PATTERN RECOGNITION ASSIGNMENT- 4
REPORT

15h November 2018

GROUP - 11

ABHIGYAN KHAUND
B16082

PRATYUSH GAURAV
B16026

SHASHWAT GARG
B16034

Faculty
Dr. A.D Dileep
Indian Institute of Technology, Mandi

Table of Contents

1. Introduction

Principal Component Analysis:

Principal Component Analysis (PCA) is a dimension-reduction tool that can be used to reduce a large set of variables to a small set that still contains most of the information in the large set. PCA transforms a number of correlated variables into smaller number of uncorrelated variables called principal components.

PCA only takes those directions in which the information content is maximum. While doing so it does not take into account the separability of data. It might so happen that, the direction in which separability of 2 classes is maximum is not the direction of maximum variance of projected data.

Data Set :

3 class scene image dataset whose 32-dimensional BoVW representation is taken which was generated in Assignment-2.

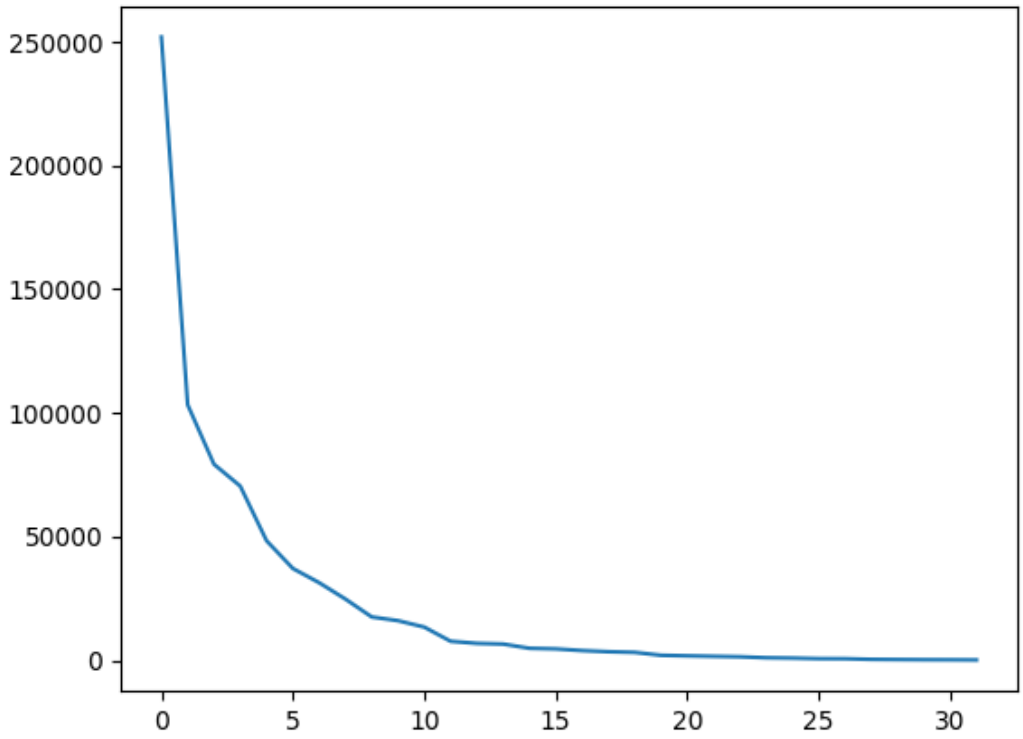
Experimental Observations

2. Principal Component Analysis

Train and test samples are classified into 3 classes. Each class contains samples of particular images. Class 1 contains images of candy store, class 2 contains images of forest broadleaf and class 3 contains images of football stadium. 32 dimesional BoVW representation of these images are taken as inputs.

2.1 Results

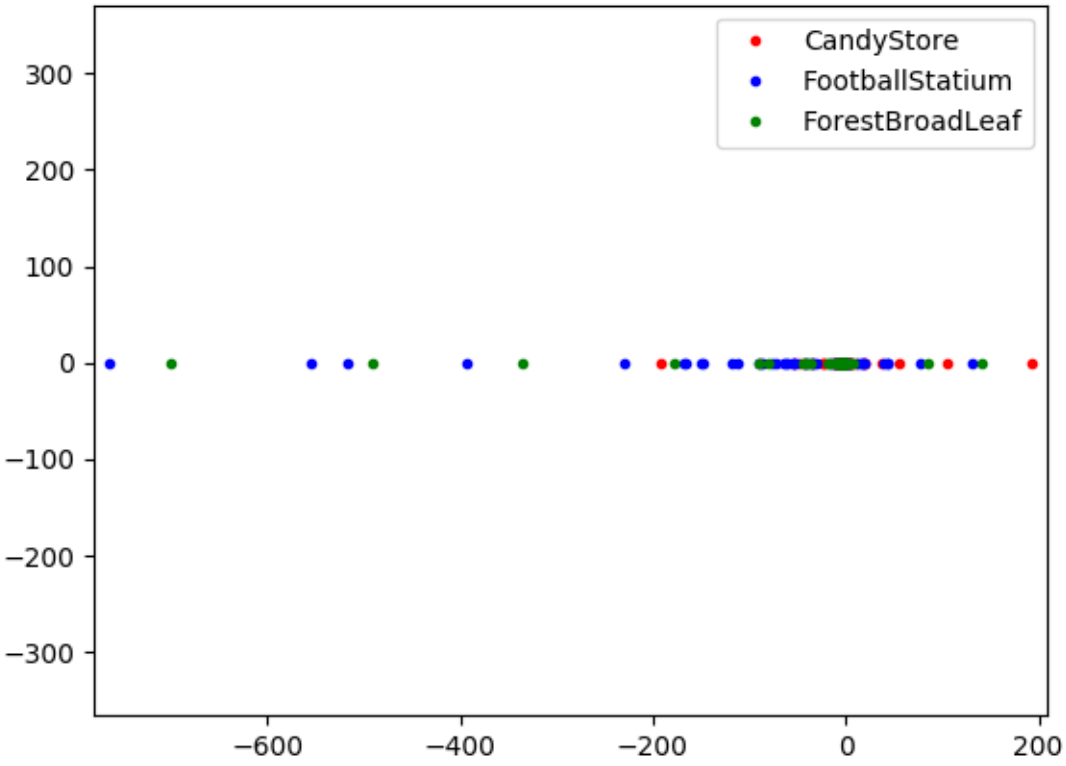
Plot of Eigen Values (Sorted) :-



Results of GMM for different values of l and K.

Class1:- Candy Store
Class2: Football Stadium
Class3: Forest Broadleaf

2.1.1 l = 1



Plot for L=1(Direction of Max Variance)

2.1.1.1 K=1

Number of Data Points = 50
Number of Parameters to be estimated = 2

Classification Accuracy (%)	38
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.362 0.818 0.166
Mean Precision	0.449
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.92 0.18 0.04
Mean Recall	0.38
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.519 0.295 0.064
Mean F-measure	0.293

Confusion Matrix :

$$C = \begin{matrix} 46 & 1 & 3 \\ 34 & 9 & 7 \\ 47 & 1 & 2 \end{matrix}$$

2.1.1.2 K=2

Number of Data Points = 50
Number of Parameters to be estimated = 6

Classification Accuracy (%)	43.3
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.457 0.431 0.422
Mean Precision	0.437
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.32 0.38 0.60
Mean Recall	0.433
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.376 0.404 0.495
Mean F-measure	0.424

Confusion Matrix :

$$C = \begin{matrix} 16 & 13 & 21 \\ 11 & 19 & 20 \\ 8 & 12 & 30 \end{matrix}$$

2.1.1.3 K=4

Number of Data Points = 50
Number of Parameters to be estimated = 12

Classification Accuracy (%)	44
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.413 0.478 0.431
Mean Precision	0.440
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.38 0.44 0.50
Mean Recall	0.44
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.395 0.458 0.463

Mean F-measure	0.480

Confusion Matrix :

$$\mathbf{C} = \begin{matrix} 19 & 14 & 17 \\ 12 & 22 & 16 \\ 15 & 10 & 25 \end{matrix}$$

2.1.1.4 K=8

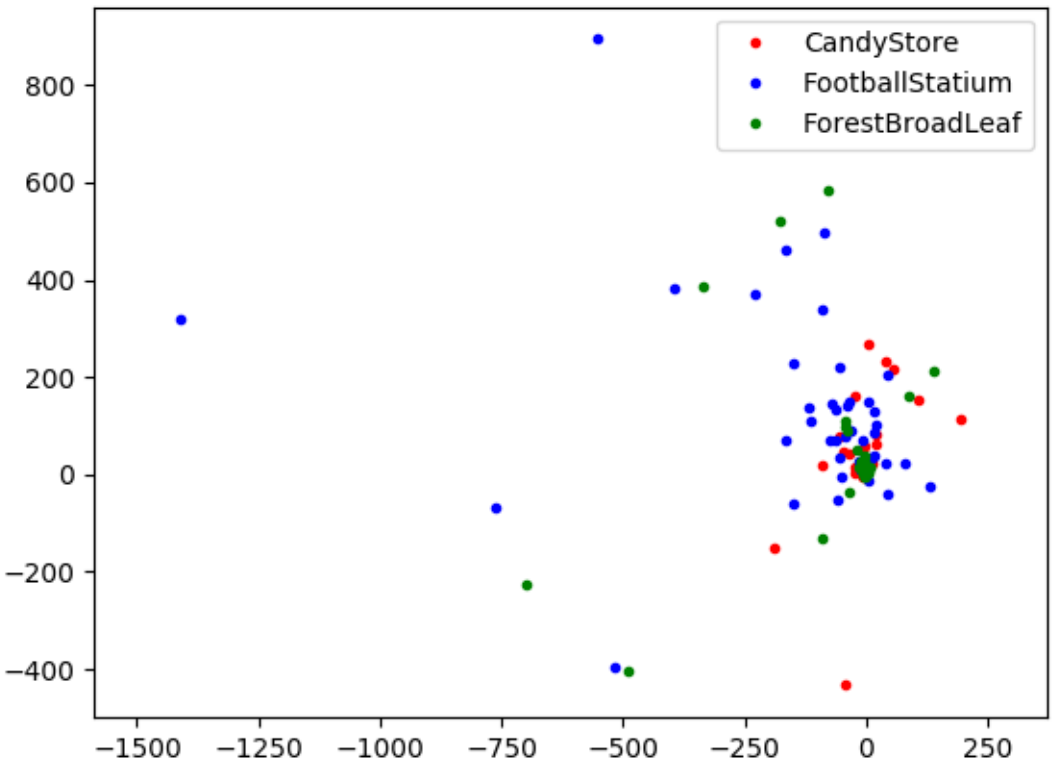
Number of Data Points = 50
Number of Parameters to be estimated = 24

Classification Accuracy (%)	48.667
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.484 0.574 0.428
Mean Precision	0.495
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.32 0.54 0.60
Mean Recall	0.4866
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.385 0.556 0.50
Mean F-measure	0.481

Confusion Matrix :

$$\mathbf{C} = \begin{matrix} 16 & 8 & 26 \\ 9 & 27 & 14 \\ 8 & 12 & 30 \end{matrix}$$

2.1.2 l = 2



Plot for L=2

2.1.2.1 K=1

Number of Data Points = 50
Number of Parameters to be estimated = 5

Classification Accuracy (%)	40.66
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.394 0.608 0.153
Mean Precision	0.385
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.90 0.28 0.04
Mean Recall	0.406
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.548 0.383 0.063
Mean F-measure	0.330

Confusion Matrix :

$$C = \begin{matrix} 45 & 3 & 2 \\ 27 & 14 & 7 \\ 42 & 6 & 2 \end{matrix}$$

2.1.2.2 K=2

Number of Data Points = 50
Number of Parameters to be estimated = 12

Classification Accuracy (%)	43.3
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.398 0.588 0.0
Mean Precision	0.328
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.90 0.40 0.0
Mean Recall	0.433
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.55 0.47 0.0
Mean F-measure	0.34

Confusion Matrix :

$$C = \begin{matrix} 45 & 4 & 1 \\ 28 & 20 & 2 \\ 40 & 10 & 0 \end{matrix}$$

2.1.2.3 K=4

Number of Data Points = 50
Number of Parameters to be estimated = 24

Classification Accuracy (%)	49.33
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.478 0.478 0.517
Mean Precision	0.491
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.44 0.44 0.6
Mean Recall	0.493
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.458 0.458 0.555

Mean F-measure	0.49

Confusion Matrix :

$$C = \begin{matrix} 22 & 15 & 13 \\ 13 & 22 & 15 \\ 11 & 9 & 30 \end{matrix}$$

2.1.2.4 K=8

Number of Data Points = 50
Number of Parameters to be estimated = 48

Classification Accuracy (%)	53.33
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.50 0.545 0.56
Mean Precision	0.535
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.56 0.48 0.56
Mean Recall	0.533
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.528 0.510 0.56
Mean F-measure	0.523

Confusion Matrix :

$$C = \begin{matrix} 28 & 10 & 12 \\ 16 & 24 & 10 \\ 12 & 10 & 28 \end{matrix}$$

2.1.3 l = 3

2.1.3.1 K=1
Number of Data Points = 50
Number of Parameters to be estimated = 9

Classification Accuracy (%)	41.6666666667
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.383 0.652 0.20
Mean Precision	0.413
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.86 0.30 0.06
Mean Recall	0.4066

F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.53 0.41 0.09
Mean F-measure	0.344

Confusion Matrix :

$$C = \begin{matrix} 43 & 3 & 4 \\ 27 & 15 & 8 \\ 42 & 5 & 3 \end{matrix}$$

2.1.3.2 K=2

Number of Data Points = 50
Number of Parameters to be estimated = 20

Classification Accuracy (%)	47.33
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.50 0.515 0.443
Mean Precision	0.486
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.38 0.34 0.7
Mean Recall	0.4733
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.431 0.409 0.542
Mean F-measure	0.46

Confusion Matrix :

$$C = \begin{matrix} 16 & 13 & 21 \\ 11 & 19 & 20 \\ 8 & 12 & 30 \end{matrix}$$

2.1.3.3 K=4

Number of Data Points = 50
Number of Parameters to be estimated = 40

Classification Accuracy (%)	41.33
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.50 0.441 0.382
Mean Precision	0.441
Recall for Class 1 Recall for Class 2	0.22 0.30

Recall for Class 3	0.72
Mean Recall	0.413
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.305 0.357 0.50
Mean F-measure	0.396

Confusion Matrix :

$$\mathbf{C} = \begin{matrix} 11 & 8 & 31 \\ 8 & 5 & 27 \\ 3 & 11 & 36 \end{matrix}$$

2.1.3.4 K=8

Number of Data Points = 50
Number of Parameters to be estimated (Diagonal Cov Matrix)= 56

Classification Accuracy (%)	42
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.464 0.714 0.382
Mean Precision	0.519
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.12 0.2 0.94
Mean Recall	0.42
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.19 0.312 0.5433
Mean F-measure	0.348

Confusion Matrix :

$$\mathbf{C} = \begin{matrix} 6 & 2 & 42 \\ 6 & 10 & 34 \\ 1 & 2 & 47 \end{matrix}$$

2.1.4 l = 4

2.1.4.1 K=1

Number of Data Points = 50
Number of Parameters to be estimated = 14

Classification Accuracy (%)	40
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.3909 0.5517 0.0909

Mean Precision	0.344
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.86 0.32 0.02
Mean Recall	0.4
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.537 0.405 0.032
Mean F-measure	0.384

Confusion Matrix :

$$C = \begin{matrix} 43 & 5 & 2 \\ 26 & 16 & 8 \\ 41 & 8 & 1 \end{matrix}$$

2.1.4.2 K=2

Number of Data Points = 50
Number of Parameters to be estimated = 30

Classification Accuracy (%)	40.667
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.4074 0.421 0.40
Mean Precision	0.409
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.22 0.32 0.68
Mean Recall	0.40667
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.285 0.363 0.503
Mean F-measure	0.384

Confusion Matrix :

$$C = \begin{matrix} 11 & 14 & 25 \\ 8 & 16 & 26 \\ 8 & 8 & 34 \end{matrix}$$

2.1.4.3 K=4

Number of Data Points = 50
Number of Parameters to be estimated(Diagonal Cov Matrix) = 36

Classification Accuracy (%)	47.33

Precision for Class 1 Precision for Class 2 Precision for Class 3	0.456 0.489 0.472
Mean Precision	0.473
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.42 0.48 0.52
Mean Recall	0.4733
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.4375 0.4848 0.495
Mean F-measure	0.472

Confusion Matrix :

$$\mathbf{C} = \begin{matrix} 21 & 14 & 15 \\ 12 & 24 & 12 \\ 13 & 11 & 26 \end{matrix}$$

2.1.4.4 K=8

Number of Data Points = 50
Number of Parameters to be estimated(Diagonal Cov Matrix) = 72

Classification Accuracy (%)	42
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.365 0.0 0.875
Mean Precision	0.4133
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.98 0.0 0.28
Mean Recall	0.42
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.532 0.0 0.424
Mean F-measure	0.318

Confusion Matrix :

$$\mathbf{C} = \begin{matrix} 49 & 0 & 1 \\ 49 & 0 & 1 \\ 36 & 0 & 14 \end{matrix}$$

2.1.5 l = 10

2.1.5.1 K=1

Number of Data Points = 50

Number of Parameters to be estimated = 55

Classification Accuracy (%)	52.66
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.50 0.596 0.484
Mean Precision	0.526
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.34 0.62 0.62
Mean Recall	0.5266
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.404 0.607 0.543
Mean F-measure	0.518

Confusion Matrix :

$$C = \begin{matrix} 17 & 9 & 24 \\ 10 & 31 & 9 \\ 7 & 12 & 31 \end{matrix}$$

2.1.5.2 K=2

Number of Data Points = 50
Number of Parameters to be estimated(Diagonal Cov Matrix) = 42

Classification Accuracy (%)	61.33
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.526 0.666 0.769
Mean Precision	0.654
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.80 0.64 0.40
Mean Recall	0.6133
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.634 0.653 0.526
Mean F-measure	0.604

Confusion Matrix :

$$C = \begin{matrix} 40 & 7 & 3 \\ 15 & 32 & 3 \\ 21 & 9 & 20 \end{matrix}$$

2.1.5.3 K=4

Number of Data Points = 50
Number of Parameters to be estimated(Diagonal Cov Matrix) = 84

Classification Accuracy (%)	53.333
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.667 0.472 0.60
Mean Precision	0.579
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.32 0.86 0.42
Mean Recall	0.533
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.432 0.6099 0.494
Mean F-measure	0.52

Confusion Matrix :

$$\mathbf{C} = \begin{matrix} 16 & 25 & 9 \\ 2 & 43 & 5 \\ 6 & 23 & 21 \end{matrix}$$

2.1.5.4 K=8

Number of Data Points = 50
Number of Parameters to be estimated(Diagonal Cov Matrix) = 150+

Not enough data Points to estimate 150+ parameters.

2.2 Overall Observations:

- Lesser is the value of l, more is the loss of information. If we take l equal to one, then the accuracy is quite less as only one projection direction is considered out of d possible directions in a d dimensional space.
- Transformation is defined in such a way that the first principal component has the largest possible variance and the plot for L=1 is shown.
- The increase in accuracy between successive values of l is directly proportional to the Eigen value of the new direction taken. Since the magnitude of eigenvalues keeps on decreasing so the percentage increase in the accuracy keeps on decreasing with increasing l.
- After a certain l, the value of Eigen values become very less or negligible compared to initial Eigen values then on increasing l no increase in accuracy is obtained.
- For larger values of l, curse of dimensionality becomes prominent.
- For a certain l the accuracy is maximum and on deviating form that value of l the accuracy will decrease.