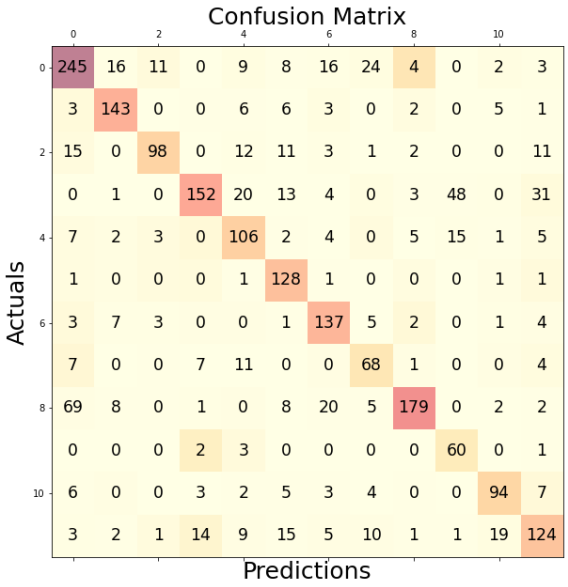
CS669
Assignment 2

REPORT

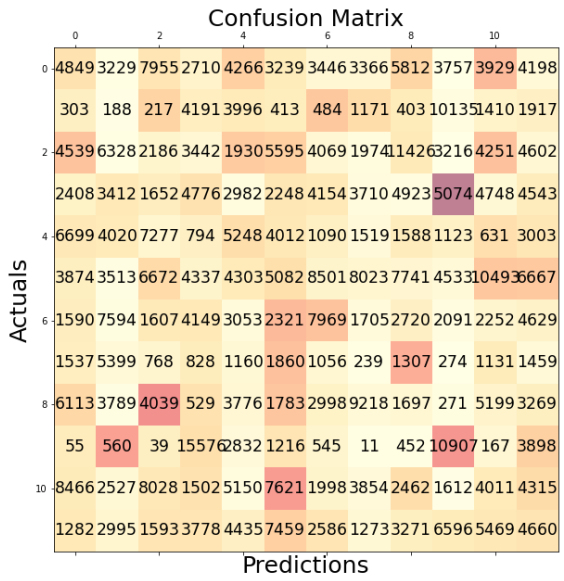
Ayush Dwivedi
Roll No. V21093

Assignment Report

1. System 1 has come out to be better than system 2 where system is applying GMM on individual language class and system is forming UBM-GMM on complete language data at once.



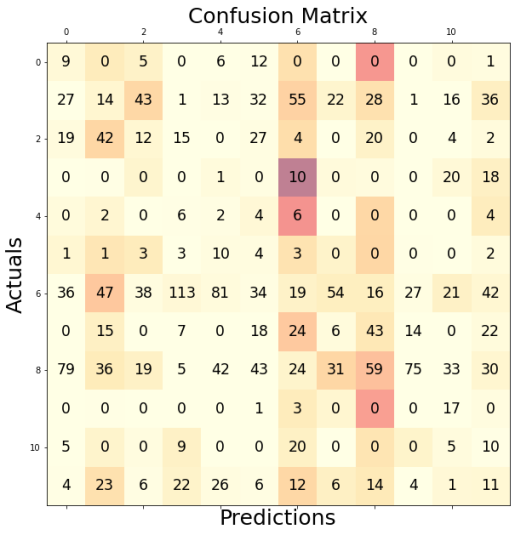
Confusion Matrix for Prasar Bharati Data. In class wise GMM.



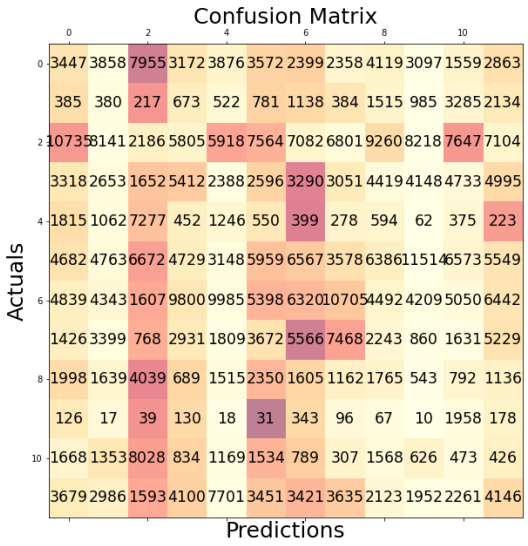
Confusion Matrix for Prasar Bharati Data in UBM-GMM

Accuracy for Prasar Bharati Data from system1 is 70%.

Accuracy for Prasar Bharati Data from System2 is 9%.

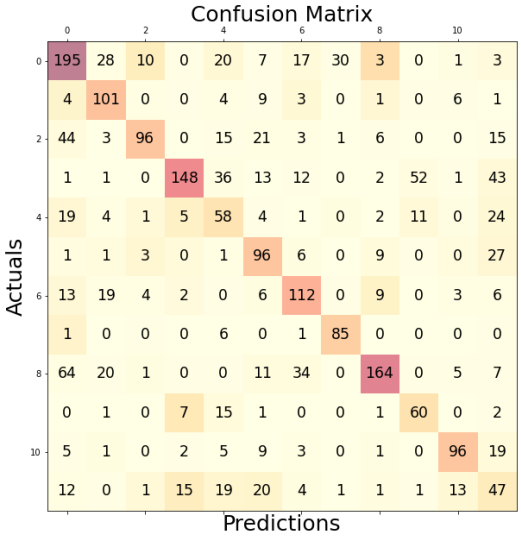


Confusion Matrix for Youtube Data Through System1

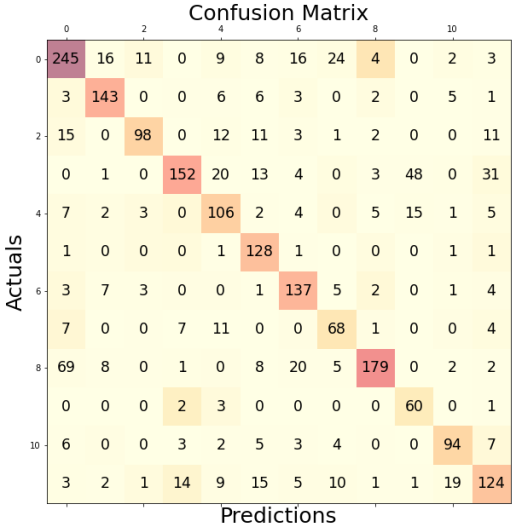


Confusion matrix for Youtube Data through system 2.

3.GMM with full covariance matrix gives better results than GMM with diagonal covariance matrix.



Confusion matrix with diagonal covariance

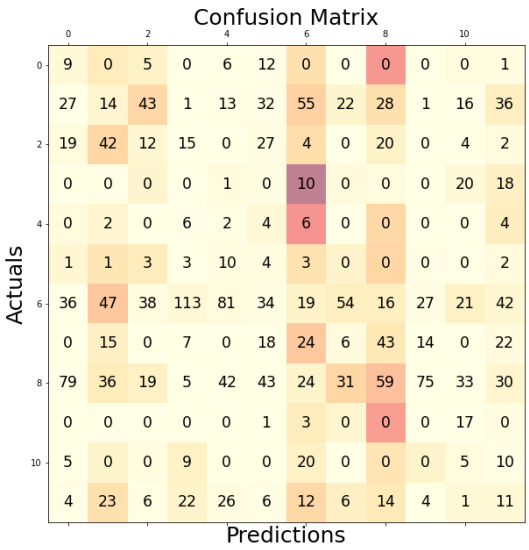


Confusion matrix with Full covariance.

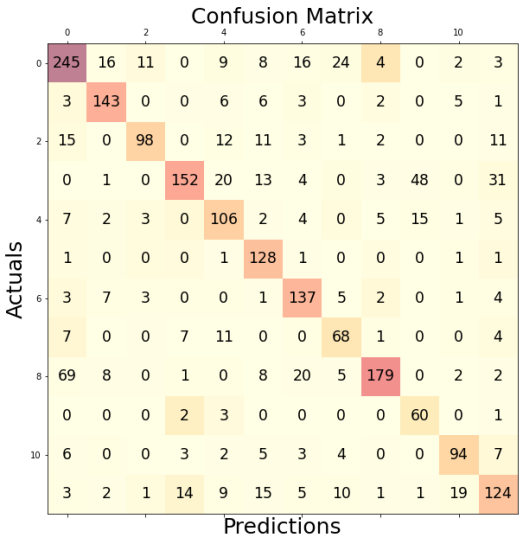
Accuracy of GMM with diagonal covariance matrix is 58%.

Accuracy of GMM with full covariance matrix is 70%.

4.Performance on PB_test is better than YT_test because we have trained the GMM with Prasar Bharati data.



YT_test Performance confusion matrix

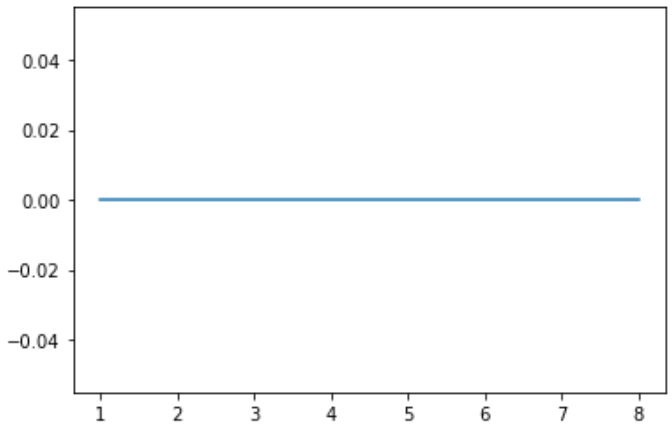


PB_test Performance confusion matrix.

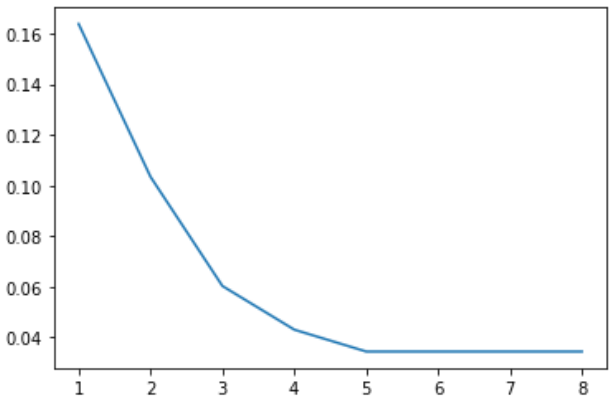
Accuracy of PB_test is 70%

Accuracy of YT_test is 8%.

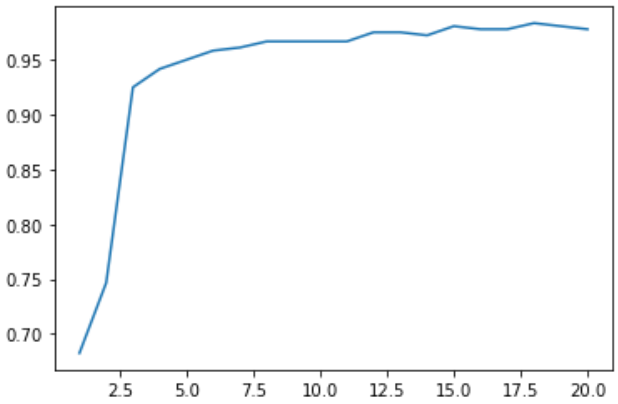
2. **Plots with change in number of clusters and the accuracy at that model for different language class are as follows.**



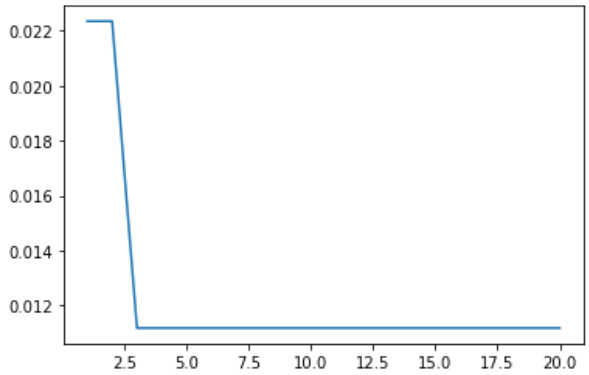
Bengali Class



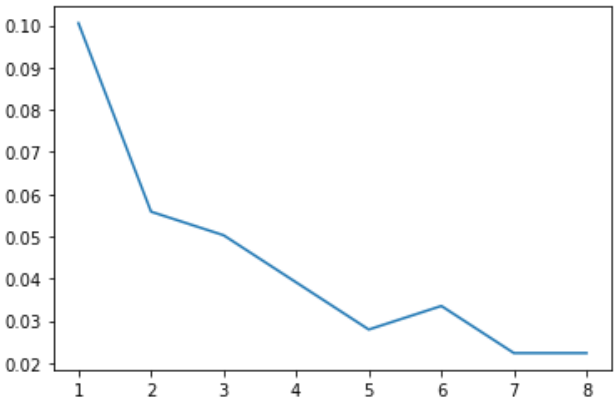
English Class



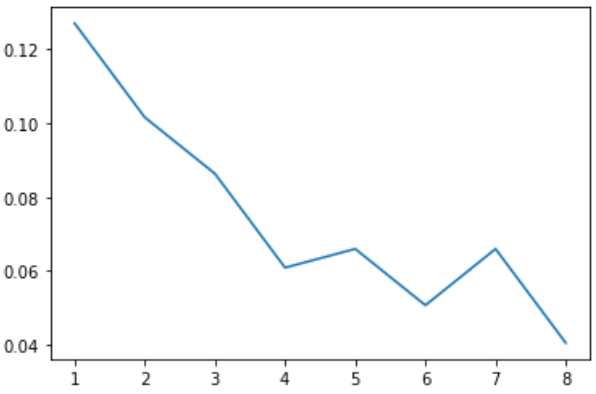
Assam Class



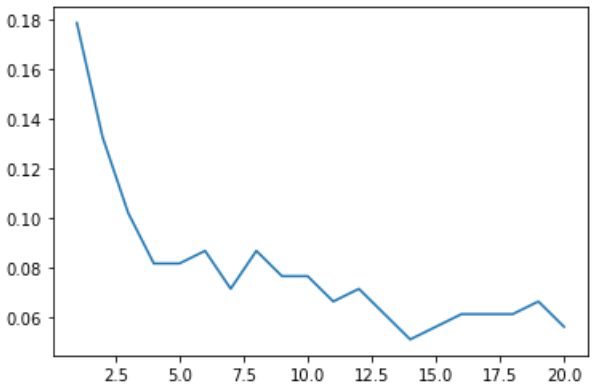
Gujarati Class



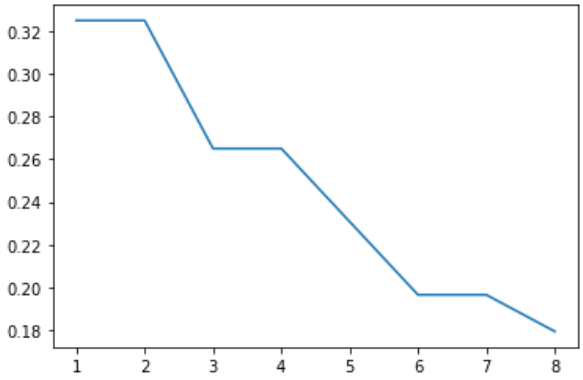
Hindi Class



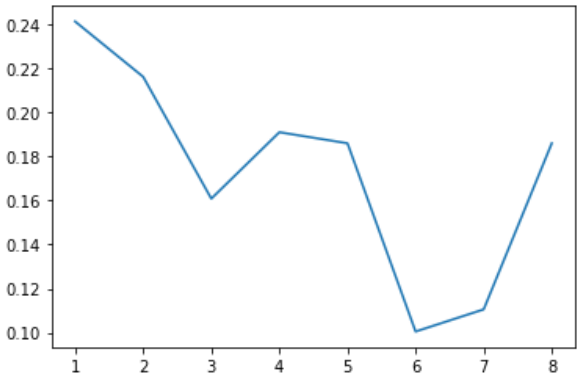
Kannad Class



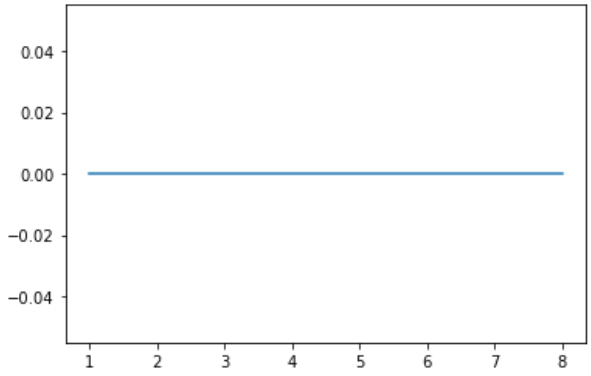
Malyalam Class



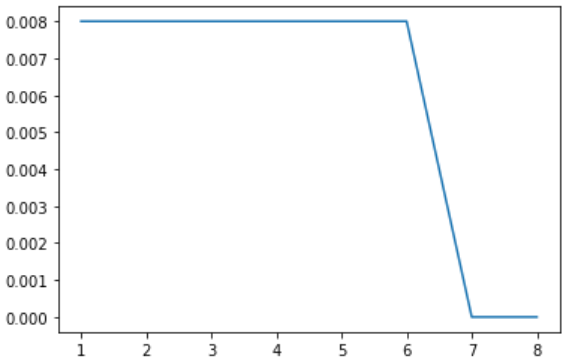
Marathi Class



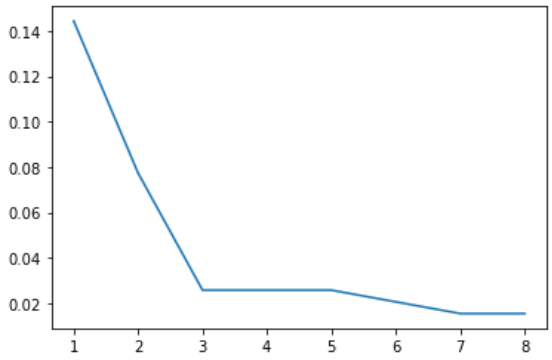
Odisi Class



Punjabi Class



Tamil Class



Telugu class

5. Languages which are confusable are Bengali and Punjabi because the their accuracy verses number clusters graph is constant at zero which means it has no accuracy at that value.

For the Google Colab Link .

Click Here