

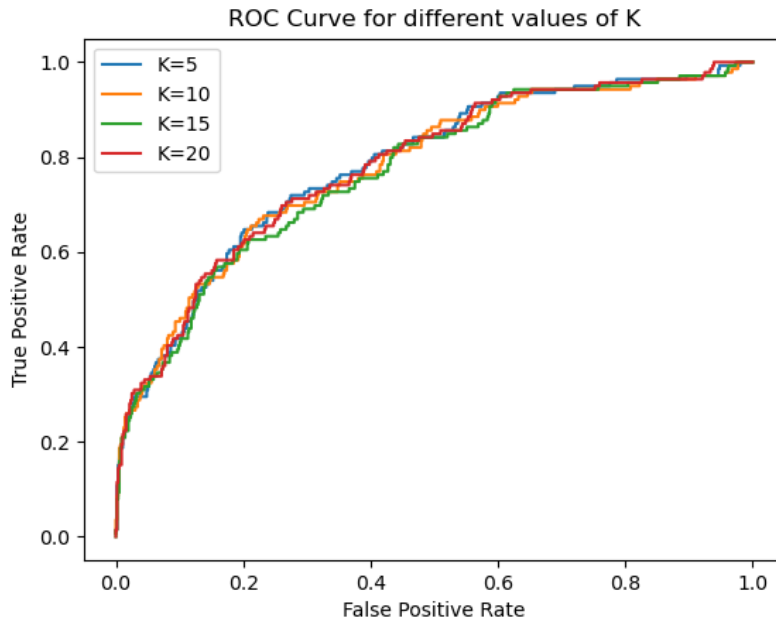
CS5691 - Assignment3

K-Means AND Gaussian Mixture Models

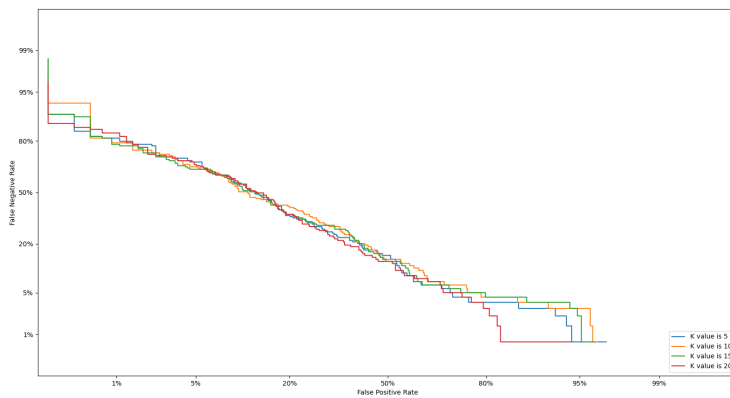
Real Data:-

ROC and DET curves for diagonal and nondiagonal:-

Varying number of clusters:-



(a) *ROC curve*



(b) *DET curve*

Inferences

1. Running Gaussian Mixture Model on the clusters calculated by k-means algorithm for small number of iterations increases accuracy very well.

2. Non diagonal covariance matrix gives better accuracy than diagonal covariance matrix for Gaussian Mixture Model since we are removing some relations between feature classes, making non diagonal elements zero, hence accuracy decreases.

Synthetic Data:- K-Means is a hard clustering unsupervised Technique. GMM is a soft clustering unsupervised technique.

Decision boundaries along with contours for k=8:-

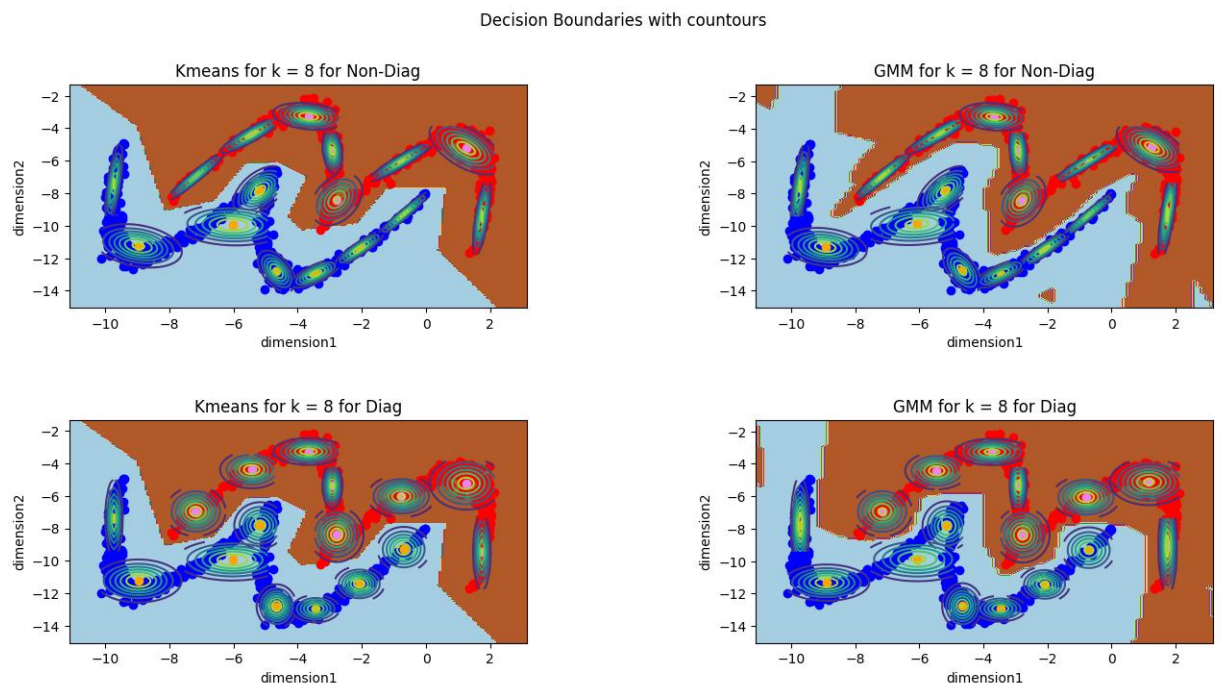
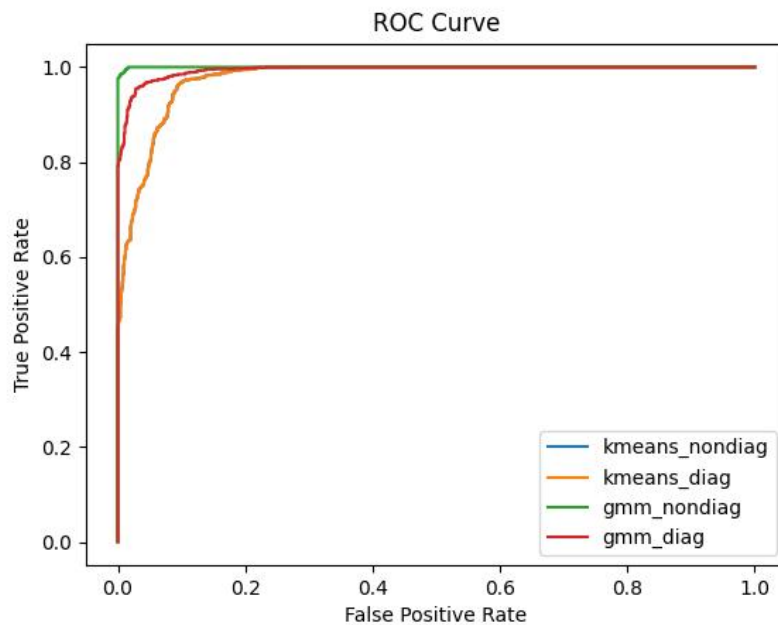
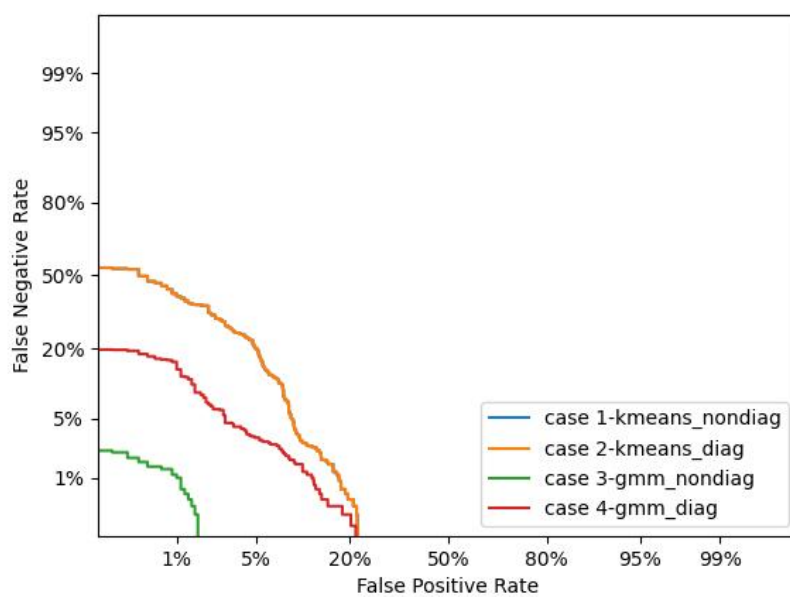


Figure 2: Decision Boundaries for K = 8 with contours

ROC and DET curves for diagonal and nondiagonal:-



(a) *ROCcurve*



(b) *DETcurve*

Nondiagonal:- Accuracy for K-Means :- 95.9 Accuracy for GMM :- 99.7

Diagonal:- Accuracy for K-Means :- 95.9 Accuracy for GMM :- 97.8

Inferences

1. Running Gaussian Mixture Model on the clusters calculated by k-means

algorithm for small number of iterations increases accuracy very well.

2. Non diagonal covariance matrix gives better accuracy than diagonal covariance matrix for Gaussian Mixture Model since we are removing some relations between feature classes, making non diagonal elements zero , hence accuracy decreases.

Best fit with maximum accuracy:-

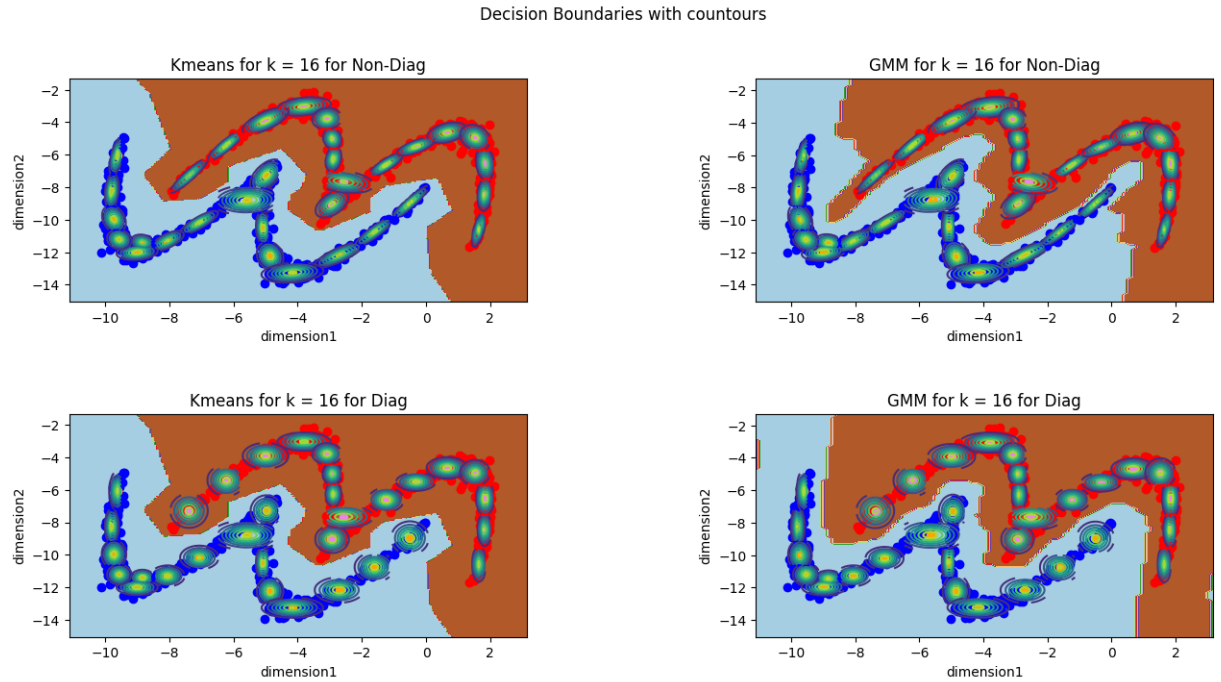
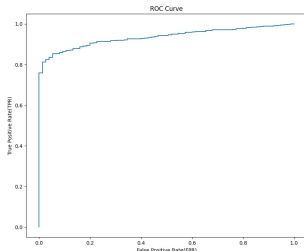
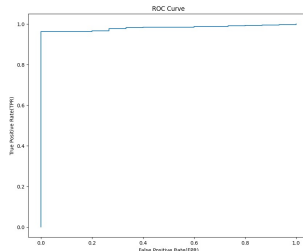


Figure 4: Decision Boundaries for K = 8 with contours

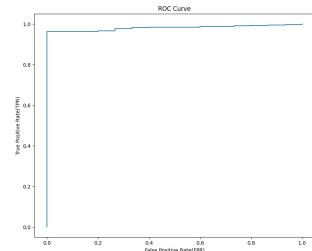
ROC variation for GMM with number of clusters:-



(a) ROCfork = 2



(b) ROCfork = 3

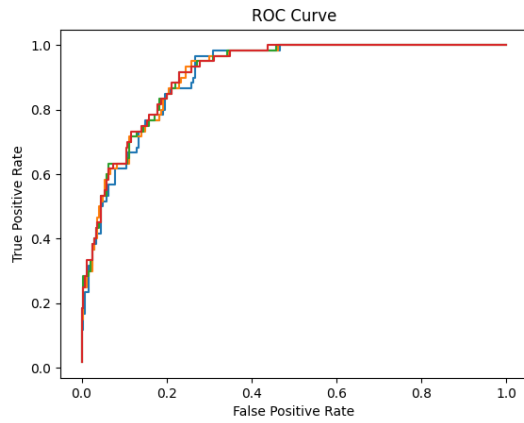


(c) ROCfork = 8

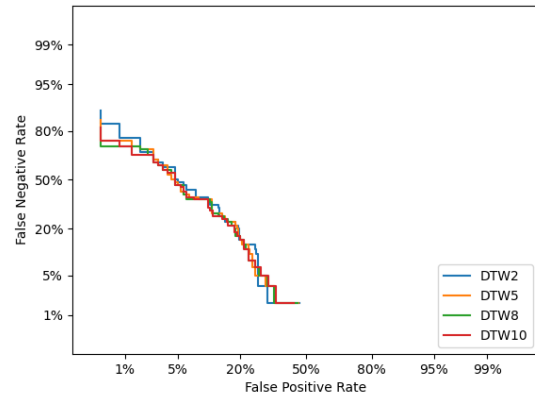
DTW and HMM

1) DTW for spoken digits:-

ROC and DET Curves



(a) ROC Curve



(b) DET curve

Confusion Matrix

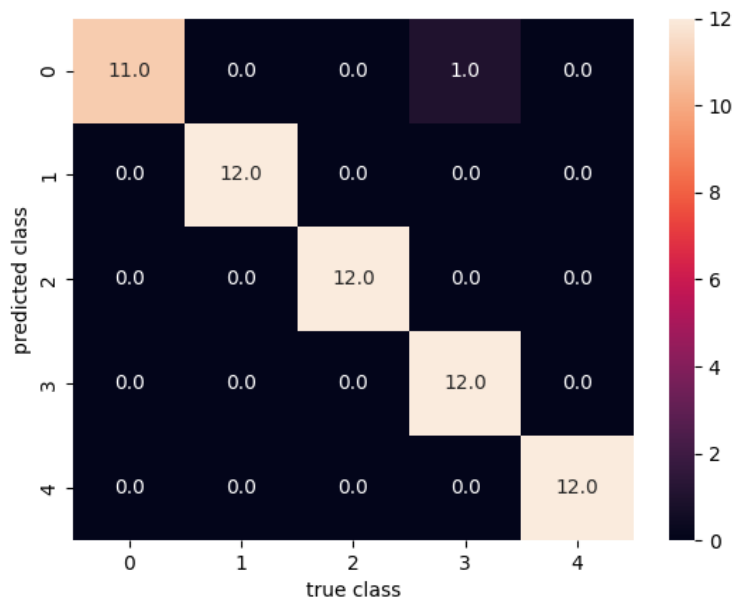


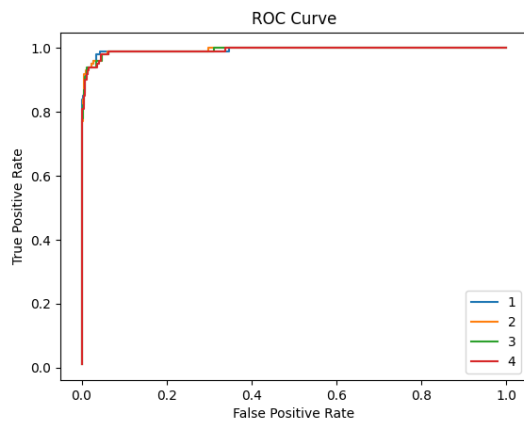
Figure 2: *ConfusionMatrix*

For these curves, we have varied number of clusters used for classification (namely $k = [2, 5, 8, 10]$) and plotted the ROC and DET curves.

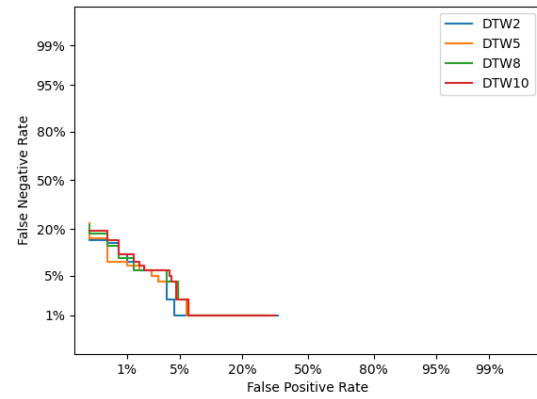
The accuracy for predicting correct spoken digit from test data set for 5 clusters is 98.33%

2) DTW for Hand written letters:-

ROC and DET Curves



(a) ROCcurve



(b) DETcurve

For hand written letters we have varied the number of clusters used in k-means algorithm , namely $k = [2, 5, 8, 10]$ and plotted the above ROC and DET curves.

The accuracy for recognising hand written letter test data by taking 5 clusters in k-means algorithm is 100%

Confusion Matrix

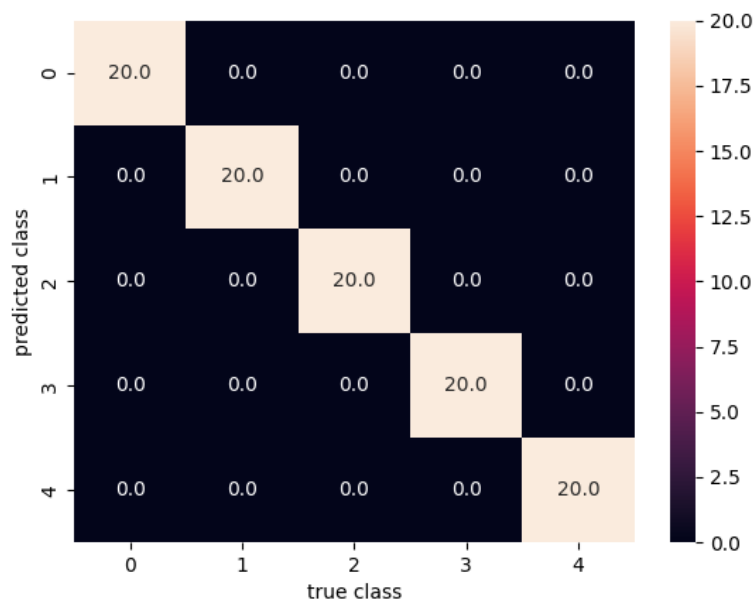


Figure 4: ConfusionMatrix

HMM

3) HMM for Spoken digits:-

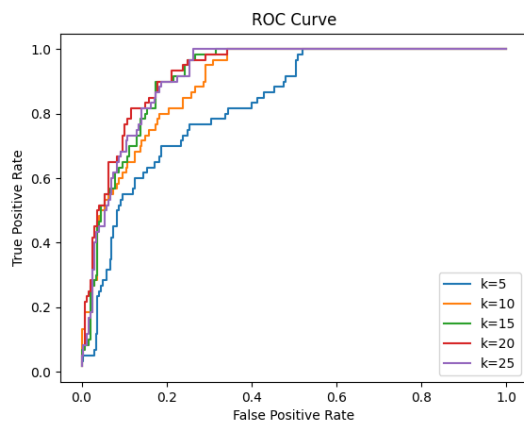
Varying number of input symbols:-

1) Dependence of Accuracy on number of symbols :-

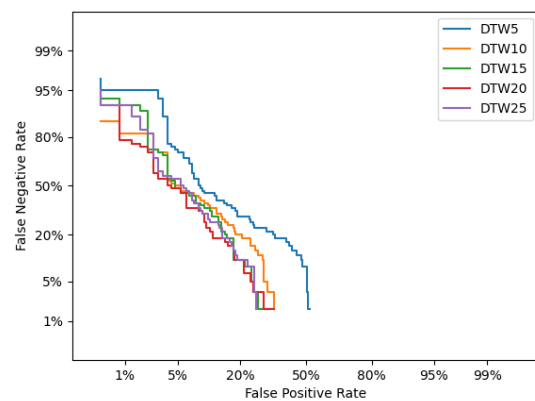
number of symbols	Accuracy
$k = 5$	56.67
$k = 10$	96.67
$k = 15$	98.33
$k = 20$	95.0
$k = 25$	93.33

Inference :- Therefore for number of input symbols = 15 , hidden states = 3, we are getting maximum accuracy.

2) ROC and DET Curves



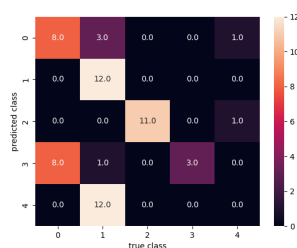
(a) ROC curve



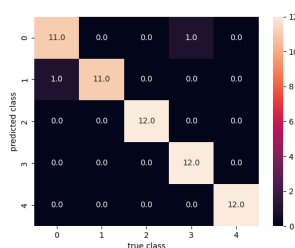
(b) DET curve

Varying number of emission symbols as 5,10,15,20,25 and keeping number of hidden states as 3

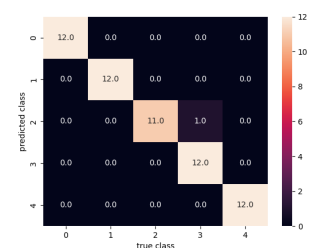
3) Confusion Matrix



(a) numsymbols = 5



(b) numsymbols = 10



(c) numsymbols = 15

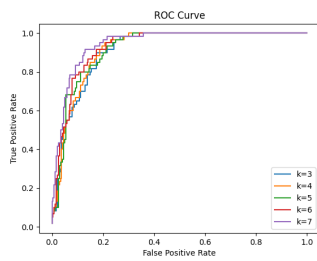
Varying number of hidden states:-

1) Dependence of Accuracy on number of hidden states :-

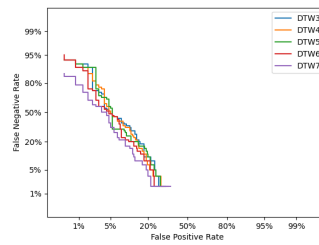
number of symbols	Accuracy
$k = 3$	98.34
$k = 4$	100
$k = 5$	98.34
$k = 6$	95.0
$k = 7$	98.34

Inference :- Therefore for number of symbols = 15, hidden states = 4 , we are getting maximum accuracy.

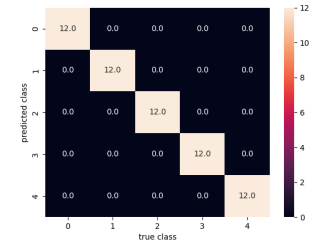
2) ROC and DET Curves



(a) ROC curve



(b) DET curve



(c) confusion matrix for $k = 4$

3) HMM for HandWritten characters:-

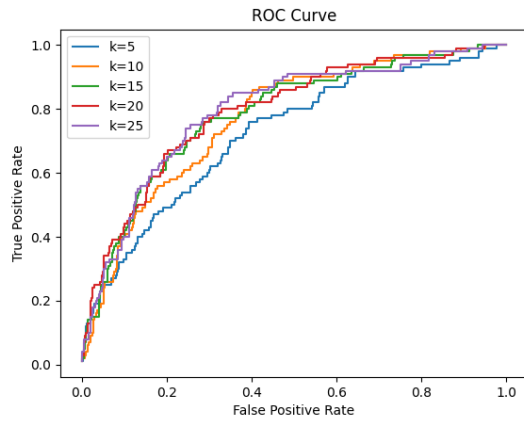
Varying number of input symbols:-

1) Dependence of Accuracy on number of symbols :-

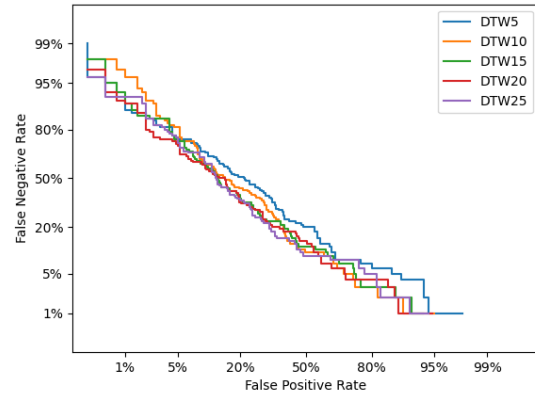
number of symbols	Accuracy
$k = 5$	49
$k = 10$	77
$k = 15$	87
$k = 20$	80
$k = 25$	80

Inference :- Therefore for number of input symbols = 15 , hidden states = 3, we are getting maximum accuracy.

2) ROC and DET Curves

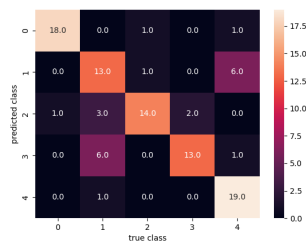


(a) ROC curve

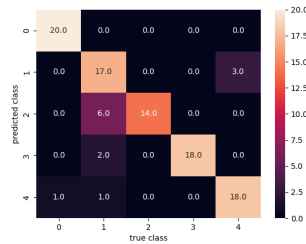


(b) DET curve

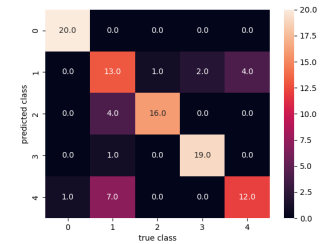
3) Confusion Matrix



(a) numsymbols = 10



(b) numsymbols = 15



(c) numsymbols = 20

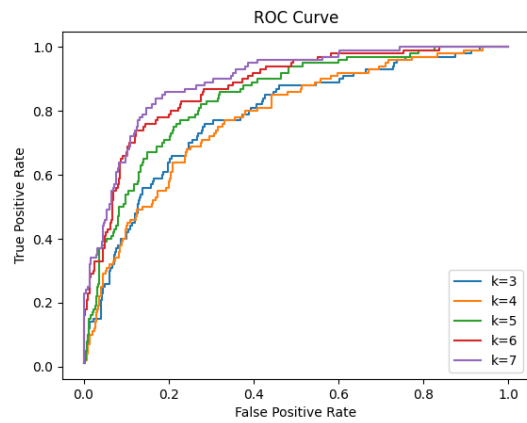
Varying number of hidden states:-

1) Dependence of Accuracy on number of hidden states :-

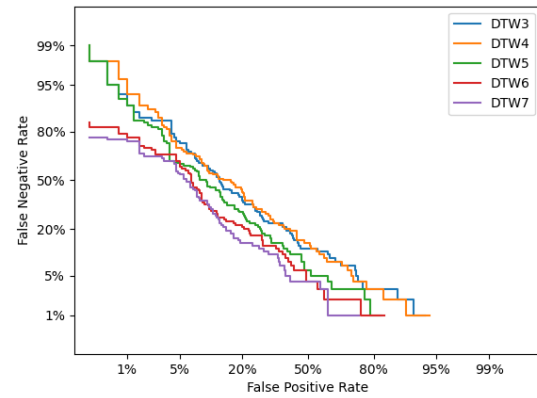
number of symbols	Accuracy
$k = 3$	87
$k = 4$	73
$k = 5$	84
$k = 6$	80
$k = 7$	81

Inference :- Therefore for number of symbols = 15, hidden states = 3, we are getting maximum accuracy.

2) ROC and DET Curves



(a) *ROCcurve*



(b) *DETcurve*

3) Confusion Matrix

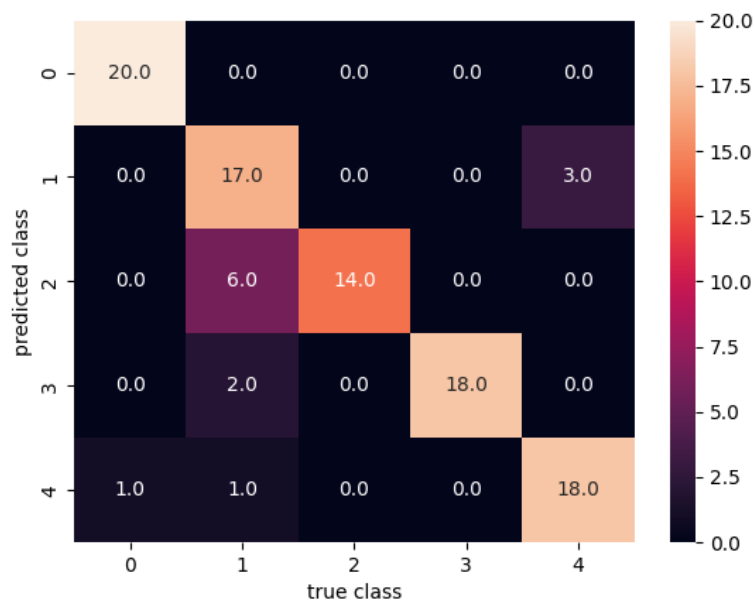


Figure 11: *ConfusionMatrixfor numhiddenstates = 3*