PATTERN RECOGNITION ASSIGNMENT-3
REPORT

07h November 2018

GROUP - 11

ABHIGYAN KHAUND	PRATYUSH GAURAV	SHASHWAT GARG
B16082	B16026	B16034

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

Table of Contents

1. Introduction

K-Nearest Neighbors:

The K-Nearest Neighbor method is a non-parametric classification method in pattern recognition.

Algorithm: First we find out the distance of a test sample from all the train sample we have. From these distances the smallest k distances are chosen. Each train sample has a class label associated with it. for the nearest k samples chosen, the most frequent class label is also assigned to the test sample, classifying it into that class. We have used Dynamic Time Warping(DTW) algorithm to calculate distances in this assignment. DTW is a dynamic programming algorithm which is used to measure the similarity between two sequences which may vary in length, speed etc. Both the time and space complexity of this algorithm are of O(nm), where n and m are the lengths of sequences. The value of k also affects the result. for a very large k, all the K samples contribute equally to the classification, hence even the train samples far away from the test sample, but in the region bounded by smallest hypersphere enveloping all the k points, will affect the classification. If we take k equal to the number of train samples, all the test samples gets classified into class with maximum train samples in it. Whereas for a very small k, the effect of noise will be high. For example, if a test sample belonging to class A is present near some samples of class B, It will get classified into class B, though it belongs to class A. This is due to the very small number of nearest neighbors affecting the classification.

Discrete Hidden Markov Model:

A Markov process is a process in which the future is independent of the past given the present. That is the future state depends only on the current state and not on the previous states. Hidden Markov model is a Markov model in which the system being modeled is assumed to be Markov process with hidden/unknown states.

Data Set :

The type of data consist of speech data divided into 3 classes – ka, kha, kA. In each class the test samples and train samples consists of many 39 dimensional feature vectors.

Feature Extraction:-

Speech recognition consists of two main modules, feature extraction and feature matching. The purpose of feature extraction module is to convert speech waveform to some type of representation for further analysis and processing, this extracted information is known as feature vector.

Commonly used techniques for feature extractions-
- MFCC (Mel-Frequency Cepstrum Coefficient)
- LPC (Linear Predictive Coding)

In the speech dataset, feature is extracted using MFCC

MFCCs are commonly derived as follows:

1. Take the Fourier transform of (a windowed excerpt of) a signal.

2. Map the powers of the spectrum obtained above onto the mel scale, using triangular overlapping windows.
3. Take the logs of the powers at each of the mel frequencies.
4. Take the discrete cosine transform of the list of mel log powers, as if it were a signal.
5. The MFCCs are the amplitudes of the resulting spectrum.

2. Vector Quantization

The Observation Sequence consists of different continuous valued observations, each of 39 dimensions for speech. In Vector Quantization,M no of clusters are formed using k-means clustering. Then each observation of a certain Observation Sequence is assigned a cluster. Thus, each observation is represented using only one value representing its cluster. Thus, the number of possible Observations becomes finite and equal to M.

Experimental Observations

3. K-Nearest Neighbors Method

Train and test samples are classified into 3 classes. Each class contains samples of particular sounds. Class 1 contains files which have feature vectors representing sound ’ka’, class 2 contains files which have feature vectors representing sound ’kha’ and class 3 contains files which have feature vectors representing sound ’kA’.

3.1 Results

3.1.1 K = 4

Classification Accuracy (%)	74.37
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.782 0.750 0.724
Mean Precision	0.752
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.635 0.2 0.889
Mean Recall	0.575
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.701 0.315 0.798
Mean F-measure	0.605

Confusion Matrix :

$$C = \begin{matrix} 61 & 0 & 35 \\ 4 & 3 & 8 \\ 13 & 1 & 113 \end{matrix}$$

3.1.2 K = 8

Classification Accuracy (%)	71.42
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.777 1.0 0.684
Mean Precision	0.820
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.510 0.266 0.921
Mean Recall	0.566
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.616 0.616 0.785
Mean F-measure	0.607

Confusion Matrix :

$$C = \begin{matrix} 49 & 0 & 47 \\ 4 & 4 & 7 \\ 10 & 0 & 117 \end{matrix}$$

3.1.3 K=16

Classification Accuracy (%)	67.22
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.775 0.666 0.645
Mean Precision	0.695
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.395 0.133 0.944
Mean Recall	0.491
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.524 0.222 0.766
Mean F-measure	0.504

Confusion Matrix :

$$C = \begin{matrix} 38 & 1 & 57 \\ 4 & 2 & 9 \\ 7 & 0 & 120 \end{matrix}$$

3.1.4 K=32

Classification Accuracy (%)	63.44
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.783 0 0.606
Mean Precision	0.463
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.302 0.0 0.960
Mean Recall	0.420
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.436 0.0 0.743
Mean F-measure	0.393

Confusion Matrix :

$$C = \begin{matrix} 29 & 0 & 67 \\ 3 & 0 & 12 \\ 5 & 0 & 112 \end{matrix}$$

3.2 Overall Observations:

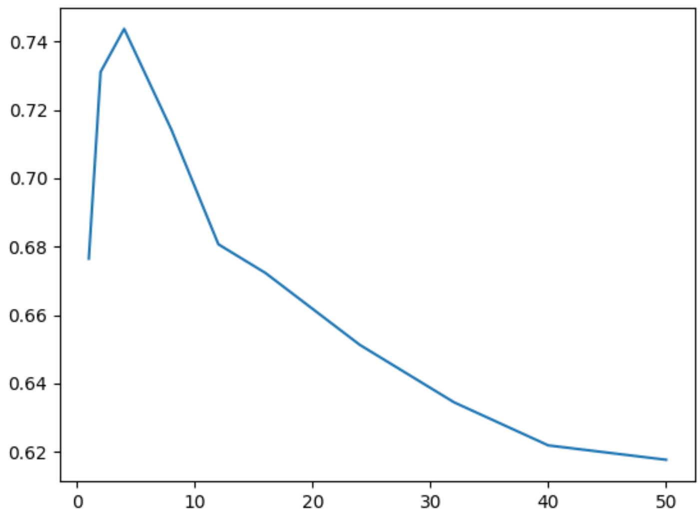

Illustration 1: Variation of accuracy w.r.t. k

- The way accuracy, precision and recall vary with k depends on the distribution of the data.
- We can see that initially with low value of k, the value of accuracy is less. This is due to noise being much more impactive at lower value of k. As k increases till about k=4, Accuracy increases up till a point, as noise becomes less and less contributing.

- After achieving a peak value, accuracy decreases again with further increase in k . As the value of k increases, almost all the train sample come into the role of contributing to the classification of a test point. Thus misclassification increases due to higher number of train samples in a class overpowering other factors.
- For class 2(kha), the number of training samples is very small as compared to other classes. For small values of k, most of its nearest neighbors belong to the same class, but as k increases, most of the nearest neighbors will come from Class 3 (kA). Thus, for sufficiently large values of k most of its test samples point will get classified into the Class 3.
- For very large values of k, the accuracy of classifier is solely due to the number of test samples of class into which entire data was classified.
- Confusion matrix for K = 50 - $\begin{matrix} 25 & 0 & 71 \\ 2 & 0 & 13 \\ 5 & 0 & 122 \end{matrix}$

4. Discrete Hidden Markov Model Method

Train and test samples are classified into 3 classes. Each class contains samples of particular sounds. Class 1 contains files which have feature vectors representing sound 'ka', class 2 contains files which have feature vectors representing sound 'kA' and class 3 contains files which have feature vectors representing sound 'kha'

N:- Number of States
M:- Number of Observation Symbols(Tokens)

(Assumption(s):- Left to Right Model)

4.1 Results

4.1.1 N=2,M=4

Classification Accuracy (%)	44.95
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.375 0.550 0.294
Mean Precision	0.406
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.437 0.472 0.333
Mean Recall	0.414
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.403 0.508 0.312
Mean F-measure	0.408

Confusion Matrix:

$$C = \begin{matrix} 42 & 45 & 9 \\ 64 & 60 & 3 \\ 6 & 4 & 5 \end{matrix}$$

4.1.2 N=3,M=4

Classification Accuracy (%)	43.277
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.318 0.493 0.600
Mean Precision	0.470
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.281 0.575 0.200
Mean Recall	0.352
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.298342541436 0.530909090909 0.3
Mean F-measure	0.376417210782

Confusion Matrix:

$$C = \begin{matrix} 27 & 68 & 1 \\ 53 & 73 & 1 \\ 5 & 7 & 3 \end{matrix}$$

4.1.3 N=4,M=4

Classification Accuracy (%)	40.235
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.287 0.446 0.400
Mean Precision	0.378
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.281 0.488 0.133
Mean Recall	0.301
F-measure for Class 1 F-measure for Class 2	0.284 0.466

F-measure for Class 3	0.2
Mean F-measure	0.316

Confusion Matrix:

$$C = \begin{matrix} 27 & 68 & 1 \\ 63 & 62 & 2 \\ 4 & 9 & 2 \end{matrix}$$

4.1.4 N=5,M=4

Classification Accuracy (%)	39.076
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.283 0.467 0.364
Mean Precision	0.371
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.271 0.496 0.267
Mean Recall	0.345
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.276 0.480 0.307
Mean F-measure	0.355

Confusion Matrix:

$$C = \begin{matrix} 26 & 67 & 3 \\ 60 & 63 & 4 \\ 6 & 5 & 4 \end{matrix}$$

4.1.5 N=2,M=8

Classification Accuracy (%)	46.218
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.391 0.553 0.158
Mean Precision	0.367

Recall for Class 1 Recall for Class 2 Recall for Class 3	0.354 0.575 0.200
Mean Recall	0.376
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.371 0.563 0.176
Mean F-measure	0.370

Confusion Matrix:

$$C = \begin{matrix} 34 & 55 & 7 \\ 45 & 73 & 9 \\ 8 & 4 & 3 \end{matrix}$$

4.1.6 N=3,M=8

Classification Accuracy (%)	46.488
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.402 0.536 0.167
Mean Precision	0.368
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.344 0.583 0.200
Mean Recall	0.375
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.370 0.558 0.181
Mean F-measure	0.370

Confusion Matrix:

$$C = \begin{matrix} 33 & 56 & 7 \\ 45 & 74 & 8 \\ 4 & 8 & 3 \end{matrix}$$

4.1.7 N=4,M=8

Classification Accuracy (%)	52.101
Precision for Class 1	0.453

Precision for Class 2 Precision for Class 3	0.571 0.200
Mean Precision	0.408
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.250 0.764 0.200
Mean Recall	0.405
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.322 0.653 0.2
Mean F-measure	0.391

Confusion Matrix:

$$C = \begin{matrix} 24 & 66 & 6 \\ 24 & 97 & 6 \\ 5 & 7 & 3 \end{matrix}$$

4.1.8 N=5,M=8

Classification Accuracy (%)	49.580
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.407 0.552 0.214
Mean Precision	0.391
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.250 0.717 0.200
Mean Recall	0.389
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.309 0.623 0.206
Mean F-measure	0.379

Confusion Matrix:

$$C = \begin{matrix} 24 & 66 & 6 \\ 31 & 91 & 5 \\ 4 & 8 & 3 \end{matrix}$$

4.1.9 N=2,M=16

Classification Accuracy (%)	46.639
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.468 0.590 0.077
Mean Precision	0.378
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.375 0.567 0.200
Mean Recall	0.381
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.416 0.578 0.111
Mean F-measure	0.368

Confusion Matrix:

$$C = \begin{matrix} 36 & 43 & 17 \\ 36 & 72 & 19 \\ 5 & 7 & 3 \end{matrix}$$

4.1.10 N=3,M=16

Classification Accuracy (%)	50.000
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.479 0.574 0.125
Mean Precision	0.393
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.365 0.638 0.200
Mean Recall	0.401
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.414 0.604 0.153
Mean F-measure	0.390

Confusion Matrix:

$$C = \begin{matrix} 35 & 53 & 8 \\ 33 & 81 & 13 \\ 5 & 7 & 3 \end{matrix}$$

4.1.11 N=4,M=16

Classification Accuracy (%)	49.580
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.477 0.564 0.125
Mean Precision	0.389
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.323 0.661 0.200
Mean Recall	0.395
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.385 0.608 0.153
Mean F-measure	0.382

Confusion Matrix:

$$C = \begin{matrix} 31 & 57 & 8 \\ 30 & 84 & 13 \\ 4 & 8 & 3 \end{matrix}$$

4.1.12 N=5,M=16

Classification Accuracy (%)	50.420
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.492 0.573 0.120
Mean Precision	0.395
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.323 0.677 0.200
Mean Recall	0.400
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.389 0.620 0.15
Mean F-measure	0.386

Confusion Matrix:

$$C = \begin{matrix} 31 & 57 & 8 \\ 27 & 86 & 14 \\ 5 & 6 & 4 \end{matrix}$$

4.1.13 N=2,M=32

Classification Accuracy (%)	50.000
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.494 0.592 0.040
Mean Precision	0.375
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.427 0.606 0.067
Mean Recall	0.367
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.458100558659 0.599221789883 0.05
Mean F-measure	0.369107449514

Confusion Matrix:

$$C = \begin{matrix} 41 & 46 & 9 \\ 35 & 77 & 15 \\ 7 & 7 & 1 \end{matrix}$$

4.1.14 N=3,M=32

Classification Accuracy (%)	48.319
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.456 0.557 0.053
Mean Precision	0.355
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.375 0.614 0.067
Mean Recall	0.352
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.411 0.584 0.058
Mean F-measure	0.351

Confusion Matrix:

$$C = \begin{matrix} 36 & 54 & 6 \\ 37 & 78 & 12 \\ 6 & 8 & 1 \end{matrix}$$

4.1.15 N=4,M=32

Classification Accuracy (%)	50.840
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.488 0.588 0.050
Mean Precision	0.375
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.417 0.630 0.067
Mean Recall	0.371
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.449 0.608 0.057
Mean F-measure	0.371

Confusion Matrix:

$$C = \begin{matrix} 40 & 50 & 6 \\ 34 & 80 & 13 \\ 8 & 6 & 1 \end{matrix}$$

4.1.16 N=5,M=32

Classification Accuracy (%)	51.681
Precision for Class 1 Precision for Class 2 Precision for Class 3	0.500 0.582 0.056
Mean Precision	0.379
Recall for Class 1 Recall for Class 2 Recall for Class 3	0.385 0.669 0.067

Mean Recall	0.374
F-measure for Class 1 F-measure for Class 2 F-measure for Class 3	0.435 0.622 0.060
Mean F-measure	0.372

Confusion Matrix:

$$C = \begin{matrix} 37 & 54 & 5 \\ 30 & 85 & 12 \\ 7 & 7 & 1 \end{matrix}$$

4.2 Overall Observations

Since each observation was a 39 dimensional continuous valued feature vector, in discrete hidden markov models, to restrict the number of tokens(observation symbols), we have used the technique of vector quantization. In vector quantization we took the entire training data for all 3 classes (“ka”,”kA” and “kha”) and applied K-Means clustering with K equal to the number of tokens we want. We applied K-Means on the training data for all the classes together rather than individual classes because It may happen that the a cluster for sounds similar to say “k” have label 1 in 1st class and some other label in another class. With vector quantization we hope that similar sounds from the data get same labels.

We observe that for less number of clusters (M), the accuracies for the DHMM classifier is less, because for low cluster number the of distinct sounds we are expecting to get will have no physical meaning as the vectors inside a cluster will not be representing a exclusive sound. The accuracy increases as M increases.

Further we can see a lot of confusion between class1 (“ka”) and class2(“kA”) due to similarity between the phoneme of these words.

Our training data contains 3 classes “ka,kA,kha” contains of majorly 4 sounds ‘k’ ,’a’,’aa’ and ‘kh’,(4 phonemes) so selection of number of states of 4 or more must be the optimal choice as we can see, the accuracies for states less than 4 are usually lesser than those with number of states more than or equal to 4 but increasing the number N will lead to more number of parameters to be estimated and since data for our classes are very less, the estimate will not be accurate.

5.Conclusions

In KNN classifier, we observed lower accuracies for higher values of K, on e way to improve this can be done by having the data points closer to the test data point contribute more to the decision rather than all K of them contributing equally, (giving closer data points more weightage).

In DHMM, we observe that because the phonemes of class ka and kA are similar so there is a lot of confusion between them hence accuracies are low further the number of training examples for class “kha” are quite less as compared to other 2 classes and for large M and N it may lead to curse of dimensionality.