CS669-Pattern Recognition

Assignment - 4

Bayes Classifier Using GMM On Reduced Dimensions Obtained Using PCA

Group 6

Team :

Akhilesh Devrari (B16005)

Aman Jain (B16044)

Yash Agrawal (B16120)

Table of Contents :

1. Objective :

1.1. Build Bayes classifier using Gaussian mixture model (GMM) with 1, 2, 4 and 8 mixtures on the reduced dimensional representations of Dataset-2 obtained using PCA.

1.2. Find confusion matrix for each case (for different datasets).

1.3. Calculate classification accuracy, precision for every class, mean precision, recall for every class, mean recall, F-measure for every class and mean F-measure.

1.4. Plot density contour for each class with data points superposed.

1.5. Plot decision region for each pair of classes and together for all classes.

2. Procedure :

2.1. Separate given data files of each class into training and test data files. Training data consists of 75% and test data consists of 25% of given data set.

2.2. To apply Bayesian classification, assume that data of all classes -follow Gaussian distribution.

2.3. Convert all the image data given in dataset 2b into 32 dimension BOVW (Bag of Visual Words) representation.

2.4. KMeans and GMM is already applied on this data in assignment 2.

2.5. Convert this 32-Dimensional data into l-dimensional data where l is the reduced dimension of the given data.

2.6. Use principal component analysis (PCA) to find the reduced dimensions of the given data.

2.7. Eigen values and eigen vectors are calculated to find the values of lambda and transformation vector.

2.8. Transformation vector is the vector which is when imposed on a d-dimensional data will convert the data into l-dimensional data.

2.9. Gaussian Mixture Model (GMM) and KMeans are applied on the corresponding reduced dimensional data.

3. Procedure For finding Principal component analysis (PCA) :

3.1. Suppose x be the given datapoint which is a high dimension (d) given data and a be the datapoint which is converted into some reduced dimension l.

3.2. Now find transformation vector (directional) which will be an orthonormal vector.

3.3. Consider Y be the mean subtracted representation of x which is the given datapoint.

3.4. Reduced dimensional datapoint will be calculated by using the multiplication of transpose of transformation vector and mean subtracted representation of datapoint.

3.5. It involve finding l directions of projection in such a way that it minimizes the error between the original representation and the approximated representation and summed over all the training examples.

3.6. With a constraint that each direction of projection should be orthonormal and the error is denoted by J.

3.7. After soving for J we get a constrained optimization problem to find the minimum of J with a constraint that each direction of projection should be orthonormal.

3.8. Lagrangian method is used to solve this constrained optimization problem.

3.9. After solving using lagrangian we get corresponding eigen vectors and corresponding eigen values.

4. Observations :

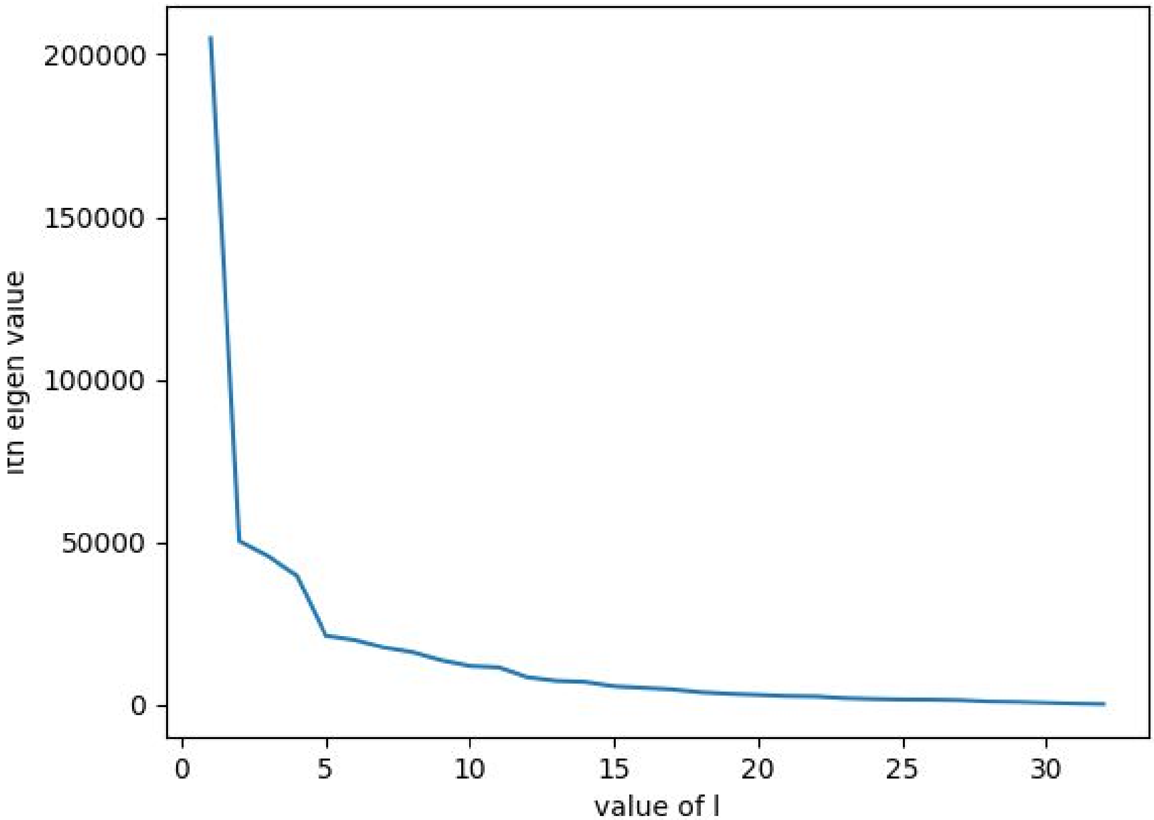

Fig : 1 Plot of Eigen Values in decreasing order

32 Eigen Values in decending order are :

204911.42371983654, 50274.91291991595, 45639.69295370778, 39675.362010630124, 21268.430981828078, 19931.34138831108, 17634.54680157979, 16124.52098115722, 13660.63758481213, 11895.450541444494, 11460.41927652893, 8453.085282423832, 7352.235022807325, 7027.20930673038, 5704.391257423952, 5232.30867694192, 4760.287042049027, 3879.8117750317324, 3409.3178648779353, 3077.767521890299, 2722.0924997674806, 2597.8594734935054, 2073.458682076009, 1830.7543815684292, 1663.6154336300363, 1592.2603827135706, 1436.4033748555703, 985.1642662749077, 853.9220119100451, 600.1328490193212, 350.0996571833756, 183.31203744581555

a) K = 1, L = 1: Accuracy : 62.6666666667 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	9	2	39
Class 2	1	41	8
Class 3	5	1	44

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.18	0.82	0.88	0.6266
Precision	0.6	0.9318	0.4835	0.8448
F-Measure	0.2769	0.8723	0.6241	0.5911

b) K = 1, L = 2: Accuracy : 38.6666666667 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	16	0	34
Class 2	2	1	47
Class 3	9	0	41

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.32	0.02	0.82	0.3866
Precision	0.5925	1.0	0.3360	0.6428
F-Measure	0.4155	0.0392	0.4767	0.3105

c) K = 1, L=3: Accuracy : 35.3333333333 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	11	0	39
Class 2	1	2	47
Class 3	10	0	40

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.22	0.04	0.8	0.3533
Precision	0.5	1.0	0.3174	0.6058
F-Measure	0.3055	0.0769	0.4545	0.2790

d) K = 1, L=4: Accuracy : 34.6666666667 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	8	0	42
Class 2	0	0	50
Class 3	6	0	44

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.16	0.0	0.88	0.3466
Precision	0.5714	nan	0.3235	nan
F-Measure	0.25	nan	0.4731	nan

e) K = 1, L=5: Accuracy : 33.3333333333 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	0	0	50
Class 2	0	0	50
Class 3	0	0	50

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.0	0.0	1.0	0.3333
Precision	nan	nan	0.3333	nan
F-Measure	nan	nan	0.5	nan

f) K = 2, L = 1: Accuracy : 56.6666666667 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	13	2	35
Class 2	2	28	20
Class 3	5	1	44

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.26	0.56	0.88	0.5666
Precision	0.65	0.9032	0.4444	0.6658
F-Measure	0.3714	0.6913	0.5906	0.5511

g) K = 2, L = 2: Accuracy : 50.6666666667 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	29	0	21
Class 2	1	1	48
Class 3	3	1	46

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.58	0.02	0.92	0.5066
Precision	0.8787	0.5	0.4	0.5929
F-Measure	0.6987	0.0384	0.5575	0.4316

h) K = 2, L = 3: Accuracy : 44.6666666667 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	20	1	29
Class 2	1	2	47
Class 3	5	0	45

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.4	0.04	0.9	0.4466
Precision	0.7692	0.6666	0.3719	0.6025
F-Measure	0.5263	0.0754	0.5263	0.3760

i) K = 2, L = 4: Accuracy : 38.0 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	10	1	39
Class 2	0	0	50
Class 3	3	0	47

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.2	0.0	0.94	0.38
Precision	0.7692	0.0	0.3455	0.3716
F-Measure	0.3174	0.0	0.5053	0.2742

j) K = 2, L = 5: Accuracy : 44.6666666667 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	15	0	35
Class 2	0	2	48
Class 3	0	0	50

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.3	0.04	1.0	0.4466
Precision	1.0	1.0	0.3759	0.7919
F-Measure	0.4615	0.0769	0.5464	0.3616

k) K = 4, L = 1: Accuracy : 75.3333333333 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	34	2	14
Class 2	1	37	12
Class 3	7	1	42

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.68	0.74	0.84	0.7533
Precision	0.8095	0.925	0.6176	0.7840
F-Measure	0.7391	0.8222	0.7118	0.7577

l) K = 4, L = 2: Accuracy : 52.6666666667 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	38	0	12
Class 2	1	0	49
Class 3	9	0	41

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.76	0.0	0.82	0.5266
Precision	0.7916	nan	0.4019	nan
F-Measure	0.7755	nan	0.5394	nan

m) K = 4, L = 3: Accuracy : 49.3333333333 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	30	0	20
Class 2	0	1	49
Class 3	7	0	43

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.6	0.02	0.86	0.4933
Precision	0.8108	1.0	0.3839	0.7315
F-Measure	0.6896	0.0392	0.5308	0.4199

n) K = 4, L = 4: Accuracy : 42.0 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	21	0	29
Class 2	1	1	48
Class 3	9	0	41

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.42	0.02	0.82	0.42
Precision	0.6774	1.0	0.3474	0.6749
F-Measure	0.5185	0.0392	0.4880	0.3486

o) K = 4, L = 5: Accuracy : 36.6666666667 %

Confusion Matrix

	Class 1	**Class 2**	**Class 3**
Class 1	4	0	46
Class 2	0	1	49
Class 3	0	0	50

Analysis

	Class 1	**Class 2**	**Class 3**	**Mean**
Recall	0.08	0.02	1.0	0.3666
Precision	1.0	1.0	0.3448	0.7816
F-Measure	0.1481	0.0392	0.5128	0.2333

p) K = 8, L = 1: Accuracy : 71.3333333333 %

Confusion Matrix

	Class 1	**Class 2**	**Class 3**
Class 1	40	2	8
Class 2	1	35	14
Class 3	17	1	32

Analysis

	Class 1	**Class 2**	**Class 3**	**Mean**
Recall	0.8	0.7	0.64	0.7133
Precision	0.6896	0.9210	0.5925	0.7344
F-Measure	0.7407	0.7954	0.6153	0.7171

q) K = 8, L = 2: Accuracy : 33.3333333333 %

Confusion Matrix

	Class 1	**Class 2**	**Class 3**
Class 1	0	0	50
Class 2	0	0	50
Class 3	0	0	50

Analysis

	Class 1	**Class 2**	**Class 3**	**Mean**
Recall	0.0	0.0	1.0	0.3333
Precision	nan	nan	0.3333	nan
F-Measure	nan	nan	0.5	nan

r) K = 8, L = 3: Accuracy : 36.0 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	40	0	10
Class 2	2	1	47
Class 3	37	0	13

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.8	0.02	0.26	0.36
Precision	0.5063	1.0	0.1857	0.5640
F-Measure	0.6201	0.0392	0.2166	0.2920

s) K = 8, L = 4: Accuracy : 33.3333333333 %

Confusion Matrix

	Class 1	Class 2	Class 3
Class 1	0	0	50
Class 2	0	0	50
Class 3	0	0	50

Analysis

	Class 1	Class 2	Class 3	Mean
Recall	0.0	0.0	1.0	0.3333
Precision	nan	nan	0.3333	nan
F-Measure	nan	nan	0.5	nan

5. Conclusion

1) From the above data, it can be seen that as we increase the reduced dimension(l) accuracy decreases in most of the cases.
2) In some cases, value of precision and recall is nan because zero number of test examples are classified in one or more classes.
3) For larger values of K, results are not that good because chances of only one data point in each cluster increases due to which the determinant becomes zero and nan values occurs in covariance matrix.
4) It is observed from above results that most of the test examples are classified in class 3.
5) It can be observed from the plot of eigen values that after certain no. of dimensions difference between two consecutive eigen values becomes very less.