Pattern Recognition

Assignment-3

Group: 10

October 21 , 2018

Team members:

1.Randheer kumar (B16139)

2.Dhrubodeep Basumatary (B16017)

3.Dilip Kumar Chauhan (B16018)

Contents:

1.Objective:

1 .Bayes classifier using K-nearest neighbour method for class-conditional density estimation using DTW distance.

- Calculate Accuracy, Precision,Recall, mean Precision,mean Recall,F-measure and Confusion Matrix

2 . Bayes classifier using Discrete HMM (DHMM)

- Calculate Accuracy, Precision,Recall, mean Precision,mean Recall,F-measure and Confusion Matrix

2.Procedure:

1.Bayes classifier using KNN method

- Take a test example and calculate the DTW distance with all the examples of all classes.
- Sort the training examples in increasing order of the DTW distance to the test example.
- Choose first K(K=4,8,16,32) examples from the sorted list as K-neighbours.
- Whichever Class(out of ‘kA’,’kha’,’khA’) has the maximum frequency out of K,the label of the test example is assigned as that class.
- Based on the above classification, Confusion matrix is built.

2.Bayes Classifier using Discrete HMM(DHMM).

- All the 39-D vectors of all the examples of all classes are clustered into 8,16 and 32 clusters using K-Means clustering.
- Take a 39-D vector of an observation and calculated the euclidean distance to all the cluster centres and from whichever cluster it has minimum distance ,that cluster number is assigned to this 39-D vector.
- Repeating the step-2 for all the 39-D vector of all the observation of all the classes from training and testing set,we quantised the 39-D vector.
- Considering the Left to Right DHMM ,Bayes classifier using DHMM is built for different combination of observation symbols or code blocks (M=8,16,32) and states(N=2,3,4,5) and corresponding Confusion Matrix is built .
- Based on Confusion Matrix,Accuracy,Recall,Precision,mean Recall ,mean Precision and F-measure is calculated.

3.observation:

3.1 Bayes classifier using KNN method

3.1.1 k=4

	kA	kha	khA
kA	126	1	0
kha	12	3	0
khA	12	0	8

	kA	kha	khA	mean
precision	0.84	0.75	1	0.86
recall	0.99	0.2	0.4	0.53
F-measure	0.91	0.32	0.57	0.66

Accuracy: 84.56%

3.1.2 k=8

	kA	kha	khA
kA	127	0	0
kha	10	3	2
khA	15	0	5

	kA	kha	khA	mean
precision	0.84	1.	0.71	0.85
recall	1	0.2	0.25	0.48
F-measure	0.91	0.33	0.37	0.62

Accuracy: 83.3%

3.1.3 k=16

	kA	kha	khA
kA	127	0	0
kha	13	1	1
khA	17	0	3

	kA	kha	khA	mean
precision	0.81	1.	0.75	0.85
recall	1.	0.067	0.15	0.46
F-measure	0.89	0.125	0.25	0.55

Accuracy: 80.2%

3.1.4 k=32

	kA	kha	khA
kA	127	0	0
kha	15	0	0
khA	20	0	0

	kA	kha	khA	mean
precision	0.81	1.	0.75	0.85
recall	1.	0.067	0.15	0.41
F-measure	0.89	0.125	0.25	0.55

Accuracy: 78.3%

3.1.5 Inference

- Here the accuracy decreases as the number of nearest neighbours(K) increases.
- Since the training examples of class ‘kA’ is very higher than that of other two classes and the training examples of all classes are very similar ,so when we increase the number of nearest neighbour,the test examples is mostly classified as one class only (i.e class ‘kA’).
- Here the precision of class ‘kA’ is very high and the reason is same as above(i.e very higher number of examples of this class as compare to other classes).
- For K=32 ,all the test examples of all the classes are classified as class ‘kA’ because more examples of class ‘kA’ are nearest(with respect to DTW distance) to test examples than the training examples of other two classes.

3.2 Bayes Classifier using Discrete HMM(DHMM):

3.2.1 M=8:

3.2.1.1 N=2

	kA	kha	khA
kA	44	41	42
kha	1	10	4
khA	6	8	6

	kA	kha	khA	mean
precision	0.862	0.169	0.115	0.38
recall	0.346	0.667	0.3	0.43
F-measure	0.494	0.270	0.166	0.408

Accuracy: 0.37

3.2.1.2 N=3

	kA	kha	khA
kA	65	22	40
kha	7	4	4
khA	7	6	7

	kA	kha	khA	mean
precision	0.85	0.125	0.137	0.37
recall	0.51	0.33	0.35	0.398
F-measure	0.64	0.18	0.197	0.385

Accuracy :0.47

3.2.1.3 N=4

	kA	kha	khA
kA	75	14	38
kha	7	4	4
khA	7	5	8

	kA	kha	khA	mean
precision	0.84	0.17	0.16	0.39
recall	0.59	0.27	0.4	0.42
F-measure	0.69	0.21	0.23	0.41

Accuracy : 0.54

3.2.1.4 N=5

	kA	kha	khA

kA	75	14	38
kha	7	4	4
khA	7	5	8

	kA	**kha**	**khA**	**mean**
precision	0.84	0.17	0.16	0.39
recall	0.59	0.27	0.4	0.42
F-measure	0.69	0.21	0.23	0.41

Accuracy : 0.54

3.2.1.5 Inference

- As we are increasing the number of states(N) accuracy is increasing.
- Here accuracy is very poor because the examples of all the classes are very similar .
- Here we can see that the precision of ‘kA’ class is very high because the number of training examples in this class is very high(around 8 times) as compare to other two classes.
- Number of examples of the class ‘kha’ and ‘khA’ is very small ,so the parameters of both of the classes are estimated very poorly.

3.2.2 M=16:

3.2.2.1 N=2

	kA	kha	khA
kA	47	36	45
kha	7	3	5
khA	6	6	8

	kA	kha	khA	mean
precision	0.78	0.07	0.14	0.33
recall	0.37	0.2	0.4	0.32
F-measure	0.5	0.1	0.21	0.325

Accuracy : 0.36

3.2.2.2 N=3

	kA	kha	khA
kA	53	41	43
kha	8	3	4
khA	8	2	10

	kA	kha	khA	mean
precision	0.760	0.06	0.17	0.34
recall	0.387	0.2	0.5	0.36
f-measure	0.515	0.098	0.26	0.34

Accuracy :0.384

3.2.2.3 N=4

	KA	kha	khA
KA	64	30	36
kha	6	3	6
khA	8	2	10

	kA	kha	khA	mean
precision	0.82	0.08	0.19	0.366
recall	0.49	0.2	0.5	0.397
F-measure	0.6	0.12	0.278	0.38

Accuracy : 0.46

3.2.2.4 N=5

	kA	kha	khA
kA	62	26	39
kha	5	4	6
khA	9	3	8

	kA	kha	khA	mean
precision	0.82	0.12	0.18	0.374
recall	0.53	0.27	0.4	0.399
F-measure	0.67	0.17	0.25	0.3862

Accuracy : 0.487

3.2.2.5 Inference

- For observation symbol equal to 16 ,increasing the number of states accuracy also increases.
- Here we can see that the accuracy is maximum for N=5 which means that number of states higher than or equal to 5 best describe the examples of training set.

3.2.3 M=32:

3.2.3.1 N=2

	kA	kha	khA
kA	62	19	46
kha	9	3	3
khA	8	3	9

	kA	kha	khA	mean
precision	0.784	0.12	0.155	0.353
recall	0.49	0.2	0.45	0.379
F-measure	0.60	0.15	0.23	0.366

Accuracy : 0.457

3.2.3.2 N=3

	kA	kha	khA
kA	62	18	41
kha	9	2	4
khA	11	1	8

	kA	kha	khA	mean
precision	0.77	0.095	0.151	0.334
recall	0.512	0.133	0.4	0.348
F-measure	0.611	0.11	0.219	0.341

Accuracy :0.46

3.2.3.3 N=4

	kA	kha	khA
kA	76	14	37
kha	5	2	8
khA	9	1	10

	kA	kha	khA	mean
precision	0.844	0.118	0.182	0.38
recall	0.598	0.133	0.5	0.411
F-measure	0.700	0.125	0.267	0.395

Accuracy : 0.54

3.2.3.4 N=5

	kA	kha	khA
kA	79	13	35
kha	6	2	7
khA	8	3	9

	kA	kha	khA	mean
precision	0.85	0.11	0.177	0.379
recall	0.622	0.133	0.45	0.40
F-measure	0.72	0.121	0.25	0.39

Accuracy : 0.56

3.2.4 Inference

- Here also increasing number of states increases the accuracy
- Out of different combination of code blocks and number of states highest accuracy(i.e 56%) has been achieved in the case of M=32 and N=5 which means that this combination of code blocks and number of states out of all combinations which we have considered here, best describe the parameters of each class .

4.conclusion:

- Even the KNN classifier is the best way to classify(in terms of accuracy) but if the data of classes overlap it gives poor accuracy as we have obtained in KNN classifier.
- we know that increasing the number of nearest neighbors first increases the accuracy and then achieves the maximum accuracy for a particular K(depends on data) and after that accuracy decreases and this maximum point depend on the distribution of data of all classes.
- When we have classified the data corresponding to consonant vowel utterances ,for nearest neighbor equal to 4 best accuracy is achieved.
- For all the combination of code blocks(i.e M=8,16,32),number of states(N) equal to 5 gave the best accuracy and for this data set the utterances of consonant and vowel is uttered using 5 events(i.e number of states).
- For code blocks(M) equal to 32, highest accuracy is achieved which means that descritising this data set into 32 symbols results into best accuracy.
- Since the training data of ‘kA’ class is very high than that of ‘kha’ and ‘khA’ class ,so the precision of ‘kA class is very high’.