SMAI Assignment-2 Technical Report

Anoushka Vyas - 20171057

***Abstract*— The objective of this assignment is to get familiarized with the problems of "classification" and "verification" with a popular problem space of "face" using three datasets - Yale Face Dataset, Indian Movie Face Dataset, IIIT Cartoon Face Dataset.**

I. REPRESENTATION

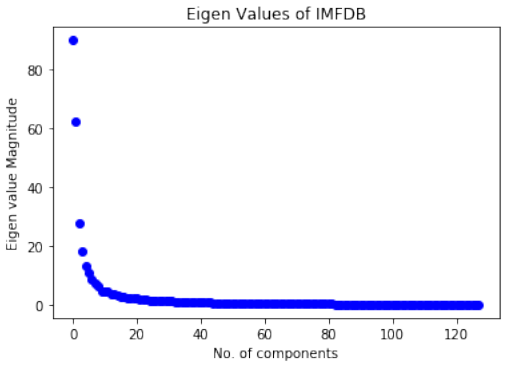
A. What are Eigen Faces?

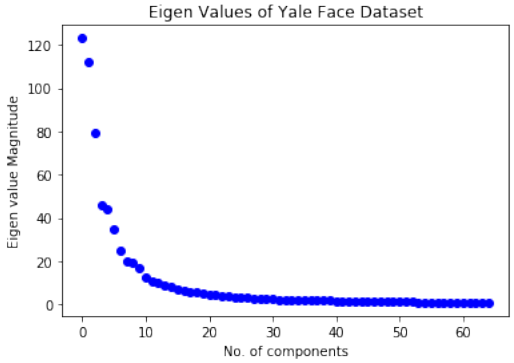
Eigen Faces are the eigen vectors of the covariance matrix of a large set of images depicting human faces which are used to represent the image. We can call them the principal components of the images. Any human face can be considered to be a combination of these standard faces.

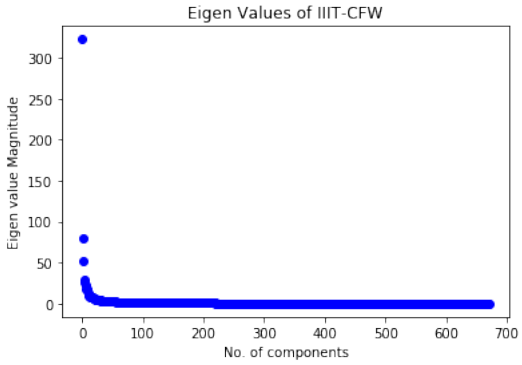
B. How many eigen vectors/faces are required to "satisfactorily" reconstruct a person in these three datasets?

To get 95 percent reconstruction of the face, find the eigen values of the covariance matrix and find the number of eigen values (descending order) required to make their sum equal to 95 percent of the sum of all eigen values. After getting the number of eigen values required, we take the corresponding eigen vectors and perform PCA to get the represenation in feature space and then performing reconstruction. These eigen vectors represent 95 percent of the variance of the data. From the graphs below we can observe that the number of components required for reconstruction in:

Dataset	Components
IMFDB	128
IIIT-CFW	360
Yale Face Dataset	65







C. Which person/identity is difficult to represent com- pactly with fewer eigen vectors? Why is that?

The reconstruction error for each class is calculated for the three data sets and the class with maximum reconstruction error is declared as the class which is most difficult to represent compactly with fewer eigen vectors. PCA is used for reconstruction with the above shown number of components for each dataset.

Dataset	Class	Reconstruction Error
IMFDB	2	0.04062
IIIT-CFW	2	0.073620
Yale Face Dataset	0	0.05877

D. Which dataset is difficult to represent compactly with fewer eigen vectors? Why is it so?

The reconstruction error for each data set is calculated using PCA reconstruction with number of components for each dataset as shown above and the maximum error observed is in the IIIT-CFW dataset. As observed above that IIIT-CFW dataset requires the maximum number of eigen vectors to represent 95%of the variance thus, it has the maximum error as the faces are of cartoon and the features of cartoons are unique in every sample.

Dataset	Reconstruction Error
IMFDB	0.03883
IIIT-CFW	0.06852
Yale Face Dataset	0.05423

II. CLASSIFICATION

A. MLP Classifier

Attributes	Value
Hidden Layers	Two of size 1000 each
Solver	Adam
Activation	relu
Max iterations	1000

B. Features

The features used for training are:

- PCA
- LDA
- Kernel PCA
- Kernel LDA
- VGG
- Resnet
- PCA of VGG
- PCA of Resnet
- LDA of VGG
- LDA of Resnet

C. Dataset formation

The dataset is split into train and validation data in the ratio 4:1. The train data is used to make the feature extraction model and the features of validation data are extracted using the model of train data. Now, the features of both data is used for training.

D. Results

1) IMFDB

	Feature	Dimension	Error	Accuracy	F1-score
0	PCA	128	11.25	88.75	0.884475
1	Kernel PCA	128	11.25	88.75	0.884942
2	LDA	7	20.00	80.00	0.798347
3	Kernal LDA	7	21.25	78.75	0.791331
4	VGG	4096	10.00	90.00	0.894657
5	Resnet	2048	5.00	95.00	0.945486
6	LDA + VGG	7	36.25	63.75	0.624544
7	LDA + Resnet	7	5.00	95.00	0.955152
8	PCA + Resnet	200	3.75	96.25	0.957629
9	PCA + VGG	200	12.50	87.50	0.873730

For this dataset PCA of Resnet with MLP classifier are giving the best accuracy.

2) Yale Face Dataset

	Feature	Dimension	Error	Accuracy	F1-score
0	PCA	65	21.212121	78.787879	0.725510
1	Kernel PCA	65	24.242424	75.757576	0.692857
2	LDA	14	0.000000	100.000000	1.000000
3	Kernal LDA	14	0.000000	100.000000	1.000000
4	VGG	4096	42.424242	57.575758	0.460000
5	Resnet	2048	3.030303	96.969697	0.969697
6	LDA + VGG	14	45.454545	54.545455	0.465368
7	LDA + Resnet	14	0.000000	100.000000	1.000000
8	PCA + Resnet	80	3.030303	96.969697	0.975510
9	PCA + VGG	80	45.454545	54.545455	0.466667

For this dataset LDA, Kernel LDA and LDA of Resnet with MLP classifier are giving the best accuracy.

3) IIIT-CFW

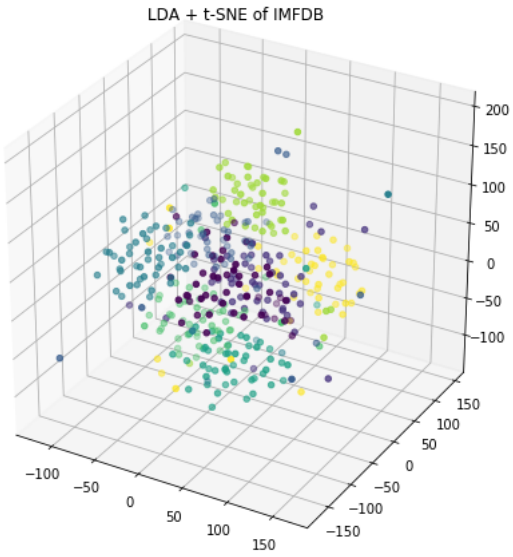
	Feature	Dimension	Error	Accuracy	F1-score
0	PCA	360	39.259259	60.740741	0.610905
1	Kernel PCA	360	39.259259	60.740741	0.600715
2	LDA	7	61.481481	38.518519	0.378212
3	Kernal LDA	7	60.740741	39.259259	0.386181
4	VGG	4096	29.629630	70.370370	0.663776
5	Resnet	2048	2.962963	97.037037	0.973366
6	LDA + VGG	7	39.259259	60.740741	0.530138
7	LDA + Resnet	7	2.222222	97.777778	0.977422
8	PCA + Resnet	300	2.222222	97.777778	0.979945
9	PCA + VGG	300	31.111111	68.888889	0.651646

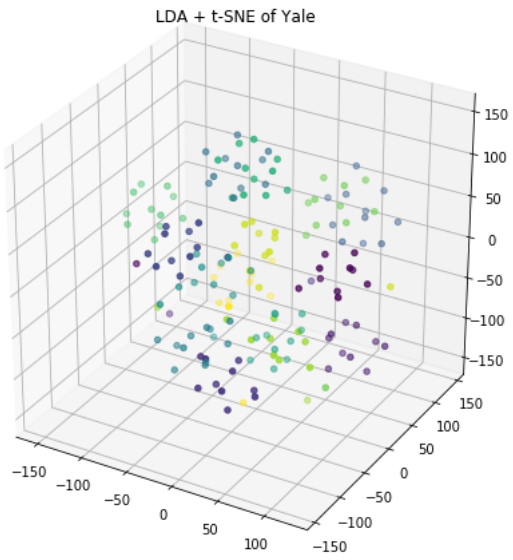
For this dataset LDA of Resnet and PCA of Resnet with MLP classifier are giving the best accuracy.

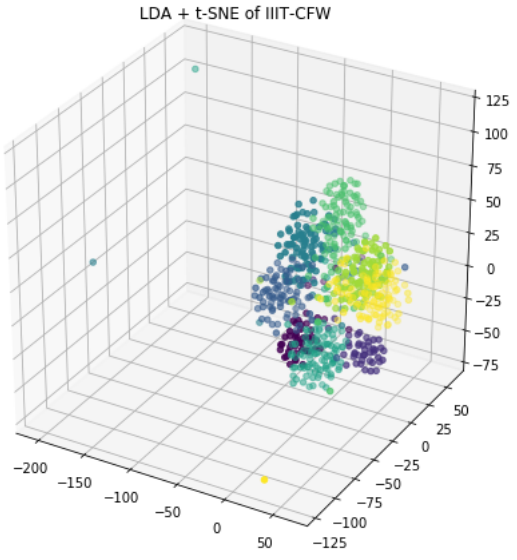
III. T-SNE

t-Distributed Stochastic Neighbor Embedding (t-SNE) is an unsupervised, non-linear technique primarily used for data exploration and visualizing high-dimensional data.

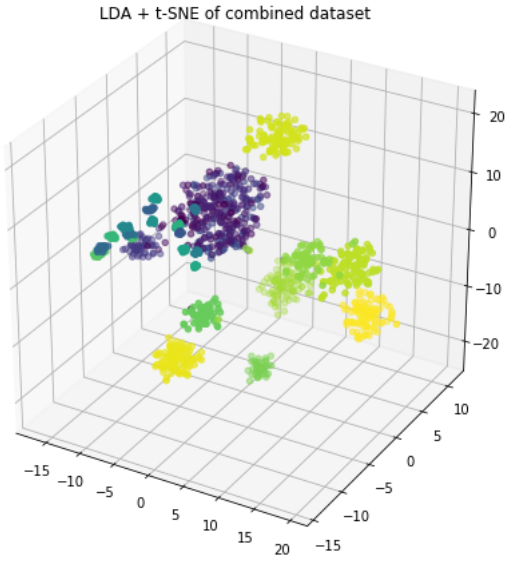
A. Visualization of Individual Datasets







B. Visualization of Combined Dataset



C. Observation

In the 3D t-SNE scatter plot for all the datasets combined, it can be seen that samples with the same label are closer to each other than others with a few exceptions.
In the 3D t-SNE scatter plot for individual datasets also, it can be seen that samples with the same label are closer to each other forming bunches of similar samples with a few exceptions like the yellow scatters.

IV. Face Verification

A. How do we formulate the problem using KNN?

The above mentioned feature extraction methods are applied to the datasets, then train the KNN classifier with the train set extracted from the dataset for each dataset.Now, we find the K (for this problem K=5) nearest neighbours of the validation set and find the nearest neighbour using majority voting and if it is matching the classID then we return T/F.

B. Analysis

1) IMDBF

With K=5:

	Feature	Dimension	Error	Accuracy	Precision
0	PCA	128	43.75	56.25	0.685039
1	Kernel PCA	128	45.00	55.00	0.680060
2	LDA	7	18.75	81.25	0.821479
3	Kernal LDA	7	18.75	81.25	0.821479
4	VGG	4096	11.25	88.75	0.879132
5	Resnet	2048	6.25	93.75	0.938636
6	LDA + VGG	7	27.50	72.50	0.722727
7	LDA + Resnet	7	6.25	93.75	0.944643
8	PCA + Resnet	200	6.25	93.75	0.938636
9	PCA + VGG	200	11.25	88.75	0.879132

Gives best result for LDA of Resnet and PCA of Resnet.
With variable K:

	Feature	K	Dimension	Error	Accuracy	Precision
0	Resnet	4	2048	6.25	93.75	0.938636
1	Resnet	6	2048	6.25	93.75	0.937354
2	Resnet	7	2048	6.25	93.75	0.937354
3	Resnet	10	2048	5.00	95.00	0.954545

2) Yale Face Database

With K=5:

	Feature	Dimension	Error	Accuracy	Precision
0	PCA	65	24.242424	75.757576	0.730952
1	Kernel PCA	65	24.242424	75.757576	0.730952
2	LDA	14	3.030303	96.969697	0.964286
3	Kernal LDA	14	3.030303	96.969697	0.964286
4	VGG	4096	48.484848	51.515152	0.413333
5	Resnet	2048	0.000000	100.000000	1.000000
6	LDA + VGG	14	45.454545	54.545455	0.462222
7	LDA + Resnet	14	3.030303	96.969697	0.964286
8	PCA + Resnet	80	0.000000	100.000000	1.000000
9	PCA + VGG	80	48.484848	51.515152	0.413333

Gives best result for Resnet and PCA of Resnet.
With variable K:

	Feature	K	Dimension	Error	Accuracy	Precision
0	Resnet	4	2048	3.030303	96.969697	0.97619
1	Resnet	6	2048	0.000000	100.000000	1.00000
2	Resnet	7	2048	0.000000	100.000000	1.00000
3	Resnet	10	2048	0.000000	100.000000	1.00000

3) IIIT-CFW

With K=5:

	Feature	Dimension	Error	Accuracy	Precision
0	PCA	360	71.851852	28.148148	0.311155
1	Kernel PCA	360	69.629630	30.370370	0.353009
2	LDA	7	54.074074	45.925926	0.474348
3	Kernal LDA	7	54.074074	45.925926	0.474348
4	VGG	4096	31.851852	68.148148	0.673355
5	Resnet	2048	2.962963	97.037037	0.973280
6	LDA + VGG	7	36.296296	63.703704	0.567677
7	LDA + Resnet	7	2.962963	97.037037	0.974877
8	PCA + Resnet	300	2.962963	97.037037	0.973280
9	PCA + VGG	300	32.592593	67.407407	0.662257

Gives best result for LDA of Resnet , PCA of Resnet and Resnet.
With variable K:

	Feature	K	Dimension	Error	Accuracy	Precision
0	Resnet	4	2048	2.962963	97.037037	0.973609
1	Resnet	6	2048	2.962963	97.037037	0.973280
2	Resnet	7	2048	2.962963	97.037037	0.974877
3	Resnet	10	2048	2.222222	97.777778	0.980400

V. EMOTION CLASSIFICATION

A. Problem Description

The problem is to classify images of two datasets Yale Face dataset and IMFDB based on emotions of the faces. Both the datasets have an emotion.txt file with actual emotion labels.

B. Uses and Applications

- Making Cars Safer and Personalized: Using facial emotion detection smart cars can alert the driver when he is feeling drowsy.
- Facial Emotion Detection in Interviews: Measure candidate's facial expressions to capture their moods and further assess their personality traits.
- Testing for Video Games: Using facial emotion detection can aid in understanding which emotions a user is going through in real-time as he is playing without analyzing the complete video manually.

C. Experiment

1) Pipeline

Load the two datasets and make a dictionary of the emotion.txt file present in the two datasets and then map the correspondnig images with the dictionary to make the new labels of emotions. The emotions are then encoded into numbers using a popular library.

2) Spilts

The dataset is split into train and validation data in the ratio 4:1. The train data is used to make the feature extraction model and the features of validation data are extracted using the model of train data. Now, the features of both data is used for training. The features used are same as those used in the MLP classifier above.

3) Evaluation Metrics

- IMFDB

	Feature	Dimension	Error	Accuracy	Precision
0	PCA	128	38.75	61.25	0.400488
1	Kernel PCA	128	37.50	62.50	0.410494
2	LDA	6	46.25	53.75	0.340580
3	Kernal LDA	6	46.25	53.75	0.340580
4	VGG	4096	40.00	60.00	0.413093
5	Resnet	2048	41.25	58.75	0.412500
6	LDA + VGG	6	73.75	26.25	0.203816
7	LDA + Resnet	6	41.25	58.75	0.538400
8	PCA + Resnet	200	37.50	62.50	0.465136
9	PCA + VGG	200	38.75	61.25	0.416198

Gives best result for Kernel PCA and PCA of Resnet.

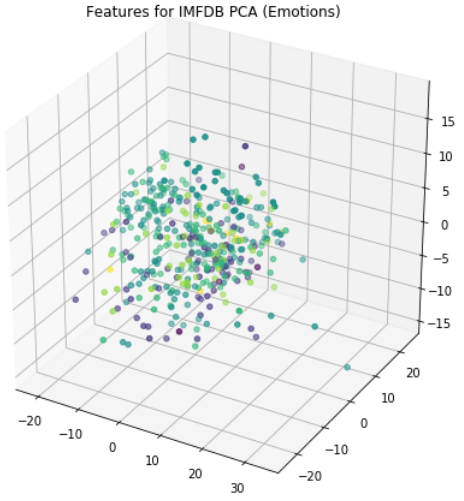
- Yale Face Dataset

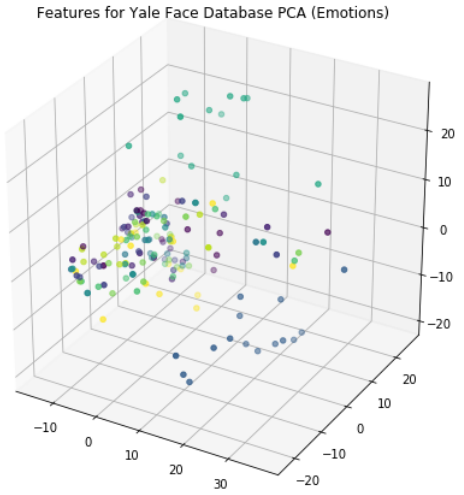
	Feature	Dimension	Error	Accuracy	Precision
0	PCA	65	54.545455	45.454545	0.410606
1	Kernel PCA	65	57.575758	42.424242	0.396970
2	LDA	10	45.454545	54.545455	0.533333
3	Kernal LDA	10	45.454545	54.545455	0.533333
4	VGG	4096	100.000000	0.000000	0.000000
5	Resnet	2048	100.000000	0.000000	0.000000
6	LDA + VGG	10	96.969697	3.030303	0.045455
7	LDA + Resnet	10	100.000000	0.000000	0.000000
8	PCA + Resnet	80	100.000000	0.000000	0.000000
9	PCA + VGG	80	93.939394	6.060606	0.113636

Gives best result for LDA and Kernel LDA.

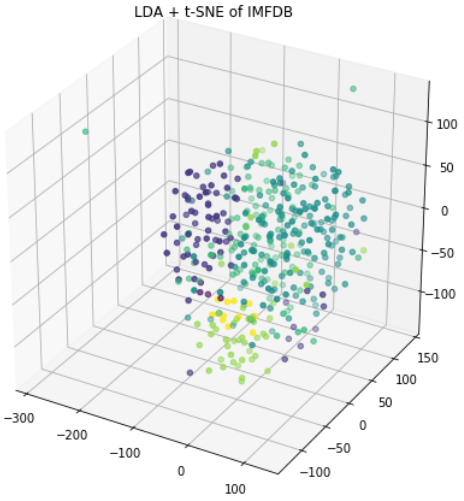
4) Qualitative Results

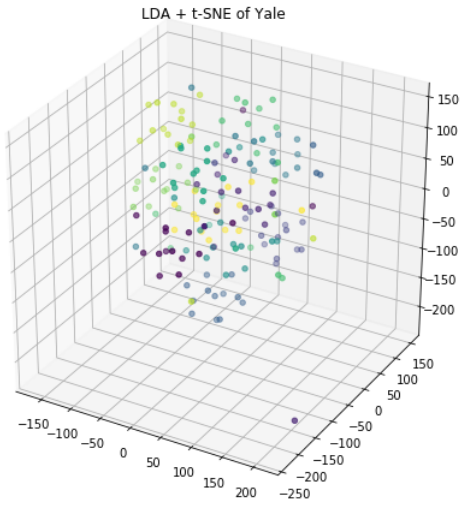
- PCA plots



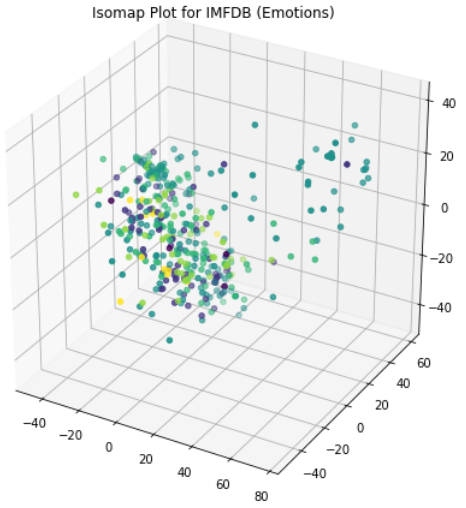


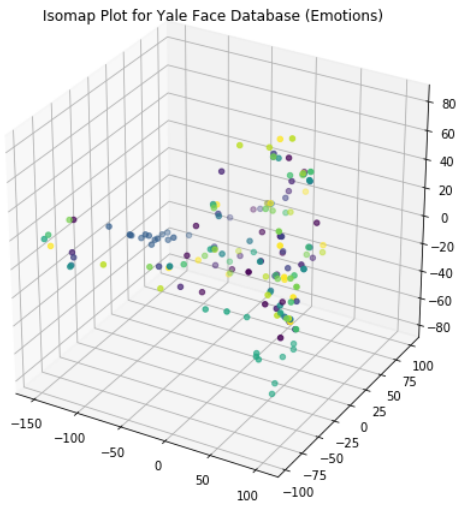
- t-SNE plots





- Isomap plots





- K-Fold validation: It is used to split dataset into train and validation into k:1 k times. So, basically the model is trained k times to find the average accuracy.

IMFDB

	Feature	Fold	Dimension	Error	Accuracy	Precision
0	PCA	4	128	45.500000	54.500000	0.325552
1	PCA	5	128	49.250000	50.750000	0.320562
2	PCA	8	128	46.500000	53.500000	0.404783
3	PCA	10	128	45.500000	54.500000	0.419002
4	PCA	12	128	44.971777	55.028223	0.428110

Yale Face dataset

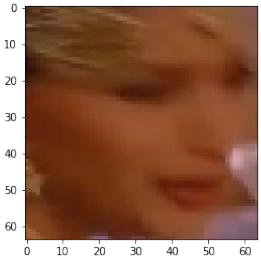
	Feature	Fold	Dimension	Error	Accuracy	Precision
0	LDA	4	10	58.188153	41.811847	0.443300
1	LDA	5	10	57.575758	42.424242	0.435022
2	LDA	8	10	51.369048	48.630952	0.479924
3	LDA	10	10	47.904412	52.095588	0.461515
4	LDA	12	10	52.655678	47.344322	0.429891

5) Quantitative Results

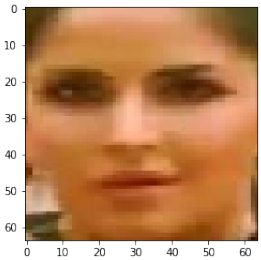
Examples of correct and wrong prediction of each dataset is calculated:

IMFDB

Actual: Happiness Predicted: Happiness

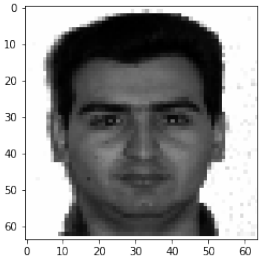


Actual: Happiness Predicted: Neutral



Yale Face dataset

Actual: centerlight Predicted: centerlight



Actual: Happy Predicted: Sad