K-Nearest Neighbour and Hidden Markov Model

Akash Agrawal, B15303, IIT Mandi
Group No: 2

Abstract:

This report presents theoretical analysis of the results obtained on implementing the K-nearest neighbour classifier and Discrete Hidden Markov Model and practically testing them on Image and Speech datasets. This gives an analysis of how both the algorithms perform on the real-world data. Models are tested on real world image and speech datasets. There are some limitations while using these models and that has been discussed in the last section of this report

1 Introduction

K-Nearest Neighbour Classifier: This is a probably the simplest classifier for classification purpose. It is a non-parametric method so there is no need of estimation of parameters. In this classifier we classify the test example by looking K -nearest neighbouring points to that test example. The class which has maximum number of points similar to that test example is considered as class label for that test example. A commonly used method to test the similarity in varying length sequence is using Dynamic Time Warping. In this particular classifier since the length of the sequence to be matched is different for all the examples of all the classes, DTW distance is used for accounting the similarity between different pairs.

DTW distance: This algorithm is used specially in speech recognition It is an algorithm which gives measure of the dissimilarity of the two sequences having varying length. For same length sequences generally, Euclidean distance is preferred.

Discrete Hidden Markov Model (DHMM classifier): HMM’s are extensively used in speech recognition, AI and pattern recognition tasks nowadays. They are used everywhere specially to model time series data.

It is a parametric method having parameters $\lambda = (N, M, A, B, \pi)$,

where N=No of states,

M=No of unique observation symbols,

A=State transition probability matrix,

B=state observation symbol probability,

Π=Initial state probability

In Discrete Hidden Markov Model, the number of state observation symbols are discrete, whereas in Continuous Hidden Markov Model the state observation symbol probability is a continuous probability density function. It is called Hidden Markov Model because the underlying state sequence of the observation sequence is not known to us. Parameters of the model are estimated using Baum-Welch algorithm which is basically an Expectation Maximisation Algorithm. We build a model for each class having parameters (N, M, A, B, π) and then checks the probability of the test sequence in all the classes and class label is assigned accordingly.

a) Ergodic HMM: In this model there is a finite path from every state to every other state.
b) Non-ergodic HMM: In this model from each state there is a path to only certain states and transition probabilities to other states is zero. For example: Left-Right HMM in which we can only go ahead that any state left behind cannot be visited.

Assumptions:

i) It is assumed that the system being modelled is first order Hidden Markov model which means that output at time 't' depends only on the what happened at time 't-1' and not before.
ii) It is Discrete Hidden Markov Model.
iii) Values of parameters N and M is same for all the different classes.

2 Baum-Welch Algorithm

This algorithm is used for estimation for parameters for Hidden Markov Model. It is an Expectation-Maximisation Algorithm.

<u>Procedure:</u>

<u>Step 1:</u> Initialise the parameters λ^{old}.

<u>Step 2</u>: Estimate ξ_t(i,j) and γ_t(i) using forward (α_t(i)) and backward variables ($\beta_t(V_k)$) corresponding to λ^{old} (Expectation Step).

ξ_t (i,j) = $\frac{\alpha t(i) * aij * bj(Ot+1)\beta t+1(j)}{\sum_{i=1}^{N}(\sum_{j=1}^{N} \alpha t(i) * aij * bj(Ot+1)\beta t+1(j))}$ = Probability of being in state 'i' at time 't' and state 'j' at time 't+1'.

γ_t (i) = $\sum_{j=1}^{N} \xi t(I, j)$ = Probability of transition from state 'i' at time 't'.

<u>Step 3</u>: Re-estimate parameters λ^{new} using ξ_t(I,j) and γ_t(i) (Maximisation Step).

$aij = \frac{\sum_{t=1}^{T-1} \xi t(i,j)}{\sum_{t=1}^{T-1} \gamma t(i)}$

$b_i (V_k) = \frac{\sum_{t=1|Ot=Vk}^{T-1} \xi t(i,j)}{\sum_{t=1}^{T-1} \gamma t(i)}$

$\pi_i = \gamma_1(i)$

<u>Step 4</u>: Repeat Step 2 and Step 3 till convergence. Convergence condition can be

a) Difference of Log likelihood of the data is less than some threshold.
b) We can repeat steps some fixed number of times.

After estimating parameters, we calculate the probability of test sequence in all the classes and the class having maximum probability is assigned as the class label.

P $(\frac{O}{\lambda})$ = $\sum_{j=1}^{N} \alpha t(j)$=Probability of an observation sequence given λ.

3 Feature Extraction

For image dataset: A image was divided into 64*64 blocks and a 24-dimension colour histogram vector is computed corresponding to each non overlapping block in a image. In this way we got a sequence of vectors or observation symbols corresponding to each image.

For Speech dataset: Speech dataset is given as the 39-dimensional mel frequency cepstral coefficient (MFCC) features extracted frame by frame from utterances for a particular CV segment by multiple people.

4 Results and Inferences for K-Nearest Neighbour Method

K represents how much nearest neighbours we have to assign the class label to a particular example.

i) Image Dataset:

a) For K=3

4	6	6
44	40	39
2	4	5

Table 1: Confusion Matrix

Accuracy: 0.32666666666666666
Precision for class 0 : 0.25
Recall for class 0 : 0.08
F-measure for class 0 ; 0.12121212121212122
Precision for class 1 : 0.3252032520325203
Recall for class 1 : 0.8
F-measure for class 1 ; 0.46242774566473993
Precision for class 2 : 0.45454545454545453
Recall for class 2 : 0.1
F-measure for class 2 ; 0.16393442622950818
Mean Precision: 0.34324956885932495
Mean Recall: 0.32666666666666666
Mean F-measure: 0.24919143103545646

b) For K=5

3	7	3
43	41	39
4	2	8

Table 2: Confusion Matrix

Accuracy: 0.3466666666666667
Precision for class 0 : 0.23076923076923078
Recall for class 0 : 0.06
F-measure for class 0 ; 0.09523809523809523
Precision for class 1 : 0.3333333333333333
Recall for class 1 : 0.82
F-measure for class 1 ; 0.47398843930635837
Precision for class 2 : 0.5714285714285714
Recall for class 2 : 0.16
F-measure for class 2 ; 0.25
Mean Precision: 0.3785103785103785
Mean Recall: 0.3466666666666666
Mean F-measure: 0.27307551151481785

c) For K=11

1	0	0
43	48	42
6	2	8

Table 3: Confusion Matrix

Accuracy: 0.38
Precision for class 0 : 1.0
Recall for class 0 : 0.02
F-measure for class 0 ; 0.0392156862745098
Precision for class 1 : 0.3609022556390977
Recall for class 1 : 0.96
F-measure for class 1 ; 0.5245901639344263
Precision for class 2 : 0.5
Recall for class 2 : 0.16
F-measure for class 2 ; 0.24242424242424243
Mean Precision: 0.6203007518796992
Mean Recall: 0.37999999999999995
Mean F-measure: 0.2687433642110595

d) For K=20

1	0	0
40	48	37
9	2	13

Table 4: Confusion Matrix

Accuracy: 0.41333333333333333
Precision for class 0 : 1.0
Recall for class 0 : 0.02
F-measure for class 0 ; 0.0392156862745098
Precision for class 1 : 0.384
Recall for class 1 : 0.96
F-measure for class 1 ; 0.5485714285714286
Precision for class 2 : 0.5416666666666666
Recall for class 2 : 0.26
F-measure for class 2 ; 0.35135135135135137
Mean Precision: 0.6418888888888888
Mean Recall: 0.41333333333333333
Mean F-measure: 0.3130461553990966

Inference:

The classification accuracy with respect to different values of K is shown below:

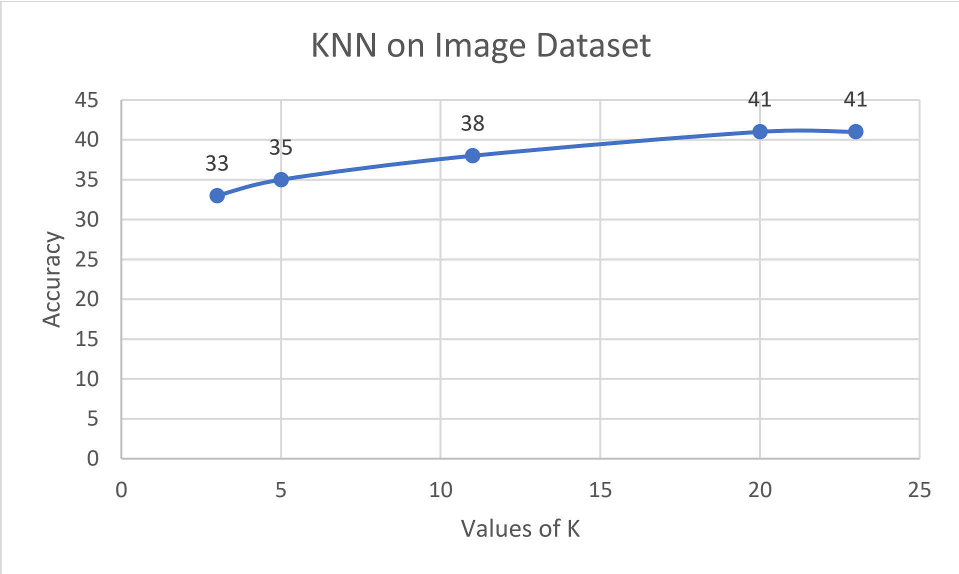

Fig 1: Chart showing variation in Accuracy v/s K

As the value of K increases the classification accuracy is not following any particular pattern or distribution. This can be much clearer in KNN on speech Dataset. This is because as we continuously expand the neighbourhood of a point it keeps on jumping in different classes. The value of K for which it maximises accuracy depends on the dataset.

ii) Speech Dataset:

a) For K=3

29	25	8
6	17	3
1	4	10

Table 5: Confusion Matrix

Accuracy: 0.5436893203883495
Precision for class 0 : 0.46774193548387094
Recall for class 0 : 0.8055555555555556
F-measure for class 0 ; 0.5918367346938775
Precision for class 1 : 0.6538461538461539
Recall for class 1 : 0.3695652173913043
F-measure for class 1 ; 0.4722222222222222
Precision for class 2 : 0.6666666666666666
Recall for class 2 : 0.47619047619047616
F-measure for class 2 ; 0.5555555555555556
Mean Precision: 0.5960849186655638
Mean Recall: 0.5504370830457788
Mean F-measure: 0.5398715041572185

b) For K=5

28	21	7
8	23	4
0	2	10

Table 6: Confusion Matrix

Accuracy: 0.5922330097087378
Precision for class 0 : 0.5
Recall for class 0 : 0.7777777777777778
F-measure for class 0 ; 0.6086956521739131
Precision for class 1 : 0.6571428571428571
Recall for class 1 : 0.5
F-measure for class 1 ; 0.5679012345679013
Precision for class 2 : 0.8333333333333334
Recall for class 2 : 0.47619047619047616
F-measure for class 2 ; 0.6060606060606061
Mean Precision: 0.6634920634920635
Mean Recall: 0.5846560846560847
Mean F-measure: 0.5942191642674736

c) For K=7

26	21	7
10	21	6
0	4	8

Table 7: Confusion Matrix

Accuracy: 0.5339805825242718
Precision for class 0 : 0.48148148148148145
Recall for class 0 : 0.7222222222222222
F-measure for class 0 ; 0.5777777777777777
Precision for class 1 : 0.5675675675675675
Recall for class 1 : 0.45652173913043476
F-measure for class 1 ; 0.506024096385542
Precision for class 2 : 0.6666666666666666
Recall for class 2 : 0.38095238095238093
F-measure for class 2 ; 0.4848484848484849
Mean Precision: 0.5719052385719051
Mean Recall: 0.519898780768346
Mean F-measure: 0.5228834530039349

d) For K=11

28	23	8
8	22	6
0	1	7

Table 8: Confusion Matrix

Accuracy: 0.5533980582524272
Precision for class 0 : 0.4745762711864407
Recall for class 0 : 0.7777777777777778
F-measure for class 0 ; 0.5894736842105264
Precision for class 1 : 0.6111111111111112
Recall for class 1 : 0.4782608695652174
F-measure for class 1 ; 0.5365853658536586
Precision for class 2 : 0.875
Recall for class 2 : 0.3333333333333333
F-measure for class 2 ; 0.48275862068965514
Mean Precision: 0.6535624607658507
Mean Recall: 0.5297906602254429
Mean F-measure: 0.5362725569179467

e) For K=17

27	19	7
9	25	6
0	2	8

Table 9: Confusion Matrix

Accuracy: 0.5825242718446602
Precision for class 0 : 0.5094339622641509
Recall for class 0 : 0.75
F-measure for class 0 ; 0.6067415730337078
Precision for class 1 : 0.625
Recall for class 1 : 0.5434782608695652
F-measure for class 1 ; 0.5813953488372093
Precision for class 2 : 0.8
Recall for class 2 : 0.38095238095238093
F-measure for class 2 ; 0.5161290322580645
Mean Precision: 0.644811320754717
Mean Recall: 0.5581435472739821
Mean F-measure: 0.5680886513763271

f) For K = 19

28	17	7
8	27	7
0	2	7

Table 10: Confusion Matrix

Accuracy: 0.6019417475728155
Precision for class 0 : 0.5384615384615384
Recall for class 0 : 0.7777777777777778
F-measure for class 0 ; 0.6363636363636364
Precision for class 1 : 0.6428571428571429
Recall for class 1 : 0.5869565217391305
F-measure for class 1 ; 0.6136363636363638
Precision for class 2 : 0.7777777777777778
Recall for class 2 : 0.3333333333333333
F-measure for class 2 ; 0.4666666666666666
Mean Precision: 0.6530321530321531
Mean Recall: 0.5660225442834138
Mean F-measure: 0.5722222222222222

g) For K = 23

24	18	9
11	27	7
1	1	5

Table 11: Confusion Matrix

Accuracy: 0.5436893203883495
Precision for class 0 : 0.47058823529411764
Recall for class 0 : 0.6666666666666666
F-measure for class 0 ; 0.5517241379310345
Precision for class 1 : 0.6
Recall for class 1 : 0.5869565217391305
F-measure for class 1 ; 0.5934065934065934
Precision for class 2 : 0.7142857142857143
Recall for class 2 : 0.23809523809523808
F-measure for class 2 ; 0.35714285714285715
Mean Precision: 0.5949579831932773
Mean Recall: 0.4972394755003451
Mean F-measure: 0.5007578628268283

h) For K = 31

21	16	8
14	30	7
1	0	6

Table 12: Confusion Matrix

Accuracy: 0.5533980582524272
Precision for class 0 : 0.4666666666666667
Recall for class 0 : 0.5833333333333334
F-measure for class 0 ; 0.5185185185185186
Precision for class 1 : 0.5882352941176471
Recall for class 1 : 0.6521739130434783
F-measure for class 1 ; 0.6185567010309279
Precision for class 2 : 0.8571428571428571
Recall for class 2 : 0.2857142857142857
F-measure for class 2 ; 0.42857142857142855
Mean Precision: 0.6373482726423902
Mean Recall: 0.5070738440303657
Mean F-measure: 0.5218822160402917

Inference:

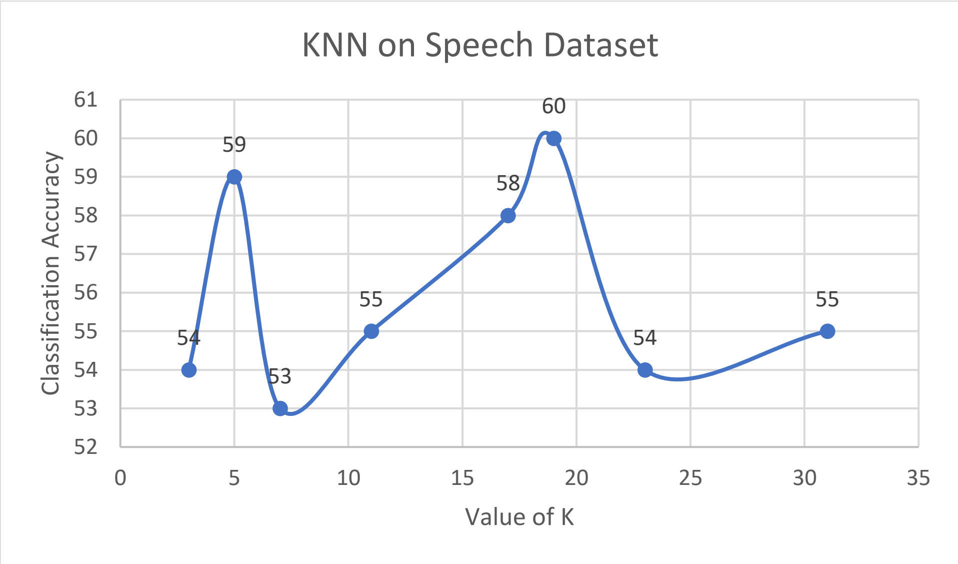

Fig 2: Chart showing variation in accuracy w.r.t. K

As the value of K increases the classification accuracy is increasing and decreasing without any particular pattern. This is because as we continuously expand the neighbourhood of a point it keeps on jumping in different classes. The value of K for which it maximises accuracy depends on the dataset.

PS. It is good practice to choose the values of K which are even or multiple of the no of classes because it may lead to the test example having equal probability in multiple classes.

5 Results and Inferences for Discrete Hidden Markov Model

Significance of N and M in both models:

In image dataset

N= This has no physical significance but it plays its role in deciding if sky(blue) can come after mountain(brown) in particular class or not.

M= No of different colour entities like sky, trees, walls etc.

For speech dataset

N= different alphabets that are part of the Speech Dataset.

M=Frequency or speed with which different people speaks.

1) Ergodic Discrete Hidden Markov Model for image dataset:

a) For N=5 and M=7

42	30	33
1	0	3
7	20	14

Table 13: Confusion Matrix

Accuracy: 0.37333333333333335
Precision for class 0 : 0.4
Recall for class 0 : 0.84
F-measure for class 0 ; 0.5419354838709678
Precision for class 1 : 0.0
Recall for class 1 : 0.0
Precision for class 2 : 0.34146341463414637
Recall for class 2 : 0.28
F-measure for class 2 ; 0.3076923076923077
Mean Precision: 0.24715447154471545
Mean Recall: 0.37333333333333335
Mean F-measure: 0.2832092638544252

b) For N=5 and M=11

41	29	33
0	0	0
9	21	17

Table 14: Confusion Matrix

Accuracy: 0.38666666666666666
Precision for class 0 : 0.39805825242718446
Recall for class 0 : 0.82
F-measure for class 0 ; 0.5359477124183006
Recall for class 1 : 0.0
F-measure for class 1 ; 0.0
Precision for class 2 : 0.3617021276595745
Recall for class 2 : 0.34
F-measure for class 2 ; 0.3505154639175258
Mean Precision: 0.25325346002891963
Mean Recall: 0.38666666666666666
Mean F-measure: 0.29548772544527546

c) For N=5 and M=17

39	28	31
0	0	1
11	22	18

Table 15: Confusion Matrix

Accuracy: 0.38
Precision for class 0 : 0.3979591836734694
Recall for class 0 : 0.78
F-measure for class 0 ; 0.5270270270270271
Precision for class 1 : 0.0
Recall for class 1 : 0.0
Precision for class 2 : 0.35294117647058826
Recall for class 2 : 0.36
F-measure for class 2 ; 0.3564356435643565
Mean Precision: 0.2503001200480192
Mean Recall: 0.38000000000000006
Mean F-measure: 0.2944875568637945

d) For N=5 and M=23

35	24	29
0	0	0
15	26	21

Table 16: Confusion Matrix

Accuracy: 0.37333333333333335
Precision for class 0 : 0.3977272727272727
Recall for class 0 : 0.7
F-measure for class 0 ; 0.5072463768115942
Recall for class 1 : 0.0
F-measure for class 1 ; 0.0
Precision for class 2 : 0.3387096774193548
Recall for class 2 : 0.42
F-measure for class 2 ; 0.375
Mean Precision: 0.24547898338220916
Mean Recall: 0.3733333333333333
Mean F-measure: 0.29408212560386476

e) For N = 5 and M = 31

37	26	22
0	0	0
13	24	28

Table 17: Confusion Matrix

Accuracy: 0.43333333333333335
Precision for class 0 : 0.43529411764705883
Recall for class 0 : 0.74
F-measure for class 0 ; 0.5481481481481482
Recall for class 1 : 0.0
F-measure for class 1 ; 0.0
Precision for class 2 : 0.4307692307692308
Recall for class 2 : 0.56
F-measure for class 2 ; 0.4869565217391305
Mean Precision: 0.2886877828054299
Mean Recall: 0.43333333333333335
Mean F-measure: 0.3450348899624262

f) For N = 3 and M = 7

42	30	33
1	0	3
7	20	14

Table 18: Confusion Matrix

Accuracy: 0.37333333333333335
Precision for class 0 : 0.4
Recall for class 0 : 0.84
F-measure for class 0 ; 0.5419354838709678
Precision for class 1 : 0.0
Recall for class 1 : 0.0
Precision for class 2 : 0.34146341463414637
Recall for class 2 : 0.28
F-measure for class 2 ; 0.3076923076923077
Mean Precision: 0.24715447154471545
Mean Recall: 0.37333333333333335
Mean F-measure: 0.2832092638544252

Inference:

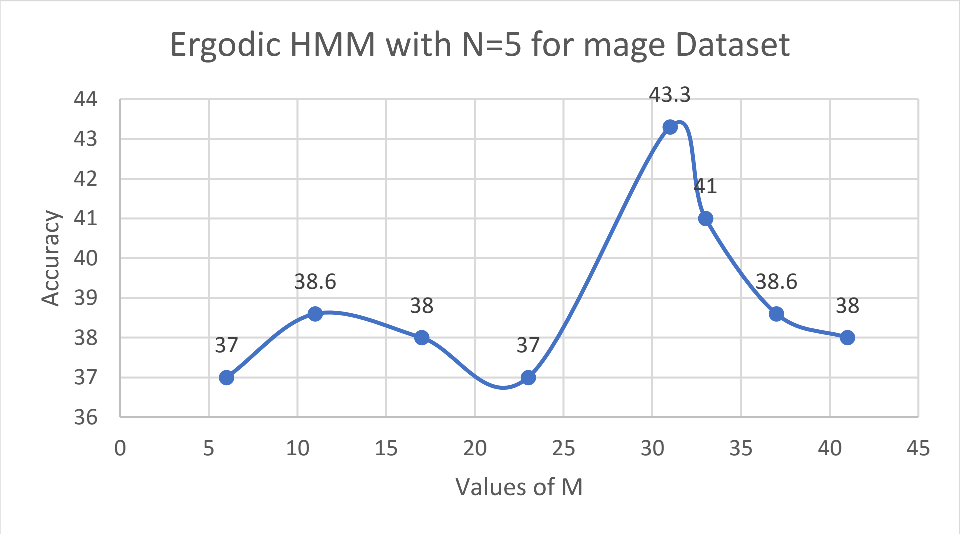

Fig 3: Ergodic HMM Accuracy varying with M at N=5

As we know that M represents different colour identities present in the dataset like sky(blue), mountains(brown), etc, so as we increase the value of M the accuracy will sometimes increase then decrease and at some particular value it will reach global maximum and that value of M separates the different colour identities in a best possible way and will lead to a good classification accuracy.

The first dataset was of Coast images, second was of Kennel outdoor images, and third was of indoor Volleyball Court images. One interesting thing to note is that the number of examples classified in the 2nd class is very less as compared to other classes. This may be because the kennel image set has background resembling to that of coast images and kennel and other thing resembling more to volleyball indoor images and we are considering only colour based features of an image which is resulting in misclassification of images.

2) Non-ergodic Discrete Hidden Markov Model for Speech Dataset:

a) For N=5 and M=11

21	20	12
10	13	4
5	13	5

Table19: Confusion Matrix

Accuracy: 0.3786407766990291
Precision for class 0 : 0.39622641509433965
Recall for class 0 : 0.5833333333333334
F-measure for class 0 ; 0.4719101123595506
Precision for class 1 : 0.48148148148148145
Recall for class 1 : 0.2826086956521739
F-measure for class 1 ; 0.3561643835616438
Precision for class 2 : 0.21739130434782608
Recall for class 2 : 0.23809523809523808
F-measure for class 2 ; 0.22727272727272724
Mean Precision: 0.3650330669745491
Mean Recall: 0.3680124223602485
Mean F-measure: 0.3517824077313072

b) For N=5 and M=17

16	16	9
19	23	7
1	7	5

Table 20: Confusion Matrix

Accuracy: 0.42718446601941745
Precision for class 0 : 0.3902439024390244
Recall for class 0 : 0.4444444444444444
F-measure for class 0 ; 0.4155844155844156
Precision for class 1 : 0.46938775510204084
Recall for class 1 : 0.5
F-measure for class 1 ; 0.4842105263157895
Precision for class 2 : 0.38461538461538464
Recall for class 2 : 0.23809523809523808
F-measure for class 2 ; 0.2941176470588235
Mean Precision: 0.4147490140521499
Mean Recall: 0.3941798941798942
Mean F-measure: 0.39797086298634293

c) For N=5 and M=19

18	19	9
9	6	2
9	21	10

Table 21: Confusion Matrix

Accuracy: 0.3300970873786408
Precision for class 0 : 0.391304347826087
Recall for class 0 : 0.5
F-measure for class 0 ; 0.4390243902439025
Precision for class 1 : 0.35294117647058826
Recall for class 1 : 0.13043478260869565
F-measure for class 1 ; 0.1904761904761905
Precision for class 2 : 0.25
Recall for class 2 : 0.47619047619047616
F-measure for class 2 ; 0.32786885245901637
Mean Precision: 0.33141517476555843
Mean Recall: 0.3688750862663906
Mean F-measure: 0.31912314439303646

d) For N=2 and M=11

10	7	6
23	34	12
3	5	3

Table 22: Confusion Matrix

Accuracy: 0.4563106796116505
Precision for class 0 : 0.43478260869565216
Recall for class 0 : 0.2777777777777778
F-measure for class 0 ; 0.3389830508474576
Precision for class 1 : 0.4927536231884058
Recall for class 1 : 0.7391304347826086
F-measure for class 1 ; 0.591304347826087
Precision for class 2 : 0.2727272727272727
Recall for class 2 : 0.14285714285714285
F-measure for class 2 ; 0.18749999999999997
Mean Precision: 0.4000878348704435
Mean Recall: 0.3865884518058431
Mean F-measure: 0.37259579955784816

e) For N=3 and M=11

9	6	4
21	34	12
6	6	5

Table 23: Confusion Matrix

Accuracy: 0.46601941747572817
Precision for class 0 : 0.47368421052631576
Recall for class 0 : 0.25
F-measure for class 0 ; 0.32727272727272727
Precision for class 1 : 0.5074626865671642
Recall for class 1 : 0.7391304347826086
F-measure for class 1 ; 0.6017699115044248
Precision for class 2 : 0.29411764705882354
Recall for class 2 : 0.23809523809523808
F-measure for class 2 ; 0.2631578947368421
Mean Precision: 0.4250881813841012
Mean Recall: 0.4090752242926156
Mean F-measure: 0.3974001778379981

f) For N=4 and M=11

14	9	9
13	16	4
9	21	8

Table 24: Confusion Matrix

Accuracy: 0.36893203883495146
Precision for class 0 : 0.4375
Recall for class 0 : 0.3888888888888889
F-measure for class 0 ; 0.411764705882353
Precision for class 1 : 0.48484848484848486
Recall for class 1 : 0.34782608695652173
F-measure for class 1 ; 0.4050632911392405
Precision for class 2 : 0.21052631578947367
Recall for class 2 : 0.38095238095238093
F-measure for class 2 ; 0.2711864406779661
Mean Precision: 0.3776249335459862
Mean Recall: 0.37255578559926383
Mean F-measure: 0.3626714792331865

Inferences:

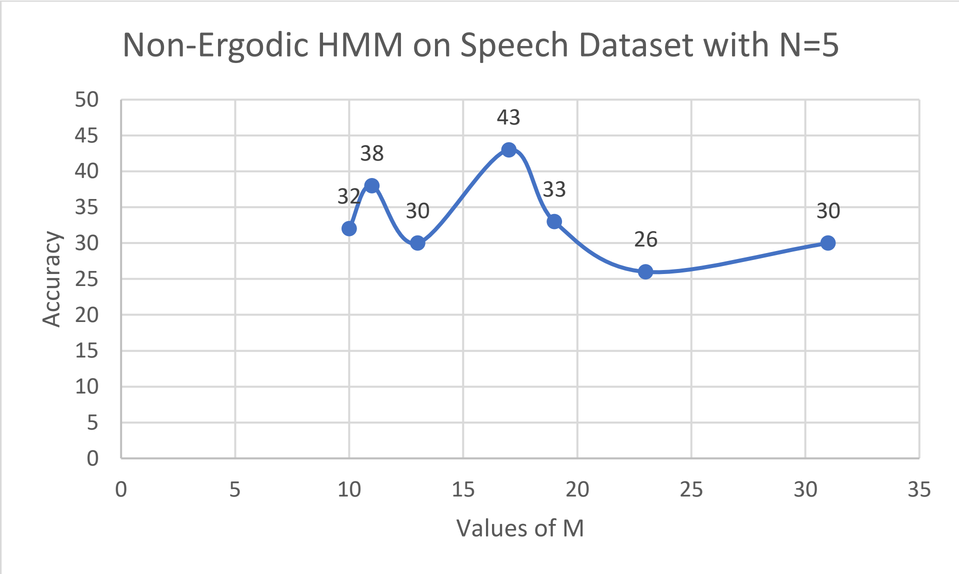

Fig 4: Non-ergodic HMM Accuracy varying with M at N=5

As we know that the significance of M is the different frequency or speed with which the alphabet was spoken. As we have dataset of CV segment spoken by different people it so happens that the value of M for some small group of people is same or very close to each other. So, at those values of M that particular group of people will be classified precisely and will lead to good classification accuracy for that particular value of M.

That's why we are getting some peaks while varying the value of M for same value of N. So maximum classification accuracy is very likely to come around these maxims as these maxims lead to precisely classify CV segment spoken by a particular group of people. This particular data justifies the claim too as the maximum accuracy with testing so far is around N=3 and M=11(where there is a maxima).

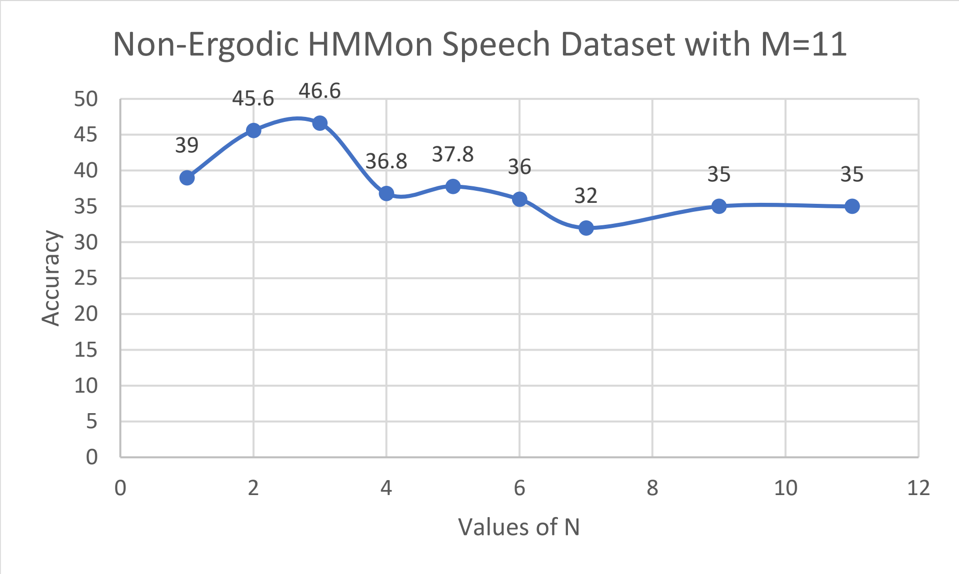

Fig 5: Non-ergodic HMM Accuracy varying with N at M=11

Inference:

As we know that significance of N is the number of alphabets spoken. Speech dataset used was having utterances for a particular CV segment by multiple people. CV segment were of two alphabets. As we can clearly see from the Fig 6 that for a constant value of M=11, we are getting accuracy maximum around N=2 and N=3 which satisfies the theoretical predictions. It also justifies the significance of N in this model.

6 Limitations while implementing hidden Markov model

Few limitations encountered while implementation:

i) Forward and backward variables tend to zero: When the time series is too long the value of the forward and backward variables approaches to zero. They become so small that they go below the precision of any machine which makes their value equal to zero. This also makes the probability of the sequence zero which lead to math domain errors while estimating parameters and also lead to misclassify test examples.

ii) Initial values of parameters: In a parametric method of classification how fast the parameters converges and how good the solution obtained is also depends on the initial values of the parameters. It might be that you start with some random values of the parameters and end up in some local minima/maxima. So, it is important to initialise parameters appropriately. Well there is no simple solution to this problem. One solution could be to give make all the transitions from one state to another and state observation probabilities equiprobable and let it left to the model to converge to a good estimation.

---THE END---------------------------------------