

CS6300 - Speech Technology

Mini Project 2 : HMMs

Dhanush Krishna : EE16B009

Vignesh S : EE16B127

September 9, 2020

Problem 1

Isolated Digit Recognition using Discrete HMMs. Use the given features also extract your own features and compare the results

Method : Training and testing

The Isolated digits we had to recognize were 1,2,3,6,7. There were 2 sets of features that were used for training. One was the given set of features and the other was the generated cepstral coefficients. The generated cepstral coefficients are defined as,

$$c[n] = \mathcal{F}^{-1}\{\log(\mathcal{F}\{x[n]\})\}$$

The method of training and testing is same for both the set of features. Only difference between them is the process of generation.

During the training the entire set of features of all the digits was used and K-means model trained. Vector Quantization was then performed on each digit's features to obtain a sequence of integers (corresponding to the centroids) for each mfcc file. These bunch of sequences (as many as the number of train files) was passed as input the already written .train executable (which performs the forward-backward algorithm) which generates HMM model for each digit. Command used for obtaining models using the executable (48 clusters, digit 7 model)

```
1 ./train_hmm seven.seq 1234 3 48 .01
```

For the test features we perform VQ over the same K-means model and obtain the set of sequences for each digit. We then stack test files of all the digits and then pass them to the executable with each digit's HMM model as the second argument. Command to run the test executable is given below (digit 2 model)

```
1 ./test_hmm file.test two.seq.hmm
```

After running this we get 5 files with log(probabilities) for each of the 5 models. We take argmax (for each test file among the 5 models) of that and classify which digit the test file belongs. For 48 clusters the model is almost 100 % accurate. The confusion matrices, ROC curves and DET curves have been plotted for different cluster values and shown below.

Visualisations

The accuracy improved as the number of clusters in the K-means model used to learn the features was increased as shown in Figure 1. Almost 100 % accuracy (98.33%) was observed with 48 clusters using the given features.

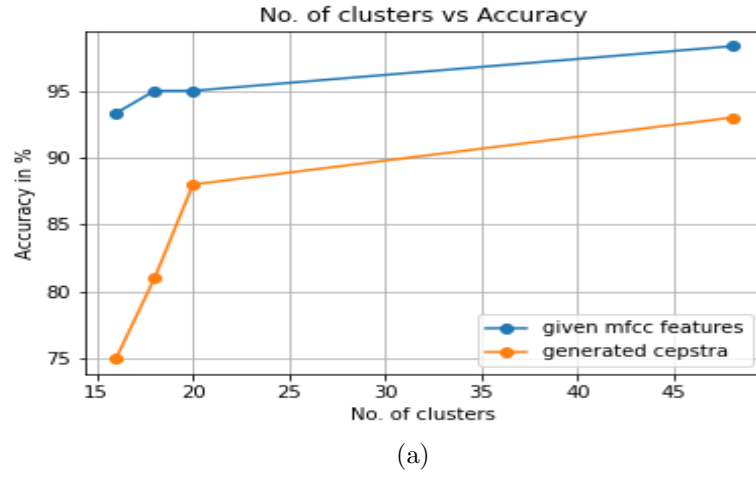


Figure 1: No. of clusters vs Accuracy

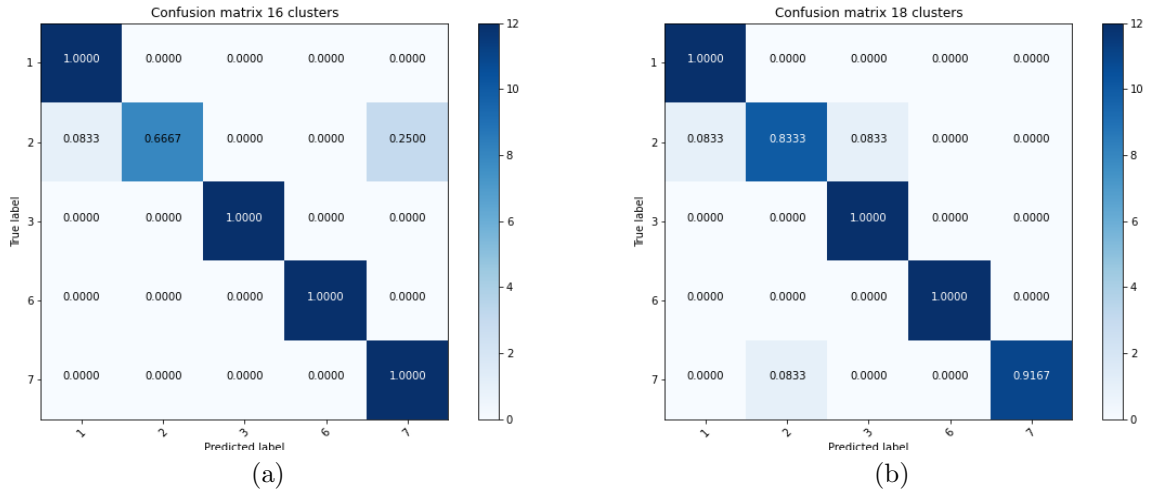


Figure 2: Confusion matrices for 16 and 18 clusters

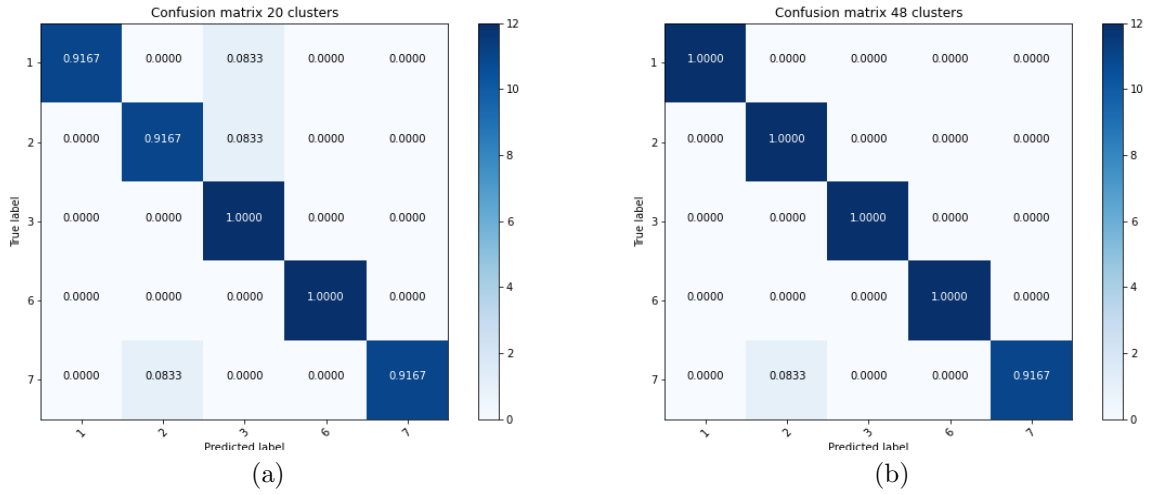
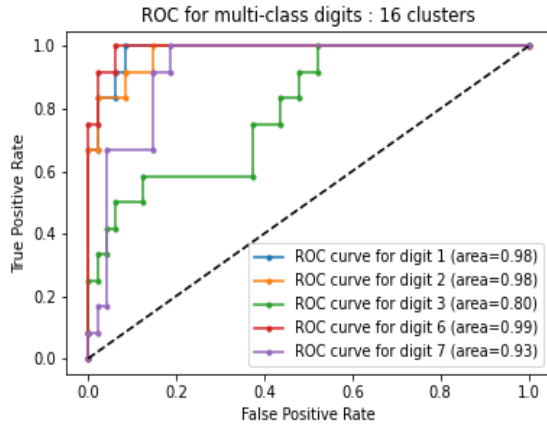
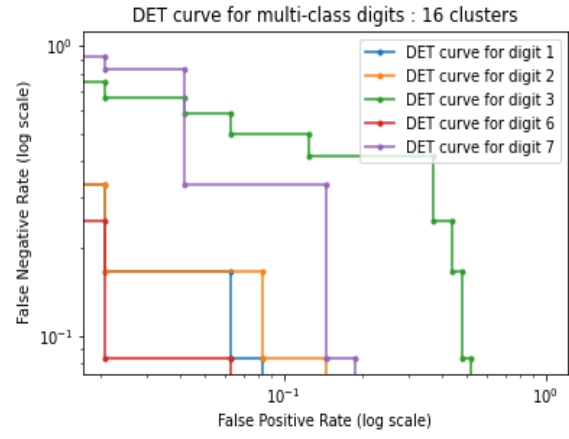


Figure 3: Confusion matrices for 20 and 48 clusters

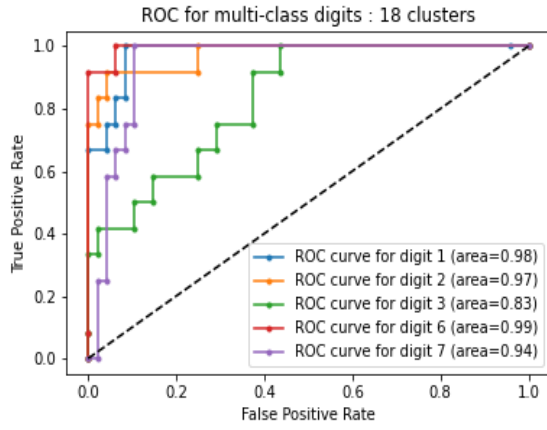


(a)

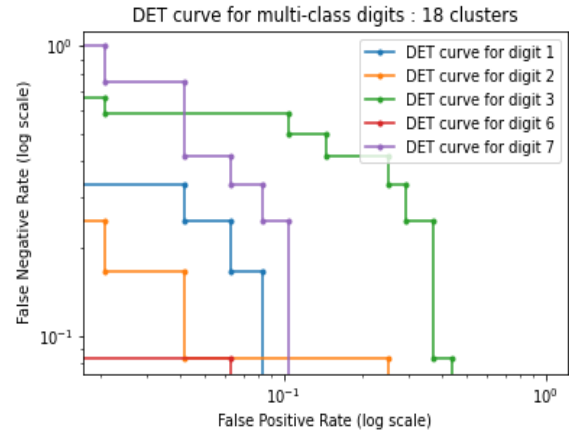


(b)

Figure 4: ROC and DET curves for 16 clusters

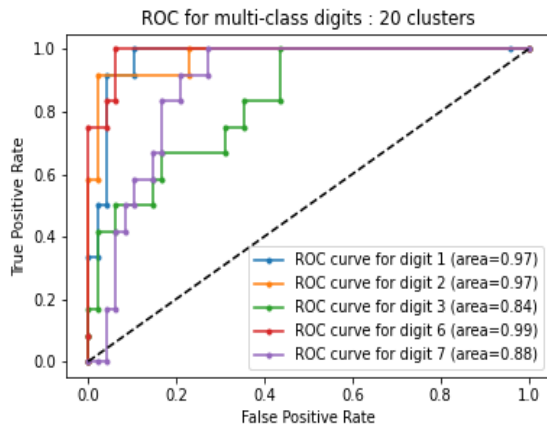


(a)

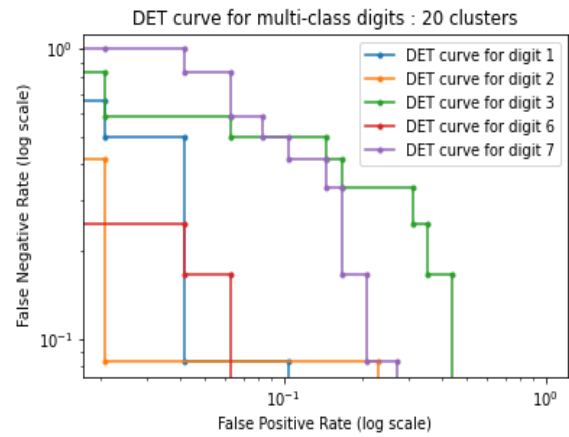


(b)

Figure 5: ROC and DET curves for 18 clusters



(a)



(b)

Figure 6: ROC and DET curves for 20 clusters

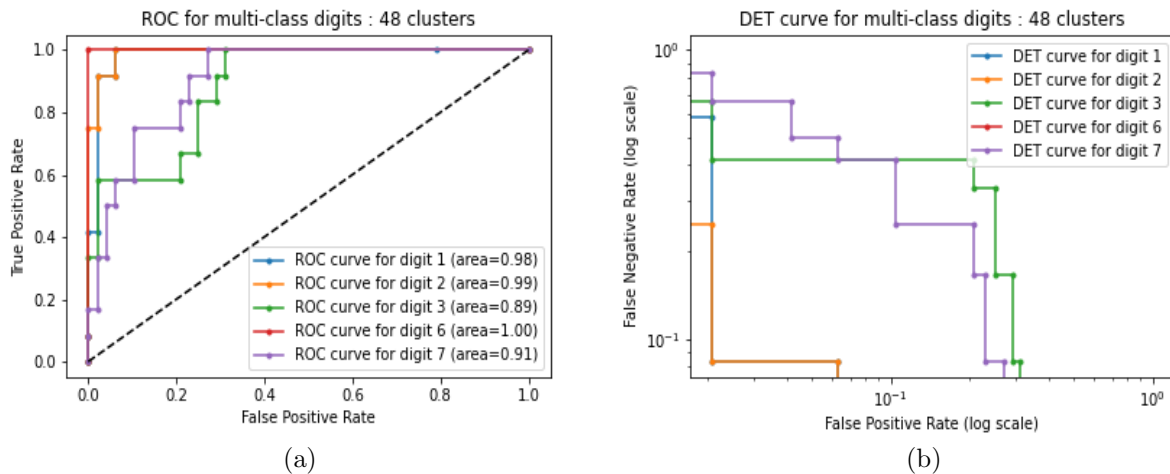


Figure 7: ROC and DET curves for 48 clusters

Problem 2

Use the HMMs trained in task 1 to recognize continuous digits. You need to concatenate the HMMs trained in task 1 to recognize continuous digits. Use only the given features.

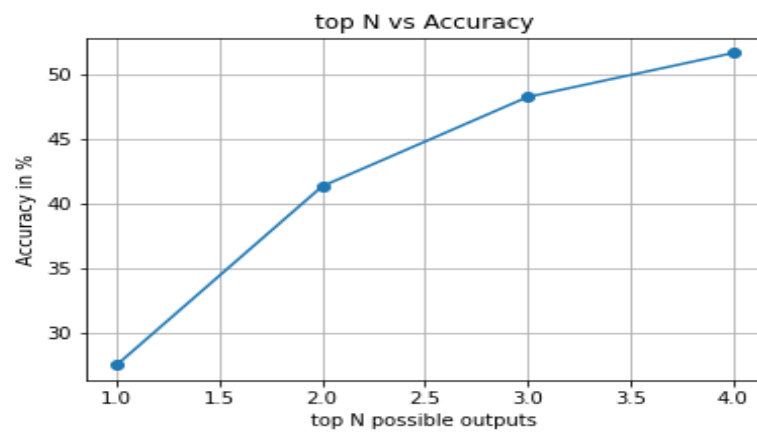
Methods and Results

We take the Isolated digit models from the above task and concatenate them to perform this task. The HMM models of all possible 2-digit and 3-digit numbers using $\{1,2,3,6,7\}$ such as $\{11,12,...,76,77,111,112,...,776,777\}$ (150 such models) have been modeled by adding a transition probability of 0.85 and 0.15 for state transition from one digit to another. Several other probabilities (0.9,0.75 etc.) were tried but 0.85 turned out to be the most accurate. The testing is similar to the previous case where we obtain VQ for all the test files and run them separately for each of the 150 models and take argmax over all the possible models and obtain the results. Since we are modeling continuous speech using Isolated models concatenation the accuracy of the model is pretty low (28 %). However of we take the top 3 results for each file then we get an accuracy of around 48.3 %. The output for variation of the accuracy with number of possible outputs is shown in Figure 8. For the test data given, we give the top 3 results obtained for each file.

1.mfcc	2.mfcc	3.mfcc	4.mfcc	5.mfcc
666	726	237	27	176
626	72	337	776	171
676	727	332	77	116

Visualisations

The plot for top N results vs accuracy for the continuous HMM model has been shown below. The top 3 result gives an accuracy of approximately 50 %



(a)

Figure 8: Top N results vs Accuracy

End of Report