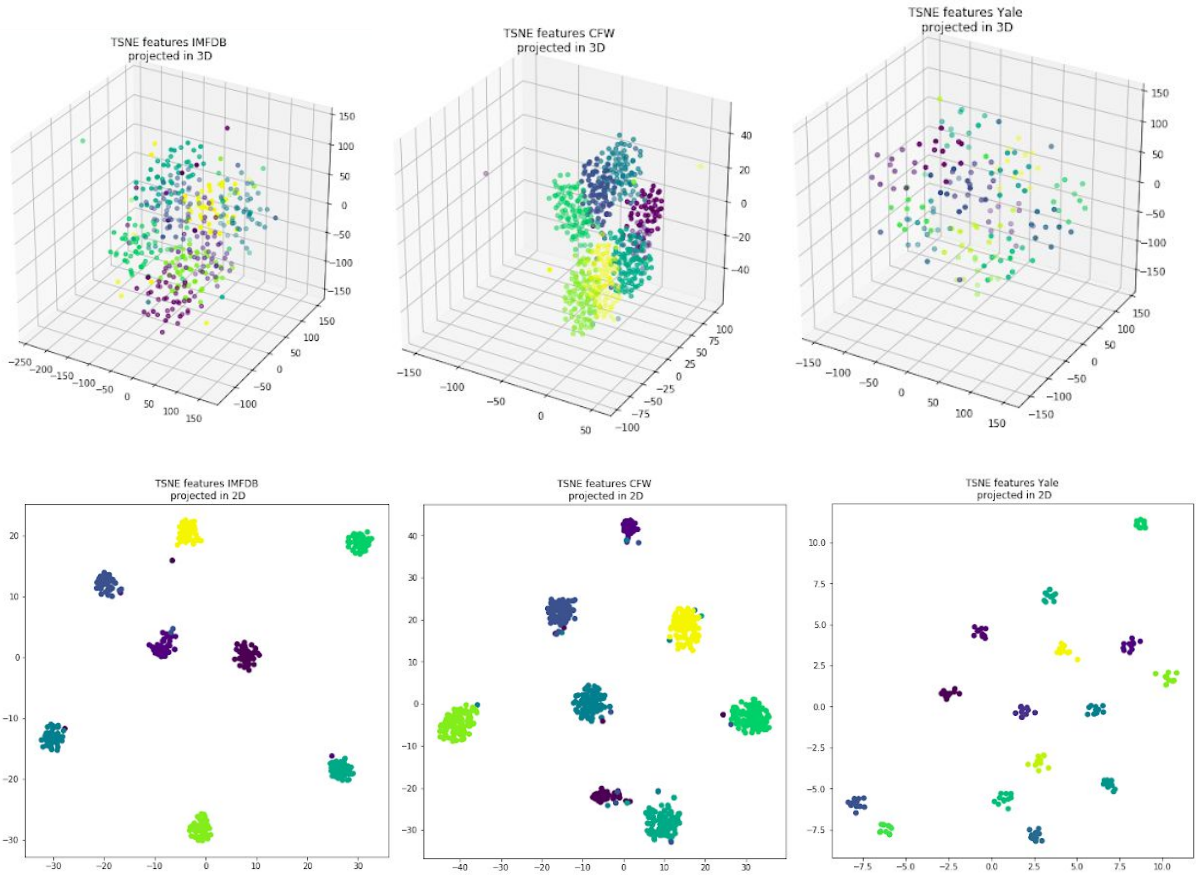

```
Using  resnet  first
Projected using  pca

Dimension =  (132, 34)
Using  Yale  dataset
Validation Accuracy = 0.93939
Precision = 0.94048
Recall = 0.94643
F1 - Score = 0.92789
```



QUESTION 3.

a) Similar to 1(b) use t-SNE based visualization of faces? Does it makesense? Do you see similar people coming together?or something else? Can you do visualization dataset wise and combined?

TSNE is a non linear dimensionality reduction technique that converts distances between high-dimensional points into probability of similarity. Then it tries to minimise the difference of similarities in higher and lower dimensional space using KL divergence of data points. Here I have used TSNE on PCA and LDA. While PCA is not properly separable, LDA is separable and hence is a good feature for classification.



QUESTION 4 - Face Verification

a) How do we formulate the problem using KNN?

We use KNN as a classifier against features generated by popular dimensionality reduction algorithms like PCA and LDA and feature extractors like VGG and ResNet. Then against a query we check the class predicted and verify it with the preset class ID.

b) How do we analyze the performance ? suggest the metrics (like accuracy) that is appropriate for this task.

To analyse the performance we can use different metrics like Classification accuracy, precision, recall, f1 score. Here accuracy is the most logical choice. The performance metrics for different datasets and features on 3NN (undeclared default) is given :-

Using Unspecified value of K for KNN

Comparison of different features on IMFDB Dataset

	Dimension	Accuracy	Error	Precision	Recall	F1 score
PCA	69.0	60.00	40.00	0.677951	0.655064	0.594332
LDA	7.0	80.00	20.00	0.799693	0.817651	0.789083
KLDA	7.0	3.75	96.25	0.009615	0.053571	0.016304
KPCA	69.0	60.00	40.00	0.718254	0.655064	0.597127
VGG	4096.0	88.75	11.25	0.879132	0.873836	0.874352
VGG+PCA	69.0	88.75	11.25	0.879132	0.873836	0.874352
VGG+LDA	7.0	72.50	27.50	0.722727	0.734419	0.720255
ResNet	2048.0	93.75	6.25	0.938636	0.939048	0.935293
ResNet+PCA	69.0	93.75	6.25	0.938636	0.939048	0.935293
ResNet+LDA	7.0	93.75	6.25	0.944643	0.939048	0.940331

Comparison of different features on CFW Dataset

	Dimension	Accuracy	Error	Precision	Recall	F1 score
PCA	205.0	33.333333	66.666667	0.388634	0.341035	0.314561
LDA	7.0	31.111111	68.888889	0.313624	0.294482	0.292665
KLDA	7.0	20.000000	80.000000	0.075557	0.144079	0.077888
KPCA	205.0	32.592593	67.407407	0.424442	0.334456	0.312065
VGG	4096.0	68.148148	31.851852	0.673355	0.647482	0.648618
VGG+PCA	205.0	68.148148	31.851852	0.673355	0.647482	0.648618
VGG+LDA	7.0	63.703704	36.296296	0.567677	0.569094	0.565641
ResNet	2048.0	97.037037	2.962963	0.973280	0.972456	0.972103
ResNet+PCA	205.0	97.037037	2.962963	0.973280	0.972456	0.972103
ResNet+LDA	7.0	97.037037	2.962963	0.974877	0.972456	0.972759

Comparison of different features on Yale Dataset

	Dimension	Accuracy	Error	Precision	Recall	F1 score
PCA	34.0	72.727273	27.272727	0.719048	0.738095	0.675850
LDA	14.0	93.939394	6.060606	0.940476	0.958333	0.937415
KLDA	14.0	3.030303	96.969697	0.002165	0.071429	0.004202
KPCA	34.0	72.727273	27.272727	0.719048	0.738095	0.675850
VGG	4096.0	51.515152	48.484848	0.413333	0.494444	0.429654
VGG+PCA	34.0	51.515152	48.484848	0.413333	0.494444	0.429654
VGG+LDA	14.0	54.545455	45.454545	0.462222	0.516667	0.464762
ResNet	2048.0	100.000000	0.000000	1.000000	1.000000	1.000000
ResNet+PCA	34.0	100.000000	0.000000	1.000000	1.000000	1.000000
ResNet+LDA	14.0	96.969697	3.030303	0.964286	0.964286	0.952381

c) Show empirical results with all the representations

I have ran the features with a KNN classifier with k varying as 3,5,7. The results for 5NN are included here (from left to right, IMFDB, CFW, Yale respectively) :

Using 5NN

	Dimension	Accuracy	Error	Precision	Recall	F1 score
PCA	69.0	60.00	40.00	0.711269	0.655064	0.589992
LDA	7.0	80.00	20.00	0.799693	0.817651	0.789083
KLDA	7.0	3.75	96.25	0.009615	0.053571	0.016304
KPCA	69.0	60.00	40.00	0.718254	0.655064	0.597127
VGG	4096.0	88.75	11.25	0.879132	0.873836	0.874352
VGG+PCA	69.0	88.75	11.25	0.879132	0.873836	0.874352
VGG+LDA	7.0	72.50	27.50	0.722727	0.734419	0.720255
ResNet	2048.0	93.75	6.25	0.938636	0.939048	0.935293
ResNet+PCA	69.0	93.75	6.25	0.938636	0.939048	0.935293
ResNet+LDA	7.0	93.75	6.25	0.944643	0.939048	0.940331

Using 5NN

	Dimension	Accuracy	Error	Precision	Recall	F1 score
PCA	205.0	32.592593	67.407407	0.417212	0.336869	0.309700
LDA	7.0	31.111111	68.888889	0.313624	0.294482	0.292665
KLDA	7.0	20.000000	80.000000	0.075557	0.144079	0.077888
KPCA	205.0	32.592593	67.407407	0.424442	0.334456	0.312065
VGG	4096.0	68.148148	31.851852	0.673355	0.647482	0.648618
VGG+PCA	205.0	68.148148	31.851852	0.673355	0.647482	0.648618
VGG+LDA	7.0	63.703704	36.296296	0.567677	0.569094	0.565641
ResNet	2048.0	97.037037	2.962963	0.973280	0.972456	0.972103
ResNet+PCA	205.0	97.037037	2.962963	0.973280	0.972456	0.972103
ResNet+LDA	7.0	97.037037	2.962963	0.974877	0.972456	0.972759

Using 5NN

	Dimension	Accuracy	Error	Precision	Recall	F1 score
PCA	34.0	72.727273	27.272727	0.719048	0.738095	0.675850
LDA	14.0	93.939394	6.060606	0.940476	0.958333	0.937415
KLDA	14.0	3.030303	96.969697	0.002165	0.071429	0.004202
KPCA	34.0	72.727273	27.272727	0.719048	0.738095	0.675850
VGG	4096.0	51.515152	48.484848	0.413333	0.494444	0.429654
VGG+PCA	34.0	51.515152	48.484848	0.413333	0.494444	0.429654
VGG+LDA	14.0	54.545455	45.454545	0.462222	0.516667	0.464762
ResNet	2048.0	100.000000	0.000000	1.000000	1.000000	1.000000
ResNet+PCA	34.0	100.000000	0.000000	1.000000	1.000000	1.000000
ResNet+LDA	14.0	96.969697	3.030303	0.964286	0.964286	0.952381

Extension / Application

- Perform Gender Classification on images from IMFDB and IIIT-CFW datasets
- Convention :-
    - Males are assigned label = 0
    - Females are assigned label = 1

Pipeline

We will search for the best features by running the features on an MLP classifier with hidden layer of size (1000,1000) and noting the accuracy for the following features - PCA, LDA, KLDA (rbf), KPCA (rbf), VGG, VGG+PCA, VGG+LDA, ResNet, ResNet+PCA, ResNet+LDA

Results

The comparison between different features are given here on the right.

Comparison of different features on IMFDB Dataset

	Dimension	Accuracy	Error	Precision	Recall	F1 score
PCA	69.0	92.50	7.50	0.926768	0.923559	0.924576
LDA	1.0	83.75	16.25	0.846032	0.841479	0.837271
KLDA	1.0	41.25	58.75	0.412779	0.419173	0.404969
KPCA	69.0	88.75	11.25	0.890476	0.885338	0.886632
VGG	4096.0	95.00	5.00	0.950000	0.951128	0.949969
VGG+PCA	69.0	97.50	2.50	0.974937	0.974937	0.974937
VGG+LDA	1.0	83.75	16.25	0.839683	0.835213	0.836246
ResNet	2048.0	98.75	1.25	0.987179	0.988095	0.987482
ResNet+PCA	69.0	98.75	1.25	0.987179	0.988095	0.987482
ResNet+LDA	1.0	98.75	1.25	0.988372	0.986842	0.987451

Comparison of different features on CFW Dataset

	Dimension	Accuracy	Error	Precision	Recall	F1 score
PCA	205.0	86.666667	13.333333	0.820946	0.771429	0.791667
LDA	1.0	71.851852	28.148148	0.607018	0.616667	0.610892
KLDA	1.0	32.592593	67.407407	0.524135	0.519048	0.323550
KPCA	205.0	87.407407	12.592593	0.847744	0.764286	0.794521
VGG	4096.0	96.296296	3.703704	0.977273	0.916667	0.942918
VGG+PCA	205.0	96.296296	3.703704	0.977273	0.916667	0.942918
VGG+LDA	1.0	88.888889	11.111111	0.841574	0.833333	0.837336
ResNet	2048.0	99.259259	0.740741	0.995283	0.983333	0.989156
ResNet+PCA	205.0	99.259259	0.740741	0.995283	0.983333	0.989156
ResNet+LDA	1.0	99.259259	0.740741	0.995283	0.983333	0.989156

We can additionally explore different methods of splitting available data into training and validation using k-fold cross validation technique. To choose a good k, we can try iterations on k. Here I have tried k = [4, 8, 12, 16]. The accompanying results are given here below. From here I have selected ResNet + PCA feature for classification.

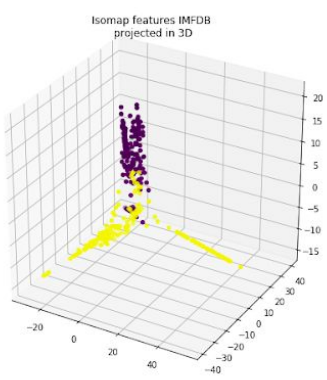
Variations in no. of folds using IMFDB dataset

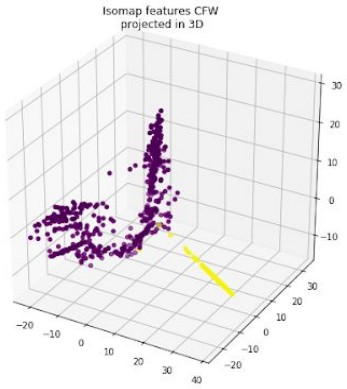
	Accuracy	Error	F1 score
4	97.0	3.0	0.969925
8	98.0	2.0	0.979928
12	100.0	0.0	1.000000
16	100.0	0.0	1.000000

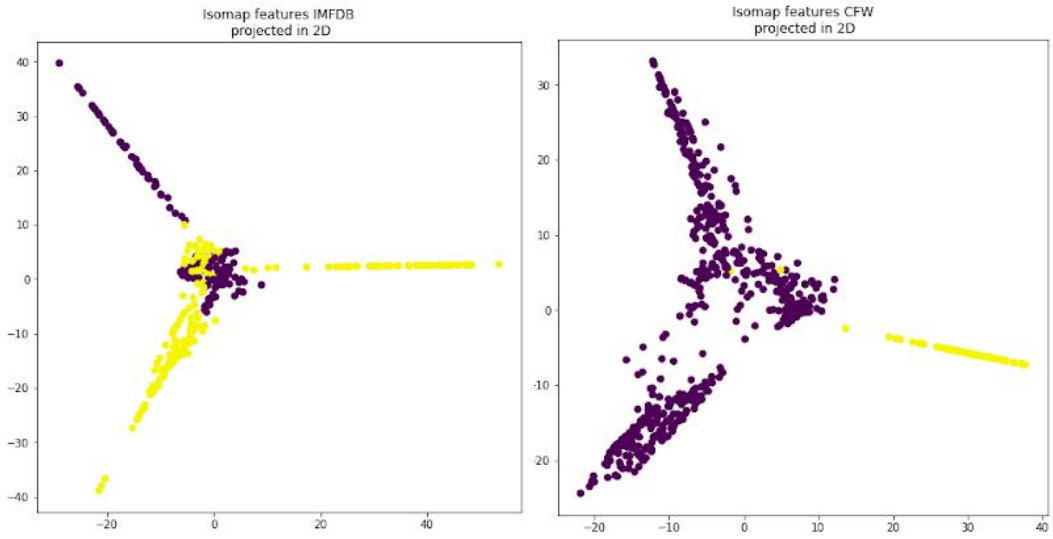
Variations in no. of folds using CFW dataset

	Accuracy	Error	F1 score
4	100.0	0.0	1.0
8	100.0	0.0	1.0
12	100.0	0.0	1.0
16	100.0	0.0	1.0

The TSNE and IsoMap plots were tried, out of which, the IsoMap plots show a good separability in 3D and 2D. The projections are given below :-

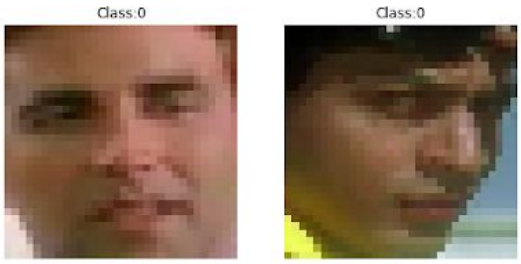






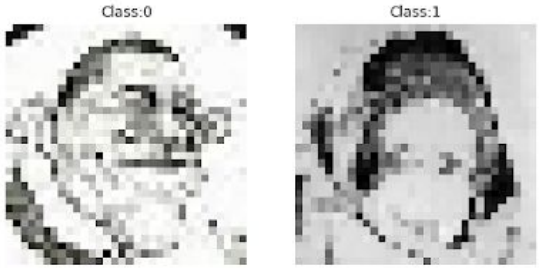
The quantitative results are as shown :-

```
Dataset used :  IMFDB
First projection by  resnet
Projected using  pca

No of wrongly classified samples =  3
No of correctly classified samples =  197
Showing Correctly predicted labels

Correct labels =  0  and  0
Correctly predicted labels =  0  and  0
```

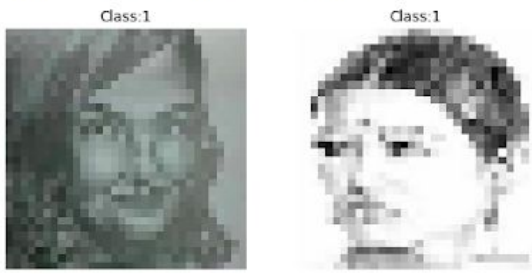



```
Showing Wrongly predicted labels

Correct labels =  0  and  0
Wrongly predicted labels =  1  and  1
```


```
Dataset used :  CFW
First projection by  resnet
Projected using  pca

No of wrongly classified samples =  2
No of correctly classified samples =  334
Showing Correctly predicted labels

Correct labels =  0  and  1
Correctly predicted labels =  0  and  1
```




```
Showing Wrongly predicted labels

Correct labels =  1  and  1
Wrongly predicted labels =  0  and  0
```



Uses and Applications :

- A gender classifier can be used on popular imaging apps like Snapchats to suggest filters based on gender.
- Online shopping apps suggest products based on gender which can be found from the account's uploaded picture.
- In offices like governmental offices, many forms have to be filled. Those can be autofilled, using the subject's image.
- Gender as a feature for human images has uses as a feature in many other problems like validating FaceID in Face Verification classifiers.