Pattern Recognition

Assignment-4

Group: 10

November , 2018

Team members:

1.Randheer kumar (B16139)

2.Dhrubodeep Basumatary (B16017)

3.Dilip Kumar Chauhan (B16018)

Contents:

1.Objective:

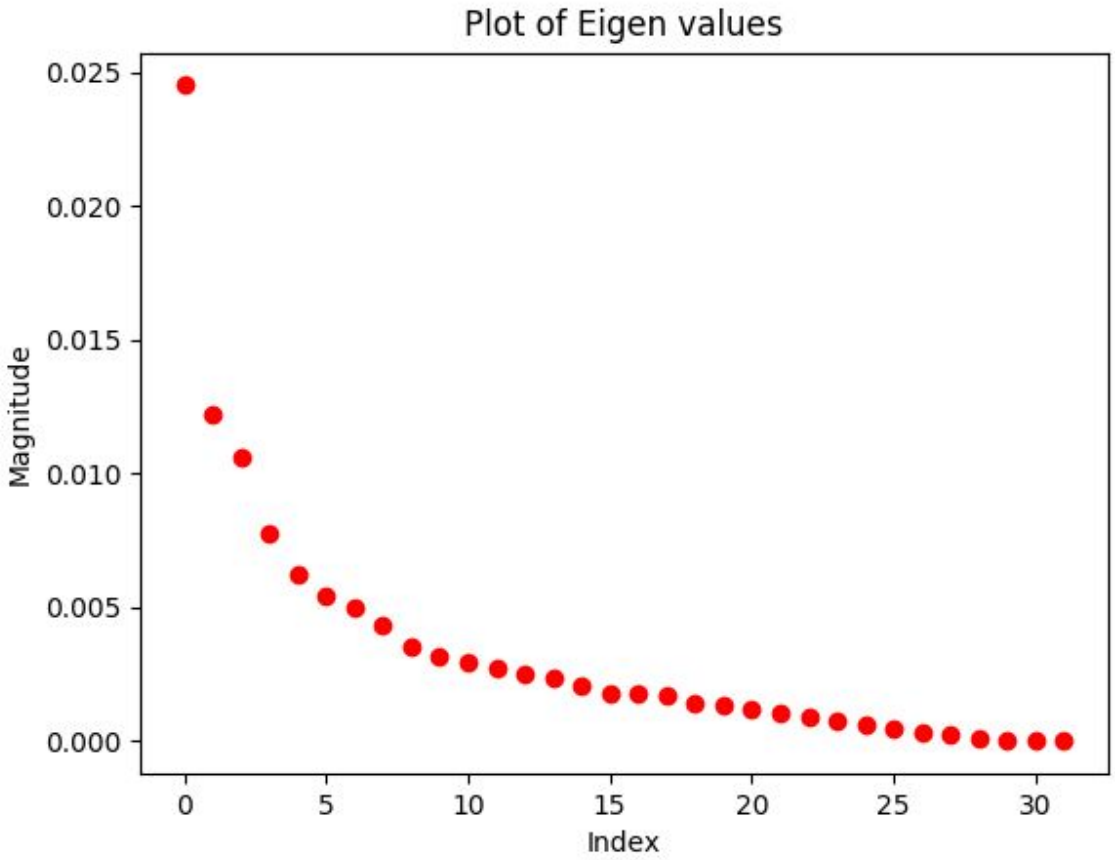
- To Build Bayes classifier using Gaussian mixture model (GMM) with 1, 2, 4 and 8 mixtures on the reduced dimensional representation of BOVW of image dataset obtained using PCA.
  - ❖ Calculate Accuracy, Precision,Recall, mean Precision,mean Recall,F-measure and Confusion Matrix

2.Procedure:

- Consider the 32D BoVW representation of image dataset.
- Apply Principal Component Analysis(PCA) on the above Dataset for different number of Principal Components(PC).
    - ❖ Take all training examples(32D BoVW representation) of all the classes and compute the covariance matrix.
    - ❖ Perform eigen analysis on covariance matrix to obtain 32 eigen values and corresponding eigen vectors.
    - ❖ Choose L eigen vectors corresponding to L leading eigen values(significant eigen values).
    - ❖ Project each y(mean subtracted representation of original data) on to the L eigen vectors to get L principal components.
    - ❖ Now the data has transformed into L dimensional feature space.
- Apply the Bayes Classifier for different number of mixtures(k=1,2,4 and 8) and corresponding Confusion Matrix is calculated.
- Accuracy, Precision,Recall, mean Precision,mean Recall,F-measure is calculated corresponding to each Confusion Matrix.

3.observation:

3.1 Bayes classifier using PCA and GMM



3.1.1 k=1

3.1.1.1 L=1

	movie	rock	valley
movie	24	4	22
rock	2	5	43
valley	7	5	38

	movie	rock	valley	mean
precision	0.72	0.35	0.368	0.48
recall	0.48	0.1	0.76	0.44
F-measure	0.578	0.15	0.49	0.464

Accuracy: 44.56%

3.1.1.2 L=2

	movie	rock	valley
kA	29	0	1
kha	3	5	42
khA	11	5	34

	movie	rock	valley	mean
precision	0.67	0.5	0.35	0.50
recall	0.58	0.1	0.68	0.45
F-measure	0.62	0.16	0.462	0.479

Accuracy: 45.3%

3.1.1.3 L = 4

	movie	rock	valley
kA	28	11	11
kha	8	14	28
khA	9	15	26

	movie	rock	valley	mean
precision	0.62	0.35	0.4	0.45
recall	0.56	0.28	0.52	0.453
F-measure	0.589	0.311	0.452	0.45

Accuracy: 45.3%

3.1.1.4 L=8

	movie	rock	valley
kA	36	11	3
kha	8	23	19
khA	10	15	25

	movie	rock	valley	mean
precision	0.66	0.46	0.53	0.55
recall	0.72	0.46	0.5	0.56
F-measure	0.69	0.46	0.51	0.55

Accuracy:56%

3.1.1.5 l=16

	movie	rock	valley
kA	34	9	7
kha	9	20	21
khA	9	15	26

	movie	rock	valley	mean
precision	0.65	0.45	0.48	0.52
recall	0.68	0.4	0.52	0.53
F-measure	0.66	0.42	0.5	0.53

Accuracy: 53.3%

Inferences :

- Generally accuracy increases with increase in dimension of projected data .
- Here for Principal Component (PC) equal to 8, maximum accuracy is achieved
- We are not getting good accuracy because data-set is projected considering only eigen vectors corresponding to significant eigen values so there is possibility of loosing discriminative information .
- Accuracy decreases after L=8 because some unnecessary information might have included.

3.1.2 K=2

3.1.2.1 L=1

	movie	rock	valley
kA	28	11	11
kha	7	14	29
khA	10	11	29

	movie	rock	valley	mean
precision	0.62	0.38	0.42	0.47
recall	0.56	0.2	0.58	0.47
F-measure	0.58	0.32	0.48	0.48

Accuracy : 47.3%

3.1.2.2 L=2

	movie	rock	valley
kA	22	22	6

kha	0	25	25
khA	7	13	30

	movie	**rock**	**valley**	**mean**
precision	0.75	0.41	0.49	0.55
recall	0.44	0.5	0.6	0.51
F-measure	0.55	0.45	0.54	0.53

Accuracy : 51.3%

3.1.2.3 L=4

	movie	**rock**	**valley**
kA	28	12	10
kha	5	24	21
khA	10	12	28

	movie	**rock**	**valley**	**mean**
precision	0.65	0.5	0.47	0.54
recall	0.56	0.48	0.56	0.533

F-measure	0.60	0.48	0.51	0.53

Accuracy : 53.3%

3.1.2.4 L=8

	movie	rock	valley
kA	33	13	4
kha	9	14	27
khA	6	16	28

	movie	rock	valley	mean
precision	0.68	0.32	0.47	0.49
recall	0.66	0.28	0.56	0.5
f-measure	0.67	0.30	0.51	0.49

Accuracy :50%

3.1.2.5 L=16

	movie	rock	valley
Movie	40	7	3

rock	10	19	21
valley	9	19	25

	movie	**rock**	**valley**	**mean**
precision	0.71	0.42	0.51	0.54
recall	0.8	0.38	0.5	0.56
F-measure	0.74	0.4	0.54	0.55

Accuracy : 56%

3.1.2.6 Inferences :

- We are getting more accuracy for k=2 as compare to k=1 . This is possibly due to better fitting of data into two gaussian distribution.
- Here the maximum accuracy is achieved for L=16 out of different dimensions considered.
- We are getting almost same accuracy in L=4,8,16 .This is due to absence of discriminative information in all the cases . This discriminative information can be obtained if we further increase the value of L .

3.1.3 k=4

3.1.3.1 L=1

	movie	**rock**	**valley**
Movie	27	12	11

rock	11	10	29
valley	12	7	31

	movie	**rock**	**valley**	**mean**
precision	0.54	0.34	0.43	0.44
recall	0.54	0.2	0.62	0.45
F-measure	0.54	0.25	0.51	0.44

Accuracy : 45.3%

3.1.3.2 L=2

	movie	**rock**	**valley**
Movie	25	14	11
rock	1	23	26
valley	9	16	25

	movie	**rock**	**valley**	**mean**
precision	0.71	0.43	0.4	0.51
recall	0.5	0.46	0.5	0.48
F-measure	0.58	0.44	0.44	0.50

Accuracy : 48.6%

3.1.3.3 L=4

	movie	rock	valley
Movie	34	8	8
rock	10	18	22
valley	10	18	22

	movie	rock	valley	mean
precision	0.62	0.40	0.42	0.48
recall	0.68	0.36	0.44	0.49
F-measure	0.65	0.38	0.43	0.49

Accuracy : 49.3%

3.1.3.4 L=8

	movie	rock	valley
Movie	27	13	10
rock	10	20	20

valley	8	12	30

	movie	**rock**	**valley**	**mean**
precision	0.6	0.4	0.5	0.514
recall	0.54	0.4	0.6	0.513
F-measure	0.56	0.42	0.54	0.513

Accuracy : 51.3%

3.1.3.5 L=16

	movie	**rock**	**valley**
movie	39	6	5
rock	12	22	16
valley	6	13	31

	movie	**rock**	**valley**	**mean**
precision	0.64	0.53	0.59	0.60
recall	0.78	0.44	0.62	0.61
F-measure	0.72	0.48	0.60	0.60

Accuracy : 61.3%

3.1.3.6 Inference :

- We are getting more accuracy for k=4 as compare to k=2 . This is due to better fitting of data in mixture of 4-Gaussian distribution as compare to 2-Gaussian distribution .
- Here accuracy is increasing with increase in number of principal component (L).

3.1.4 k=8

3.1.4.1 L=1

	movie	**rock**	**valley**
Movie	26	12	12
rock	11	18	21
valley	18	12	20

	movie	**rock**	**valley**	**mean**
precision	0.47	0.42	0.37	0.42
recall	0.52	0.36	0.4	0.42
F-measure	0.49	0.39	0.388	0.426

Accuracy : 42.6%

3.1.4.2 L=2

	movie	rock	valley
kA	25	15	10
kha	3	19	28
khA	10	19	21

	movie	rock	valley	mean
precision	0.65	0.35	0.35	0.45
recall	0.5	0.38	0.42	0.43
F-measure	0.56	0.36	0.38	0.44

Accuracy : 43.3%

3.1.4.3 L=4

	movie	rock	valley
kA	32	12	6
kha	8	23	19

khA	15	9	26

	movie	**rock**	**valley**	**mean**
precision	0.65	0.35	0.355	0.45
recall	0.5	0.38	0.42	0.433
F-measure	0.56	0.36	0.38	0.44

Accuracy : 43.3%

3.1.4.4 L=8

	movie	**rock**	**valley**
kA	32	6	12
kha	6	18	26
khA	11	16	23

	movie	**rock**	**valley**	**mean**
precision	0.65	0.45	0.37	0.49
recall	0.64	0.36	0.46	0.48
F-measure	0.64	0.4	0.41	0.48

Accuracy : 48.6%

3.1.4.5 L=16

	movie	rock	valley
kA	30	16	4
kha	3	22	25
khA	5	23	22

	movie	rock	valley	mean
precision	0.78	0.36	0.43	0.52
recall	0.6	0.44	0.44	0.49
F-measure	0.68	0.39	0.43	0.50

Accuracy : 49.3%

3.1.4.6 Inference

- Here increasing the number of Principal components accuracy is increasing and maximum accuracy is found for L=16.
- We are getting poor accuracy due to overlapping of gaussian distribution of one with gaussian distribution of other class .
- Here accuracy is lesser than accuracy of k=4 .

4.conclusion:

- As we can see from the plot of eigen values only few eigen values are significant, others are close to zeros(less significant).
- The accuracy is almost same when data was projected on 8D considering eigen vectors corresponding to 8 leading eigen values and for the original 32D data.
- Accuracy corresponding to smaller number of Principal Components is smaller than that of higher number of Principal components.
- Here the accuracy has decreased for Gaussian mixtures equal to 8 than mixtures equal to 4 due to overlapping of Gaussian mixture of one class to the Gaussian mixture of other class.
- For the given data set accuracy is maximum when the number of Gaussian mixtures is 4 because the data of each class are best fit in 4 mixtures Gaussian.