SMAI ASSIGNMENT 2 - TECHNICAL REPORT

Author - Eesha Dutta [20171104]

1 Setup and How to Run

You need to have installed the following python packages : *numpy*, *pandas* , *sklearn* , *scipy* , *seaborn* and *jupyter-notebook* (or any other equivalent).

2 Datasets And Preprocessing

Three Data Sets are used. All the images have been reduced to 32*32 from the original sizes. Also the 3 channels have been kept intact.
1.IIIT-CFW Dataset : 672,32,32,3 (cfw_dict).
2.IMFDB Dataset : 400,32,32,3 (imfdb_dict).
3.YALE Dataset : 165,32,32,3. A dictionary for each dataset is made with the class name and class index which helps in identification of classes later. Each dataset is flattened to make further processing easy.

3 Question 1 - Feature Extraction

We have used 6 feature extraction methods: PCA, LDA, Kernel PCA, Kernel LDA, VGG and RESNET.

For each dataset we then compute the extracted features for each method and show them in a 3D scatter plot.

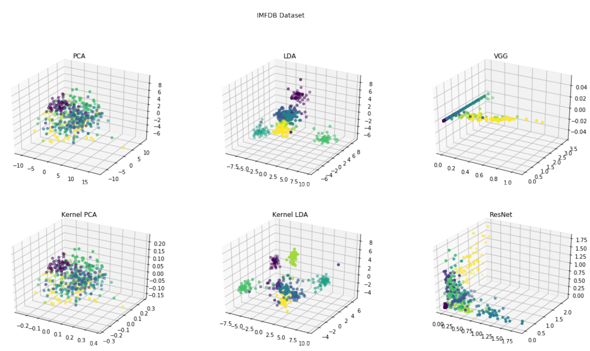

Figure 1: Extracted Features for IMFDB

3.1 What are Eigen Faces?

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).

Eigen Faces are a set of vectors which help to visualise images in a compressed form. They form a set of basis vectors to represent a set of images. We do Principal Component Analysis on all the three datasets and find the compressed form of the images. We then reconstruct back the images using this compressed set (Eigen Faces).

3.2 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.

Considering 95% variance preservation, CFW data being sparse requires 304 Principal Components to satisfactorily reconstruct the images back. Yale requires around 61 PC's and IMFDB requires 118 PC's. We plot the eigen value spectrum of the datasets to infer this.

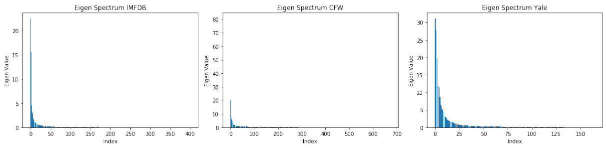

Figure 2: Eigen Spectrum

3.3 Reconstruct the image back for each case

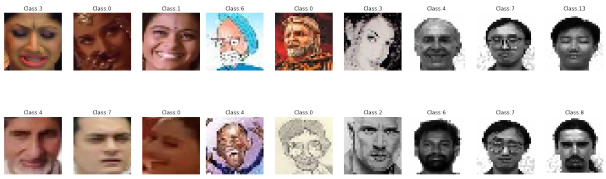

Figure 3: Reconstruction using PCA for IMFDB, CFW and Yale dataset

3.4 Which person/identity is difficult to represent compactly with fewer eigen vectors? 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. In IMFDB dataset - Shilpa Shetty, in CFW dataset - Narendra Modi and in Yale dataset - Class 7. Intuitively, 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.

4 Question 2 - Classification

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

We use MLP classifier with hidden layer size (100, 100). Adam optimiser is used and the hidden layers have ReLU activations.

4.2 Which method works well? Do a comparative study

We use the following combinations of Feature Extraction with MLP classifier - PCA, KPCA, LDA, KLDA, VGG, ResNet, VGG + PCA, VGG + LDA, ResNet + PCA, ResNet + LDA. The comparisons across datasets and different methods are as follows -

	Reduced Dimension	Error	Accuracy	F1-Score
PCA	60.0	16.25	83.75	0.832501
KPCA	60.0	18.75	81.25	0.806230
LDA	7.0	22.50	77.50	0.779937
KLDA	7.0	45.00	55.00	0.543511
VGG	4096.0	11.25	88.75	0.887446
ResNet	2048.0	5.00	95.00	0.945486
VGG + PCA	60.0	10.00	90.00	0.899668
VGG + LDA	7.0	33.75	66.25	0.651257
ResNet + PCA	60.0	6.25	93.75	0.934250
ResNet + LDA	7.0	3.75	96.25	0.964451

Figure 4: IMFDB DATASET-MLP

	Reduced Dimension	Error	Accuracy	F1-Score
PCA	75.0	45.925926	54.074074	0.544740
KPCA	75.0	39.259259	60.740741	0.574574
LDA	7.0	62.962963	37.037037	0.372640
KLDA	7.0	83.703704	16.296296	0.176851
VGG	4096.0	33.333333	66.666667	0.626000
ResNet	2048.0	2.222222	97.777778	0.979289
VGG + PCA	75.0	30.370370	69.629630	0.665453
VGG + LDA	7.0	37.777778	62.222222	0.559427
ResNet + PCA	75.0	2.222222	97.777778	0.979945
ResNet + LDA	7.0	3.703704	96.296296	0.963369

Figure 5: CFW DATASET-MLP

	Reduced Dimension	Error	Accuracy	F1-Score
PCA	40.0	18.181818	81.818182	0.789796
KPCA	40.0	9.090909	90.909091	0.865986
LDA	14.0	6.060606	93.939394	0.949206
KLDA	14.0	24.242424	75.757576	0.664444
VGG	4096.0	45.454545	54.545455	0.454444
ResNet	2048.0	3.030303	96.969697	0.952381
VGG + PCA	40.0	48.484848	51.515152	0.449495
VGG + LDA	14.0	48.484848	51.515152	0.444444
ResNet + PCA	40.0	3.030303	96.969697	0.952381
ResNet + LDA	14.0	0.000000	100.000000	1.000000

Figure 6: YALE DATASET-MLP

The heatmaps of Confusion Matrices for the best methods for each dataset -

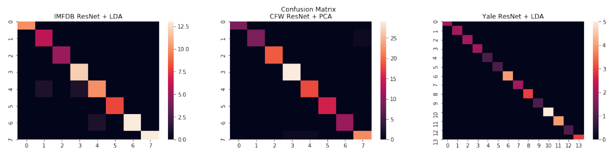

Figure 7: Reconstruction using PCA for IMFDB, CFW and Yale dataset

5 Question 3 - t-SNE

t-Distributed Stochastic Neighbor Embedding (t-SNE) is a non-linear technique for dimensionality reduction that is particularly well suited for the visualization of high-dimensional datasets.

t-Distributed stochastic neighbor embedding (t-SNE) minimizes the divergence between two distributions: a distribution that measures pairwise similarities of the input objects and a distribution that measures pairwise similarities of the corresponding low-dimensional points in the embedding.

In this way, t-SNE maps the multi-dimensional data to a lower dimensional space and attempts to find patterns in the data by identifying observed clusters based on similarity of data points with multiple features. However, after this process, the input features are no longer identifiable, and you cannot make any inference based only on the output of t-SNE. Hence it is mainly a data exploration and visualization technique.

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

t-SNE 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 we use t-SNE on PCA and LDA. While PCA is not properly separable, LDA is separable and hence is a good feature for classification.

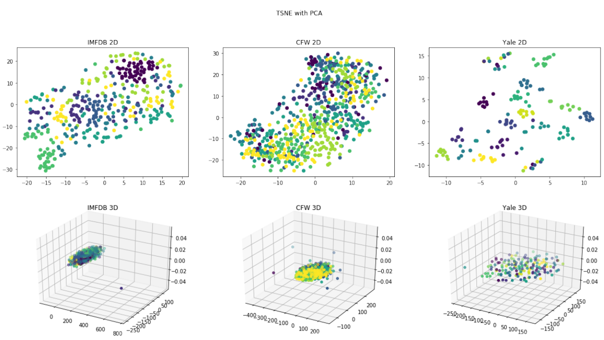

Figure 8: PCA + t-SNE

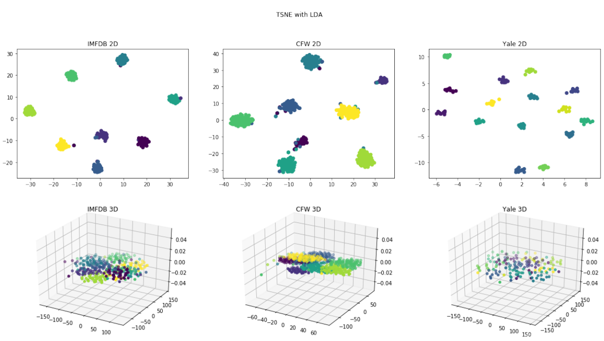

Figure 9: LDA + t-SNE

6 Question 4 - Face Verification

In practice "face" is used for verification. i.e., input is "identity/classID) and the face image" and response is "Yes" or "No" depending on whether the face matches with the given class.

6.1 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.

6.2 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 KNN (k=3) are given -

	Reduced Dimension	Error	Accuracy	Precision
None	3072.0	42.50	57.50	0.688767
PCA	60.0	38.75	61.25	0.711142
LDA	7.0	18.75	81.25	0.821622
VGG	4096.0	8.75	91.25	0.908780
ResNet	2048.0	5.00	95.00	0.954545
VGG + PCA	60.0	8.75	91.25	0.908780
VGG + LDA	7.0	27.50	72.50	0.728896
ResNet + PCA	60.0	5.00	95.00	0.954545
ResNet + LDA	7.0	6.25	93.75	0.944643

Figure 10: IMFDB-KNN

	Reduced Dimension	Error	Accuracy	Precision
None	3072.0	67.407407	32.592593	0.477535
PCA	75.0	60.000000	40.000000	0.476267
LDA	7.0	68.888889	31.111111	0.320640
VGG	4096.0	32.592593	67.407407	0.663551
ResNet	2048.0	2.222222	97.777778	0.975815
VGG + PCA	75.0	32.592593	67.407407	0.663551
VGG + LDA	7.0	37.037037	62.962963	0.560701
ResNet + PCA	75.0	2.222222	97.777778	0.975815
ResNet + LDA	7.0	2.962963	97.037037	0.974877

Figure 11: CFW-KNN

	Reduced Dimension	Error	Accuracy	Precision
None	3072.0	24.242424	75.757576	0.754762
PCA	40.0	24.242424	75.757576	0.754762
LDA	14.0	9.090909	90.909091	0.916667
VGG	4096.0	54.545455	45.454545	0.354444
ResNet	2048.0	0.000000	100.000000	1.000000
VGG + PCA	40.0	54.545455	45.454545	0.354444
VGG + LDA	14.0	42.424242	57.575758	0.494444
ResNet + PCA	40.0	0.000000	100.000000	1.000000
ResNet + LDA	14.0	3.030303	96.969697	0.964286

Figure 12: YALE-KNN

We obtain best accuracy for feature extraction using ResNet and Dimensionality Reduction using PCA.

6.3 Show emperical results with all the representations and variations in k

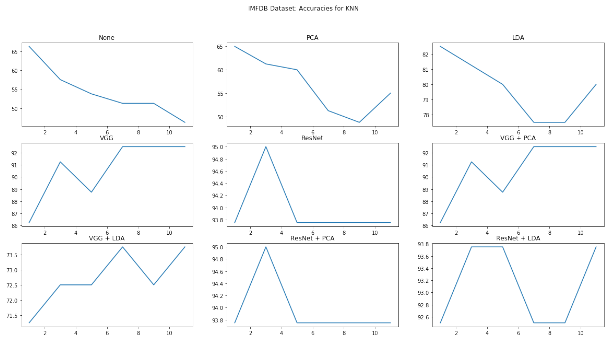

Figure 13: IMFDB KNN Variations

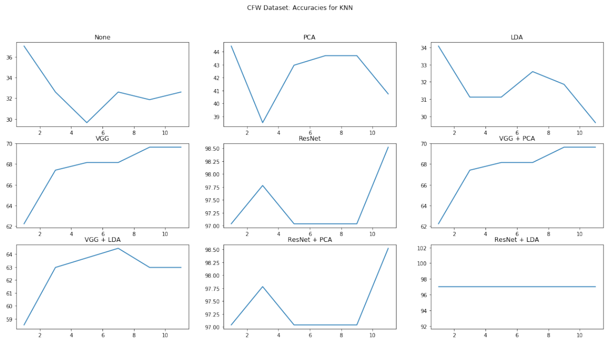

Figure 14: CFW KNN Variations

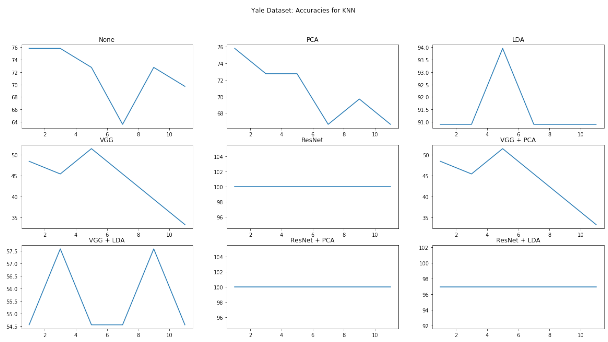

Figure 15: YALE KNN Variations

7 Extension/Application - Perform Gender Classification on IMFDB and IIIT-CFW dataset

Males are assigned label 0 and females are assigned label 1 for this task. We search for the best features by running the features on an MLP classifier with hidden layer of size (100,100) and noting the accuracy for the following features - PCA, LDA, KLDA (rbf), KPCA (rbf), VGG, VGG+PCA, VGG+LDA, ResNet, ResNet+PCA, ResNet+LDA

	Reduced Dimension	Error	Accuracy	Precision	F1-Score
PCA	60.0	5.00	95.00	0.952020	0.949717
KPCA	60.0	11.25	88.75	0.887806	0.887059
LDA	1.0	13.75	86.25	0.863352	0.862479
KLDA	1.0	23.75	76.25	0.762099	0.761569
VGG	4096.0	5.00	95.00	0.950000	0.949969
ResNet	2048.0	0.00	100.00	1.000000	1.000000
VGG + PCA	60.0	3.75	96.25	0.962164	0.962447
VGG + LDA	1.0	16.25	83.75	0.839683	0.836246
ResNet + PCA	60.0	0.00	100.00	1.000000	1.000000
ResNet + LDA	1.0	1.25	98.75	0.988372	0.987451

Figure 16: IMFDB Gender Classification

	Reduced Dimension	Error	Accuracy	Precision	F1-Score
PCA	75.0	10.370370	89.629630	0.871622	0.837963
KPCA	75.0	13.333333	86.666667	0.820946	0.791667
LDA	1.0	25.185185	74.814815	0.635714	0.635714
KLDA	1.0	45.185185	54.814815	0.537912	0.504662
VGG	4096.0	4.444444	95.555556	0.957833	0.932410
ResNet	2048.0	1.481481	98.518519	0.978571	0.978571
VGG + PCA	75.0	4.444444	95.555556	0.972973	0.930556
VGG + LDA	1.0	11.111111	88.888889	0.841574	0.837336
ResNet + PCA	75.0	0.740741	99.259259	0.995283	0.989156
ResNet + LDA	1.0	0.740741	99.259259	0.995283	0.989156

Figure 17: CFW Gender Classification

7.1 k-fold validation

We can 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. The results are as follows using ResNet and PCA for classification -

	Avg error	Avg acc	Avg F1
4	10.750000	89.250000	0.891343
8	9.500000	90.500000	0.901973
12	10.264409	89.735591	0.894196
16	8.500000	91.500000	0.912355

Figure 18: k-fold cross validation for IMFDB

	Avg error	Avg acc	Avg F1
4	12.351190	87.648810	0.729601
8	11.309524	88.690476	0.770538
12	12.053571	87.946429	0.741655
16	11.011905	88.988095	0.776010

Figure 19: k-fold cross validation for CFW

7.2 Quantitative Results

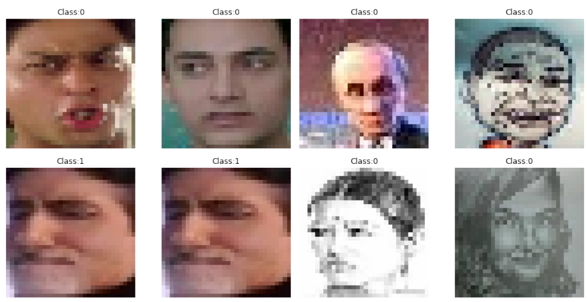

Figure 20: The 4 images on the top are the correct predictions of IMFDB and CFW dataset respectively (2 each) and the 4 images at the bottom are the wrong predictions of IMFDB and CFW dataset respectively (2 each)

7.3 IsoMap Plots

IsoMap plots show a good separability in 2D and 3D.

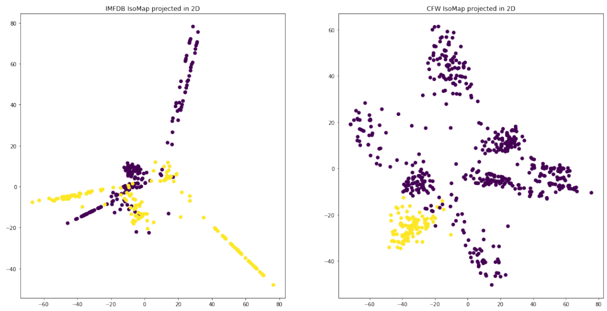

Figure 21: Isomap features projected in 2D

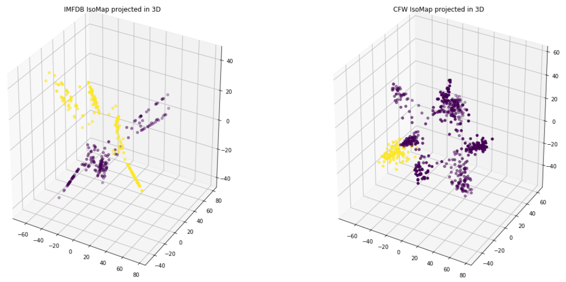

Figure 22: Isomap features projected in 3D