CS669 Pattern Recognition

Assignment II

Rajesh R
S21005
21 November 2022

INDIAN LANGUAGE IDENTIFICATION SYSTEM

DATASET:

The IIT Mandi LID dataset which has audio data from Prasar Bharati (PB) and from YouTube (YT). There are 12 languages to be classified. For each class, the data in PB train is used to train the model. There are two test sets for each class: PB test and YT test. MFCCs feature is used.

Goal is to build two LID systems (System I and II) with GMMs and UBM-GMM and compare the performance.

SYSTEM I:

This system uses a GMM to model each class conditional density. This is done using the EM algorithm.

SYSTEM II:

This is a UBM-GMM system. Pooled the data of all classes to form a large GMM, called the universal background model (UBM.) From the UBM, class-specific GMMs are built using MAP adaptation. Only the means are adapted, and other parameters (Σk, πk) are used as such from the UBM.

Given the training data for class c as Xc = {x1, x2, . . . , xT }, adapt the UBM to get MAP estimates of the mean vectors as :

$\tilde{\mu}_k = \alpha_k \tilde{x}_k + (1 - \alpha_k)\mu_k$

where $\alpha_k = \dfrac{N_k}{N_k + r}$

Here, $\tilde{x}_k$ is the partial estimate of the mean vector using Xc, as in the E-step of the EM algorithm, Nk is the effective number of examples from the kth component using Xc, μk is the mean from the UBM, and r is a relevance factor, which is taken as 0.7.

DATA USED & CLASS INFORMATION

CLASS DETAILS:

There are 12 languages to be classified. There is a variable number of audio examples and of different length.

Number of data in each class

Class	Labels	Train	PB Test	YouTube Test
Assamese	0	719	359	180
Bengali	1	379	179	180
English	2	244	116	126
Gujarati	3	358	179	181
Hindi	4	359	179	181
Kannada	5	391	197	181
Malayalam	6	398	196	180
Marati	7	241	117	119
Odissa	8	398	199	190
Punjabi	9	242	124	121
Tamil	10	241	125	117
Telugu	11	388	194	178

There are total 4,358 training examples of all languages.

IMPORTANT NOTE:

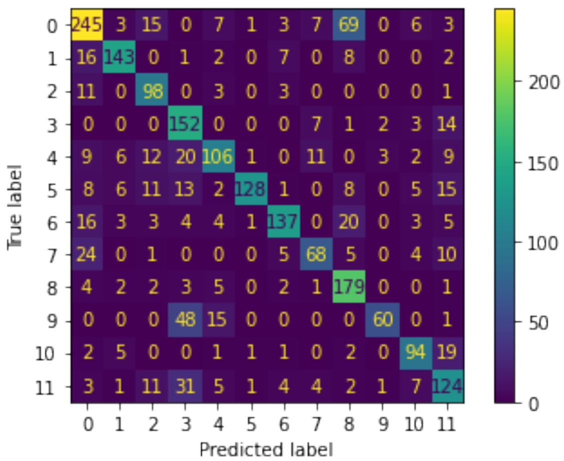
The classifier is trained using the PB train data. It is tested using PB test data and real world YouTube test data.

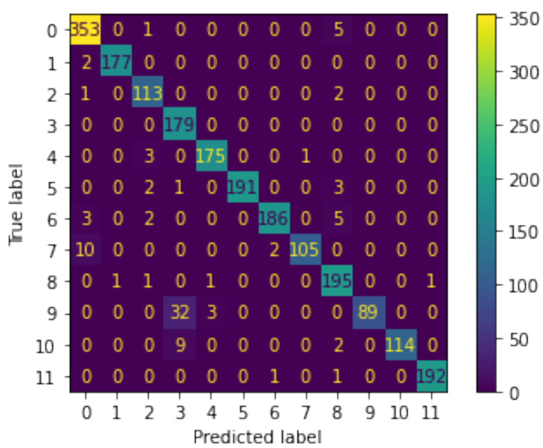
OBSERVATIONS FROM THE SYSTEM OUTPUT

1) Which system performs better and why?

System II: UBM - GMM model works better than the traditional model in terms of accuracy and/or confusion matrix. The performance of this system is better compared to the traditional GMM - System I.

Performance Parameters

<table>
<tr><th>System I</th><th>System II</th></tr>
<tr><td colspan="2">PB Test Data Results</td></tr>
<tr><td colspan="2">1) Accuracy</td></tr>
<tr><td>
Accuracy: 0.7088724584103512
Recall: 0.7088724584103512
Precision: 0.7368703582313847
F1 Score: 0.7091140212315188
</td><td>
Accuracy: 0.961828282827968
Recall: 0.9667879622001
Precision: 0.9701989827961
F1 Score: 0.9596857670979667
</td></tr>
<tr><td colspan="2">2) Confusion Matrix</td></tr>
<tr><td>


</td><td>


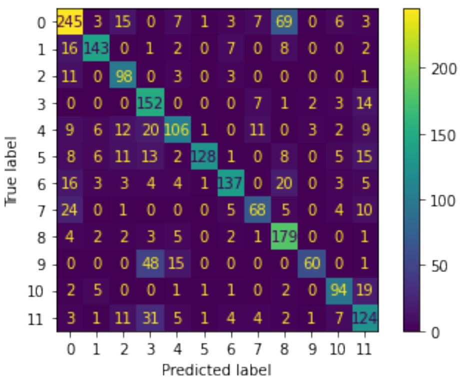
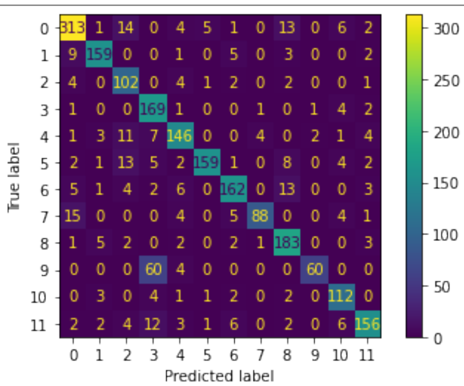
</td></tr>
</table>

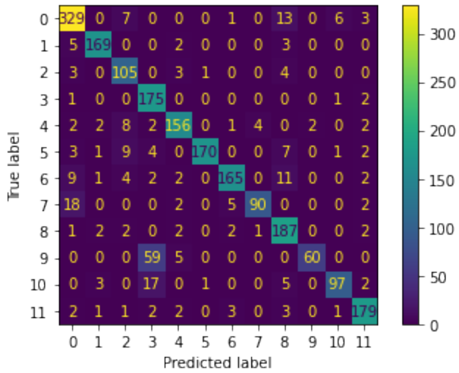
Due to the tighter coupling between the UBM and the trained models the performance of System II are better than the decoupled traditional GMM System I. Due to the use of coupled approaches, the performance is not affected by unseen language events i.e., when an unseen language event occurs the mixture parameters of that unseen language classes are directly copied from the UBM. So the during the testing phase unseen acoustic event produces almost zero likelihood ratio which does not contribute evidence either towards or away from the hypothesized language. Though GMM represents a distribution over a large space but a single vector can influence only a few components of the GMM, adaptation and estimation of likelihood are be done faster by considering best scoring mixture components among all the components.

2) How does performance vary with the number of mixtures in the GMM? Give a meaningful plot.

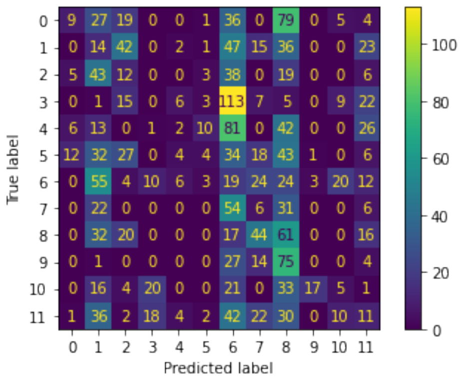
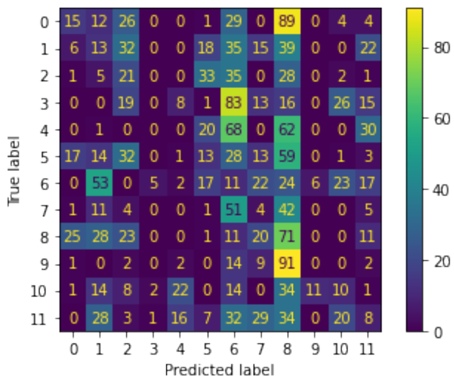
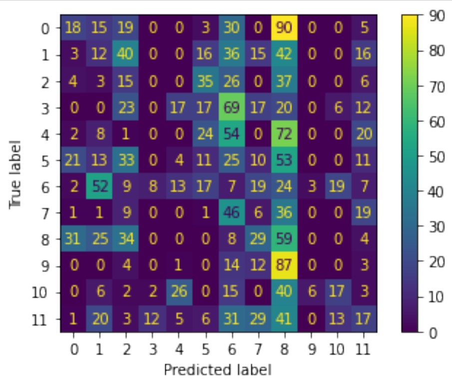
- Performance of the system increases greatly when the number of mixtures in the GMM increased.
- Since addition of two gaussian is more gaussian, the model performance increased when we increase the number of components.

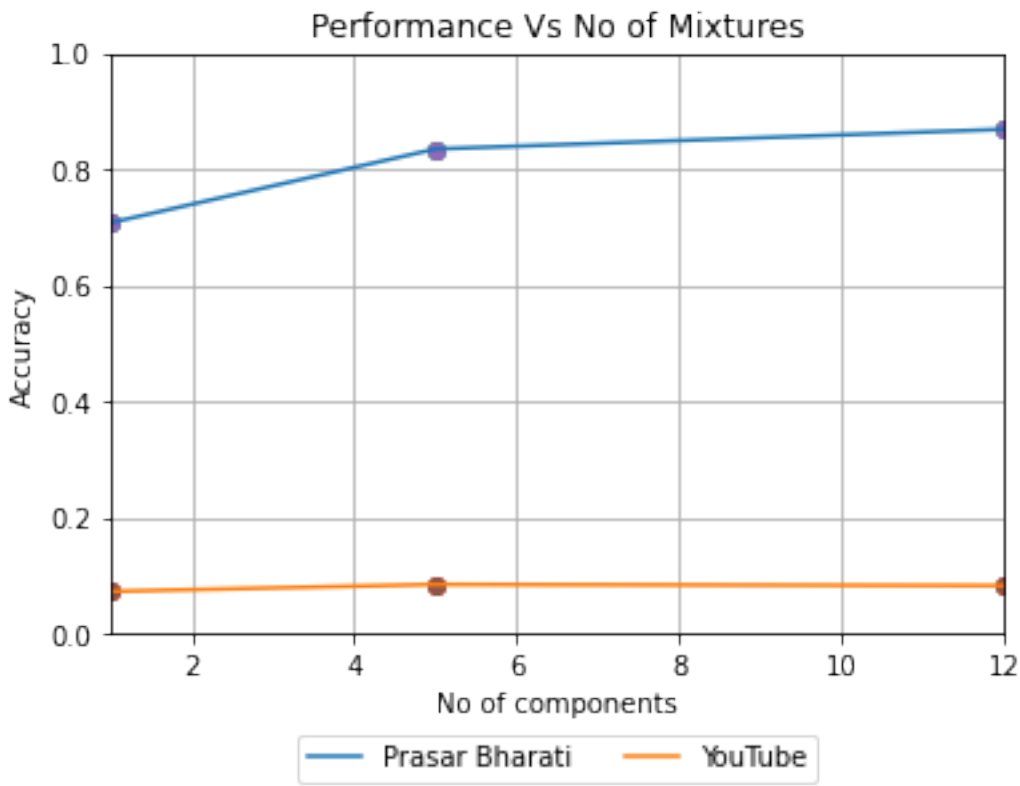
RESULTS ON PB TEST DATA

<table>
<tr><th>MIXTURES</th><th>CONFUSION MATRIX</th><th>PARAMETERS</th></tr>
<tr><td>1</td><td>
</td><td>Accuracy: 0.7088724584103512
Recall: 0.7088724584103512
Precision: 0.7368703582313847
F1 Score: 0.7091140212315188</td></tr>
<tr><td>5</td><td>
</td><td>Accuracy: 0.8359519408502772
Recall: 0.8359519408502772
Precision: 0.8524191547249562
F1 Score: 0.8351577721951564</td></tr>
</table>

MIXTURES	CONFUSION MATRIX	PARAMETERS
12		Accuracy: 0.8696857670979667 Recall: 0.8696857670979667 Precision: 0.8850499972075281 F1 Score: 0.8681031374893899

RESULTS ON YOUTUBE TEST DATA

MIXTURES	CONFUSION MATRIX	PARAMETERS
1		Accuracy: 0.07394002068252327 Recall: 0.07394002068252327 Precision: 0.08880528284666632 F1 Score: 0.060329694702621325
5		Accuracy: 0.08583247156153051 Recall: 0.08583247156153051 Precision: 0.07603130785254986 F1 Score: 0.06990226690175955
12		Accuracy: 0.08376421923474664 Recall: 0.08376421923474664 Precision: 0.0859618693082005 F1 Score: 0.07447735785271418

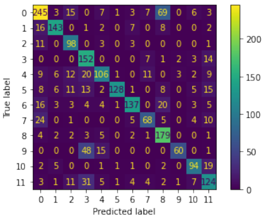
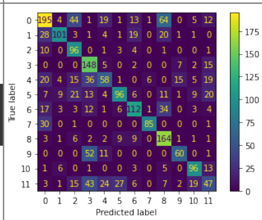
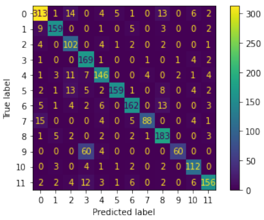
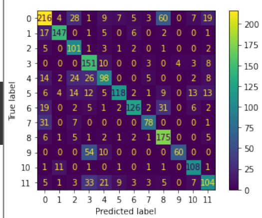
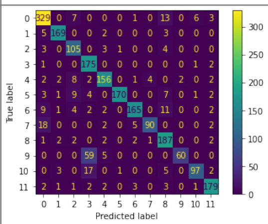
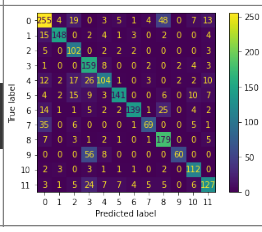


The plot shows the Accuracy vs Number of mixture in GMMs. It is clearly seen that as the number of mixtures increase, the accuracy increases and tends towards 1 and saturates at certain value.

3) Is it better to use a full covariance matrix or a diagonal covariance matrix in the GMM?

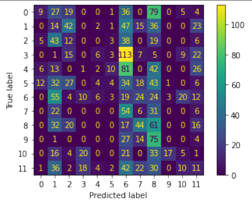
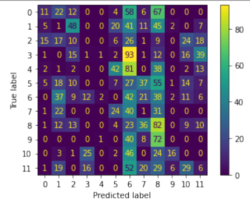
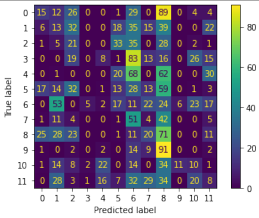
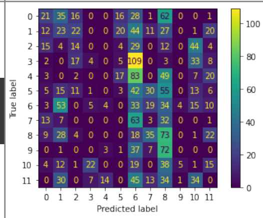
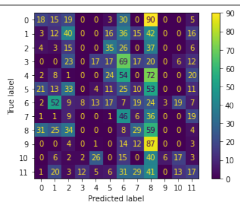
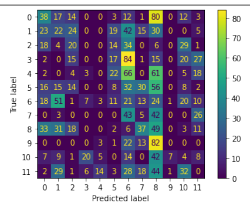
- It is better to use full covariance matrix than a diagonal covariance matrices in the GMM. The performance decreased when diagonal covariance is used in the model.
- Unlike other types of covariances(Diagonal/Symmetry), Full covariance matrix allows for correlation between our two random variables.

COMPARISON B/W FULL & DIAGONAL COVARIANCES

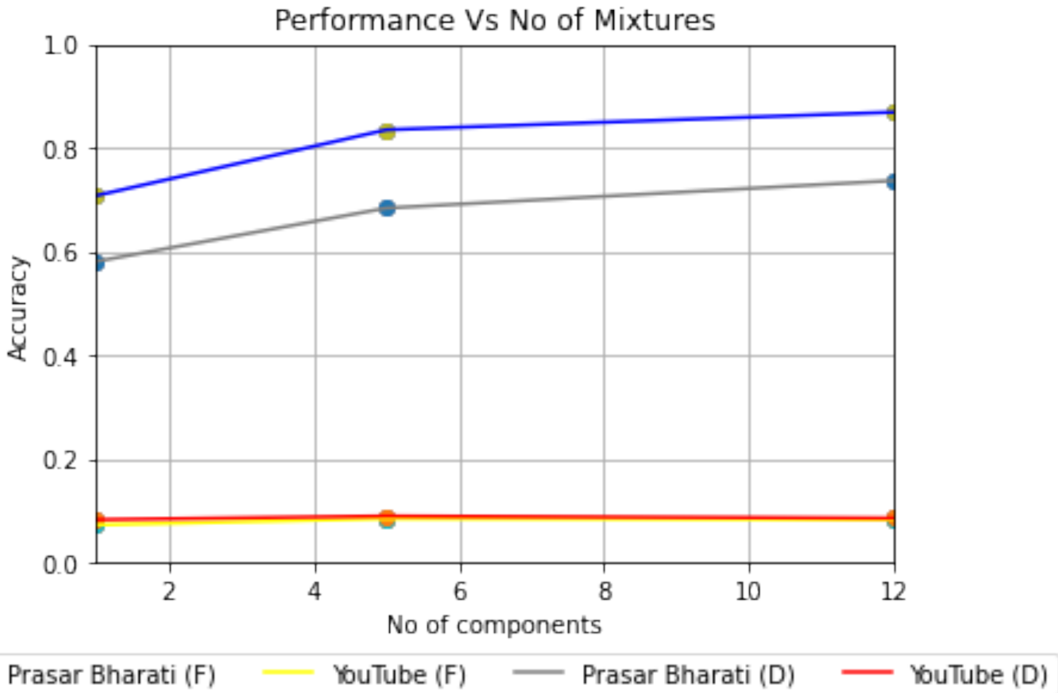
PB TEST DATA				
Mixtures	Full Covariance (F)		Diagonal Covariance (D)	
	Confusion Matrix	Parameters	Confusion Matrix	Parameters
1		Accuracy: 0.7088724584103512 Recall: 0.7088724584103512 Precision: 0.7368703582313847 F1 Score: 0.7091140212315188		Accuracy: 0.5813308687615527 Recall: 0.5813308687615527 Precision: 0.597804233811644 F1 Score: 0.5734355800029346
5		Accuracy: 0.8359519408502772 Recall: 0.8359519408502772 Precision: 0.8524191547249562 F1 Score: 0.8351577721951564		Accuracy: 0.6848428835489834 Recall: 0.6848428835489834 Precision: 0.714919420499643 F1 Score: 0.6837436183748604
12		Accuracy: 0.8696857670979667 Recall: 0.8696857670979667 Precision: 0.8850499972075281 F1 Score: 0.8681031374893899		Accuracy: 0.7370609981515711 Recall: 0.7370609981515711 Precision: 0.7648625324914246 F1 Score: 0.7364953005976088

Accuracy Improvements

Mixtures	Full Covariance	Diagonal Covariance
1	70.88%	58.13%
5	83.59%	68.48%
12	86.96%	73.70%

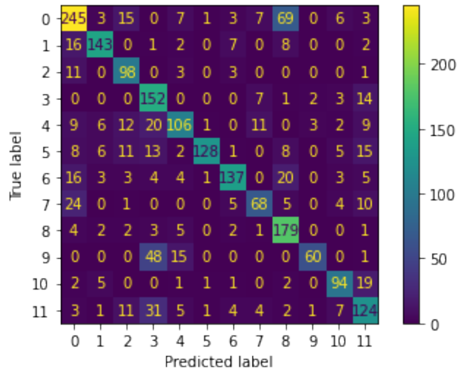
YOUTUBE TEST DATA				
Mixtures	Full Covariance (F)		Diagonal Covariance (D)	
	Confusion Matrix	Parameters	Confusion Matrix	Parameters
1		Accuracy: 0.07394002068252327 Recall: 0.07394002068252327 Precision: 0.08880528284666632 F1 Score: 0.06032969470262l325		Accuracy: 0.08324715615305067 Recall: 0.08324715615305067 Precision: 0.06661280137698586 F1 Score: 0.05844540182373808
5		Accuracy: 0.08583247156153051 Recall: 0.08583247156153051 Precision: 0.07603130785254986 F1 Score: 0.06990226690175955		Accuracy: 0.0904860392967942 Recall: 0.0904860392967942 Precision: 0.07924150298897438 F1 Score: 0.07139847919633592
12		Accuracy: 0.08376421923474664 Recall: 0.08376421923474664 Precision: 0.0859618693082005 F1 Score: 0.07447735785271418		Accuracy: 0.08634953464322648 Recall: 0.08634953464322648 Precision: 0.07071313055598412 F1 Score: 0.0729122188743797

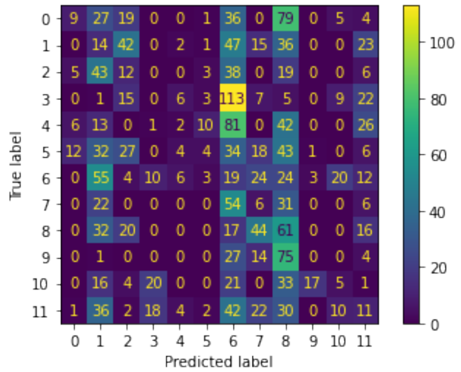
Overall performance vs number of mixtures in GMMs are plotted below. It is clear that performance increase when we increase the number of mixtures and when we use full covariance matrix.

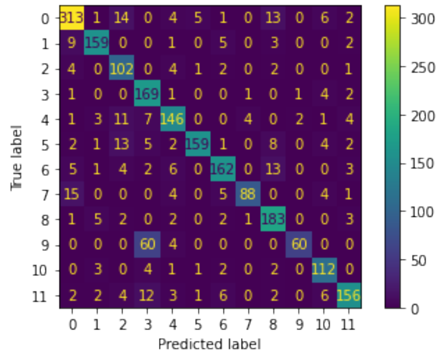


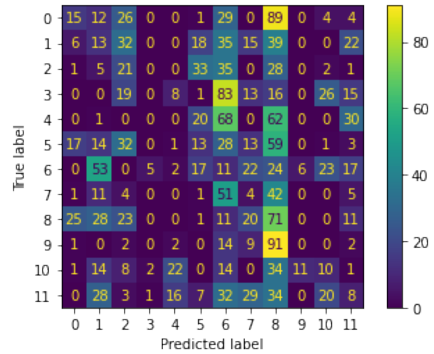
4) Compare the performance on PB test and on YT test. Why is there a difference?

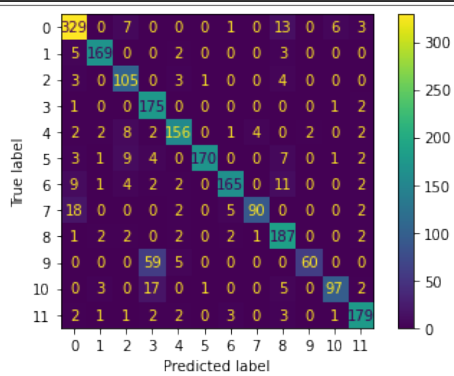
There is a vast difference in the system prediction for YouTube test data. Comparison is shown below.

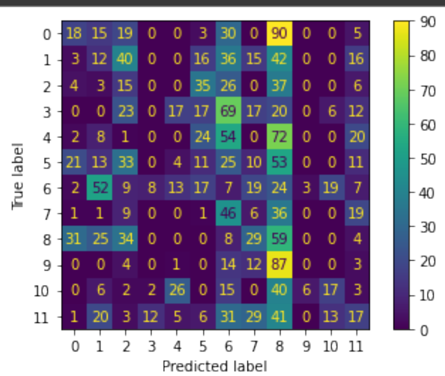
<table>
<tr><th>Mixtures</th><th>PB Test Results</th><th>YouTube Test Results</th></tr>
<tr><td>1</td><td>

Accuracy: 0.7088724584103512
Recall: 0.7088724584103512
Precision: 0.7368703582313847
F1 Score: 0.7091140212315188</td><td>

Accuracy: 0.07394002068252327
Recall: 0.07394002068252327
Precision: 0.08880528284666632
F1 Score: 0.060329694702621325</td></tr>
<tr><td>5</td><td>

Accuracy: 0.8359519408502772
Recall: 0.8359519408502772
Precision: 0.8524191547249562
F1 Score: 0.8351577721951564</td><td>

Accuracy: 0.08583247156153051
Recall: 0.08583247156153051
Precision: 0.07603130785254986
F1 Score: 0.06990226690175955</td></tr>
<tr><td>12</td><td>

Accuracy: 0.8696857670979667
Recall: 0.8696857670979667
Precision: 0.8850499972075281
F1 Score: 0.8681031374893899</td><td>

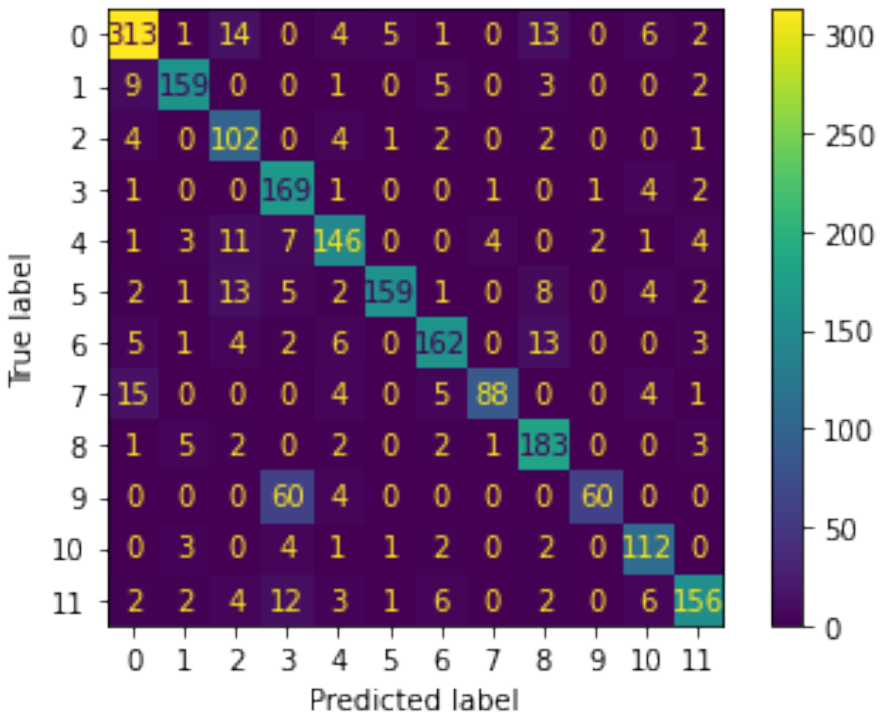
Accuracy: 0.08376421923474664
Recall: 0.08376421923474664
Precision: 0.08596186930820054
F1 Score: 0.07447735785271418</td></tr>
</table>

The model is trained using the PB Training data alone. It is to be noted and assumed that the PB training data and test data comes from the same distribution. Same recording setup or etc. Since the training is done in PB data, the test results on PB outputs better when compared to other data.

Real world data like YouTube audio will have very high varying data distribution. The model performs really bad in predicting those languages since it is not trained on such data.

5) Which languages are confusable and why?

For reference, let us consider the confusion matrix with number of mixing equal to 5 on PB test data.



0	1	2	3	4	5	6	7	8	9	10	11
Asm	Ben	Eng	Guj	Hin	Kan	Mal	Mar	Odi	Pun	Tam	Tel

As per the results obtained from System I,

- Extremely confusing languages are **Punjabi** and **Gujarati**. The data of Punjabi are misclassified as Gujarati.
- Following above, **Marathi** and **Assamese** were highly misclassified. Almost 15 data of Marathi is misclassified as Assamese.
- Next, 14 files of Assamese is misclassified with the language English. **Assamese** and **English** also have higher rate here.
- **Assamese** and **Odissa**, **Kannada** and **English**, **Malayalam** and **Odissa**, **Telugu** and **Gujarati** are also having higher misclassification rate.

Owing to decent from the same origin, most of the Indian languages have overlapping phoneme sets. Despite the similarity in the phoneme sets, every language has its influence on the phonotactic constraints of that language. Causing confusion in the classification.

OVERCOME BY:

For discriminating a language using its Phonotactic information in the presence of similar phoneme sets need a large amount of training data for developing a language model. The modelling technique should have a large number of mixture components to account for the slight variation in Phonotactics imparted by the language. GMM-UBM can be used to develop the language models. In GMM-UBM modelling certain amount of data from all the classes is pooled to build a universal background model with a large number of mixture components and this UBM model is adapted to all the classes. By this, a system with a large number of mixture components can be developed though data in each class is inadequate to support a large model independently.

REFERENCE

"Significance of GMM-UBM based Modelling for Indian Language Identification" by Ravi Kumar V., Hari Krishna Vydana and Anil Kumar Vuppala - Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015)

CODE LINK

Github Link

System I	https://github.com/its-rajesh/Pattern-Recognition/blob/main/PRA2Q1.ipynb
System II	https://github.com/its-rajesh/Pattern-Recognition/blob/main/PRA2Q2.ipynb

Google Colab Link

System I	https://colab.research.google.com/drive/1gPL3G1Kvn1n-KbtFF3Mz39lVlhGbm-Pf?usp=share_link
System II	https://colab.research.google.com/drive/1p6pc2Mwjo7ghC3HGKQ3BtaTAGKKJfKqA?usp=share_link