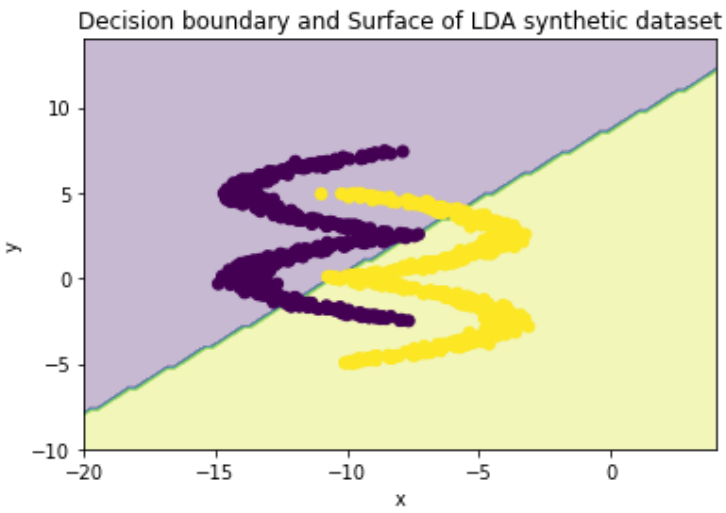
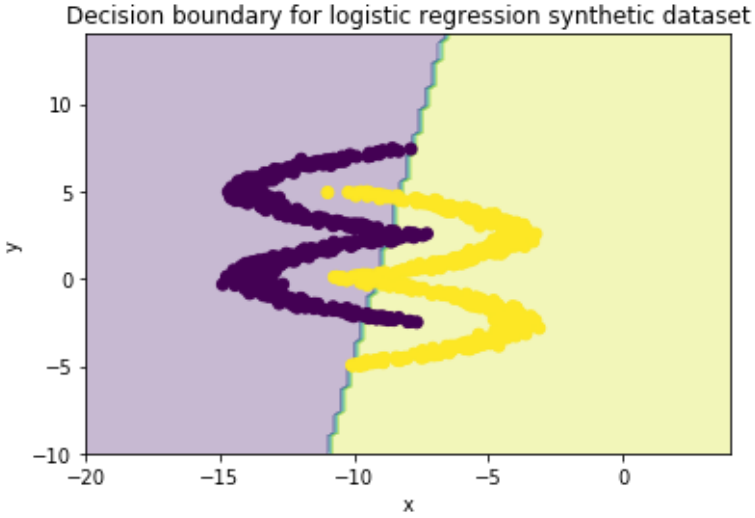
Assignment 4

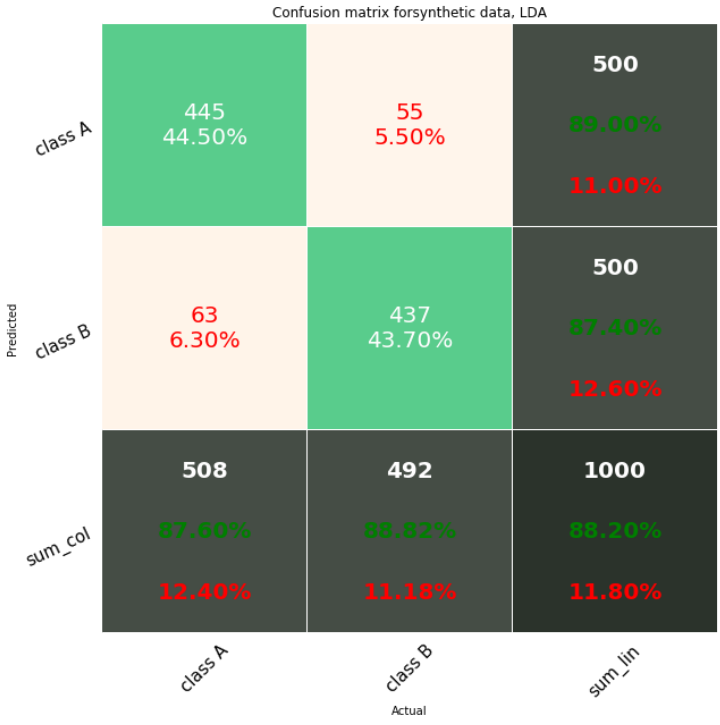
Team 6- Arabhi Subhash(CS17B005), Abdul Mooizz(CS17B034)

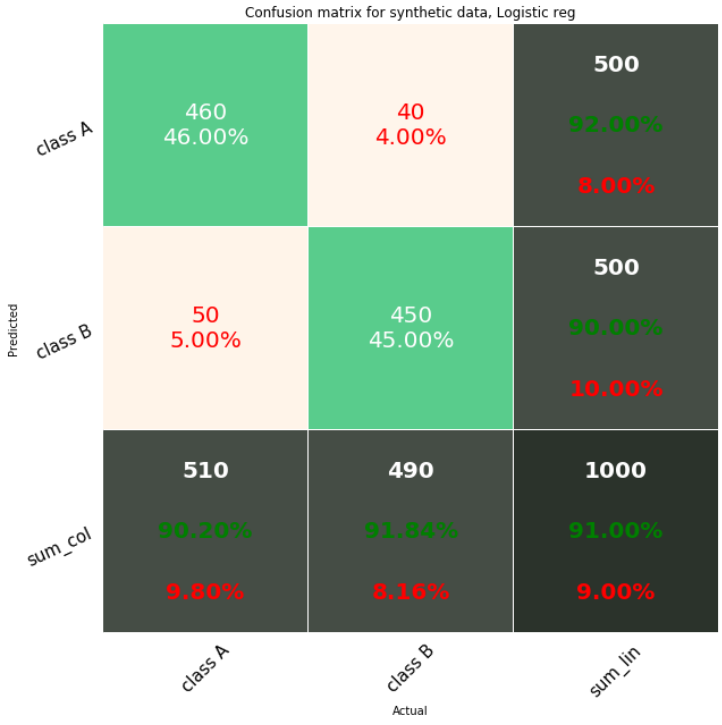
Comparison of different classifiers on different dataset

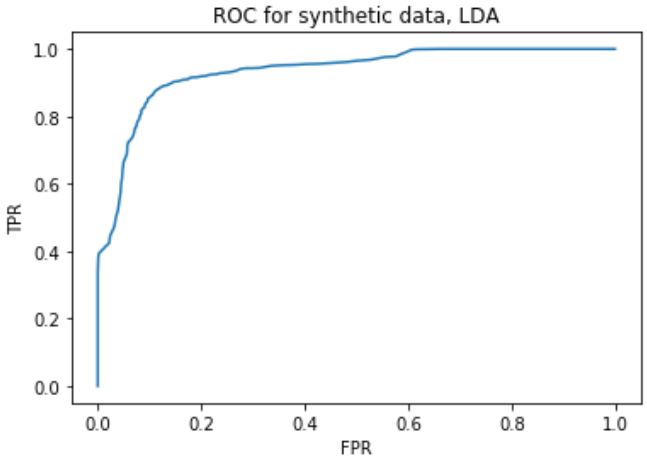
Synthetic Dataset

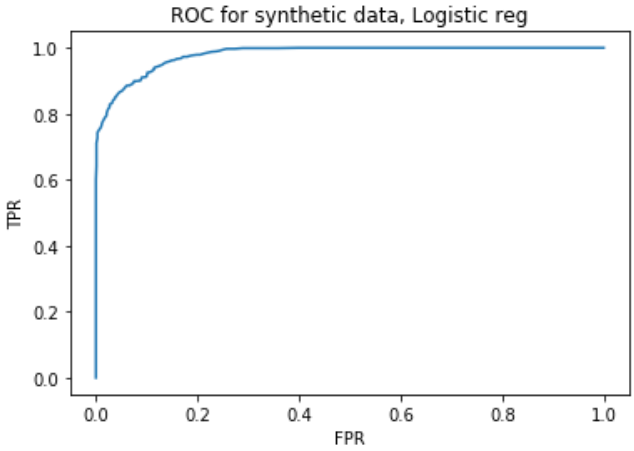


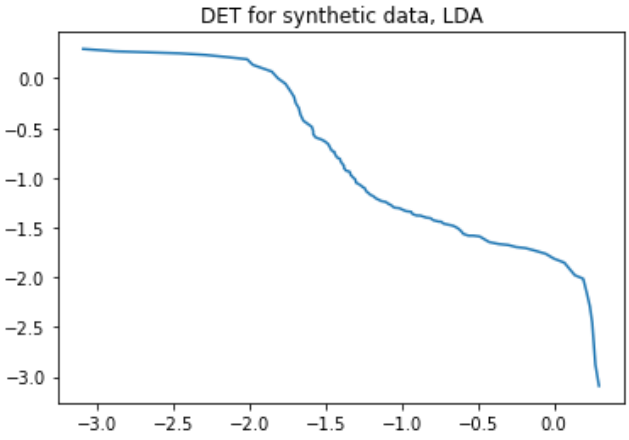


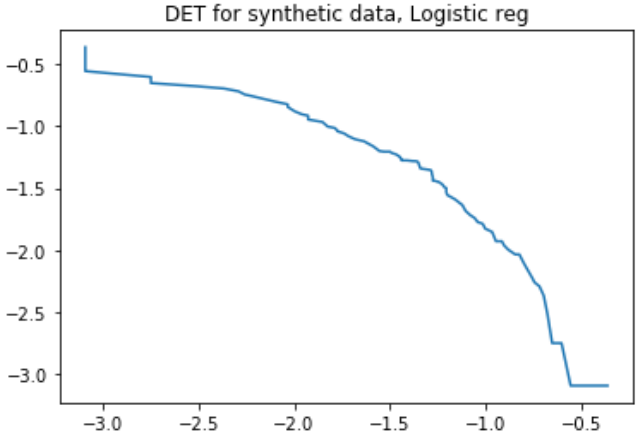


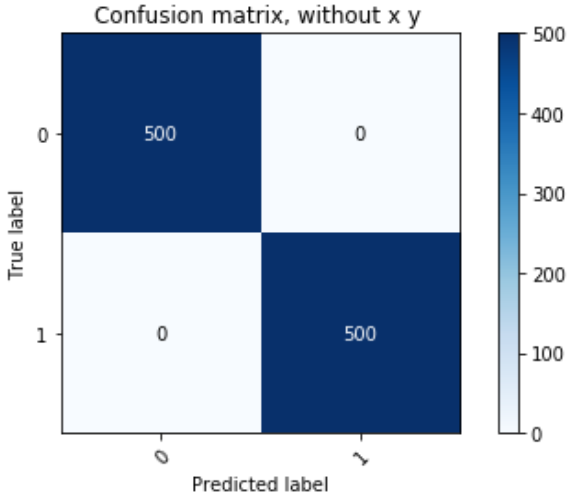


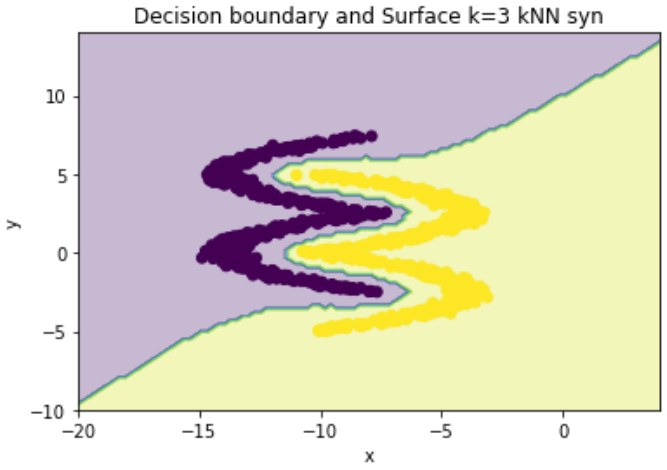


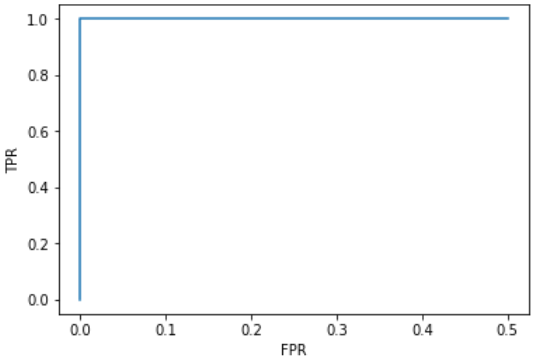






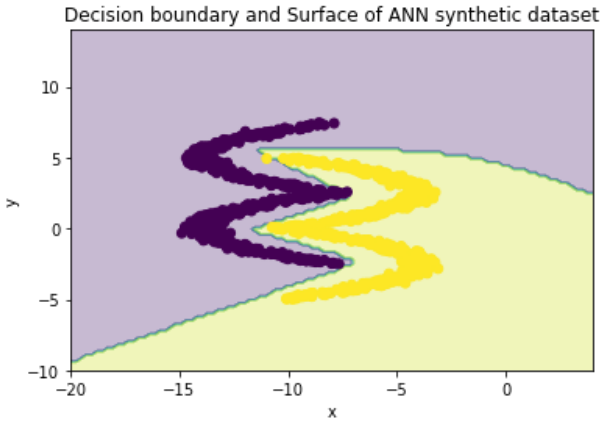


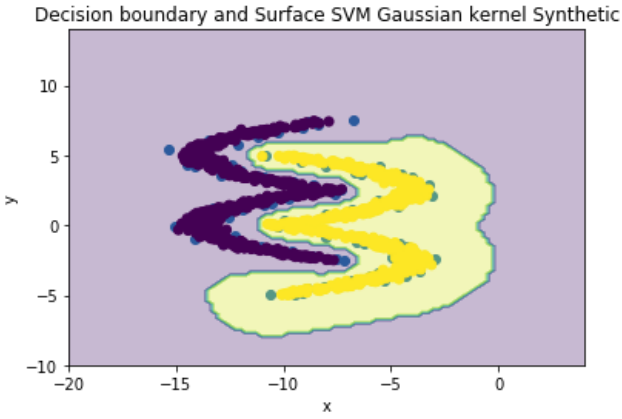


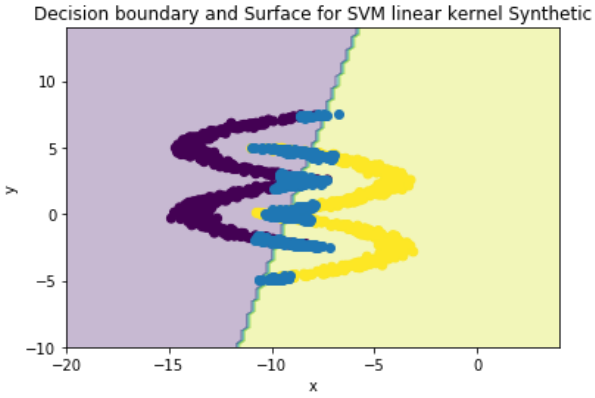


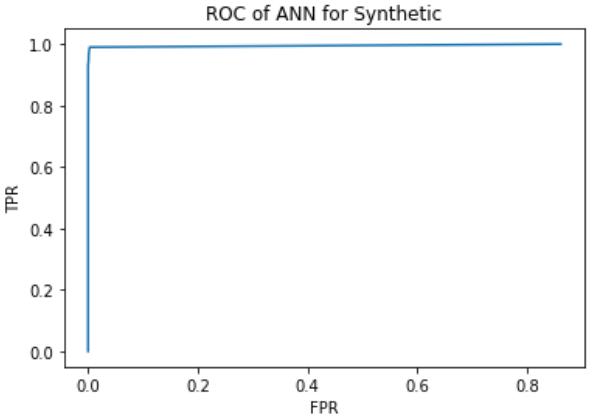
Observations:

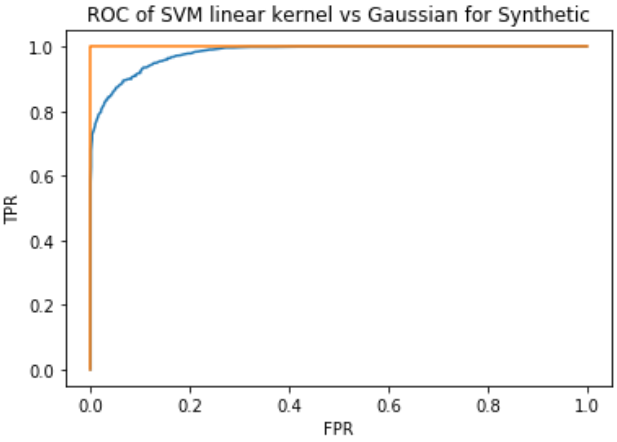
1. In logistic regression best output is obtained after 30,000 iterations with a learning parameter 5e(-7).
2. k=1 was enough to classify the data correctly in KNN

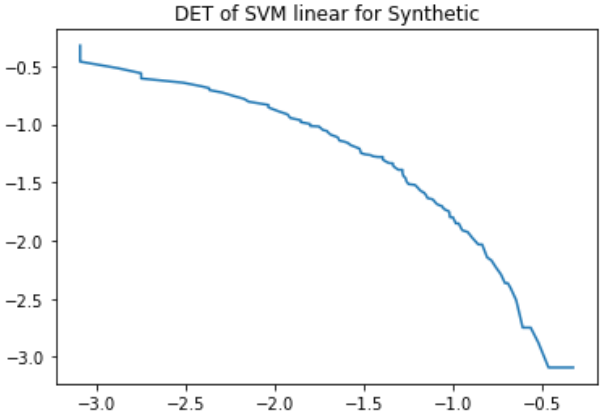


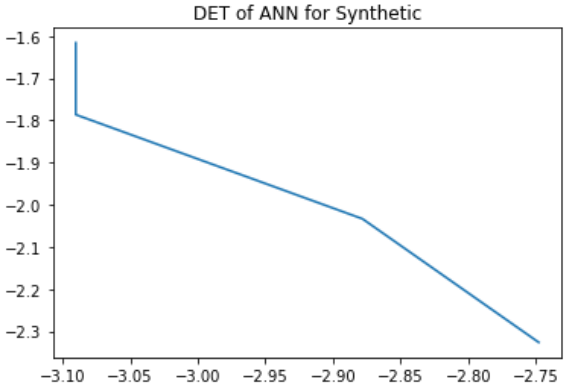










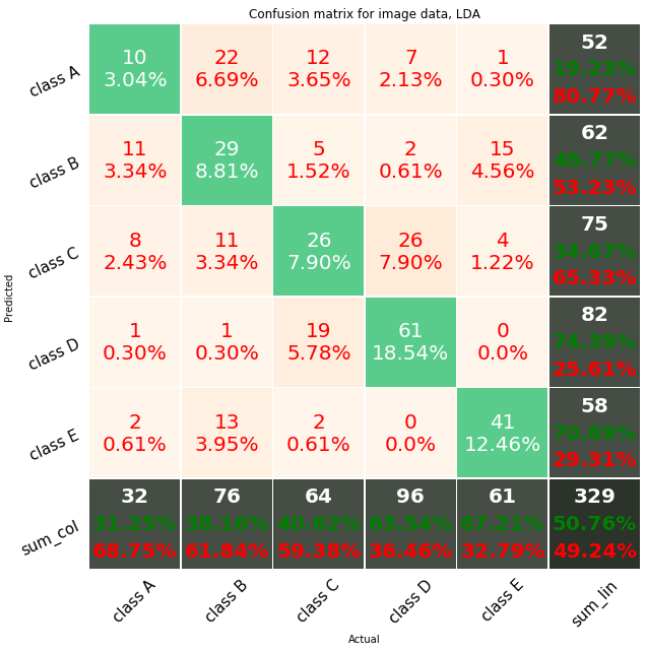


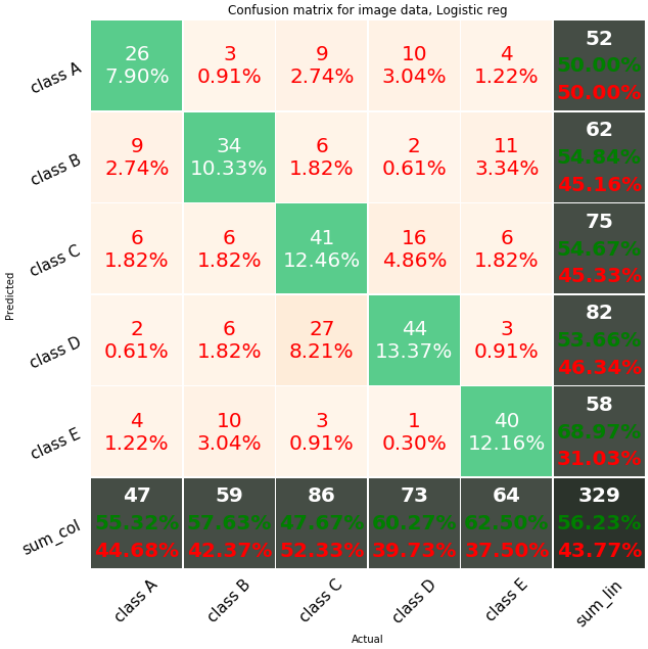
Observations:

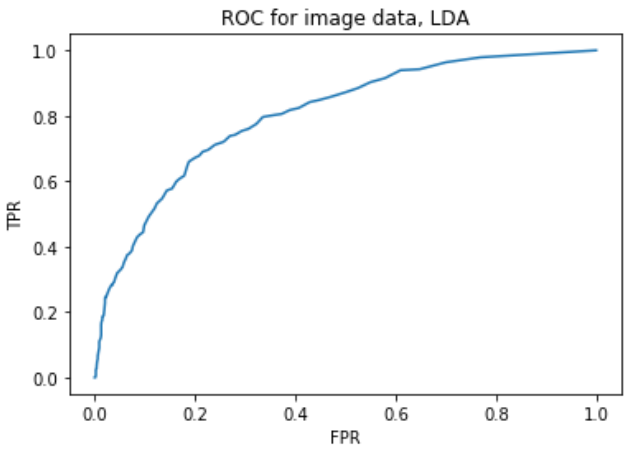
1.Blue dots on the boundary surface represent the support vectors.

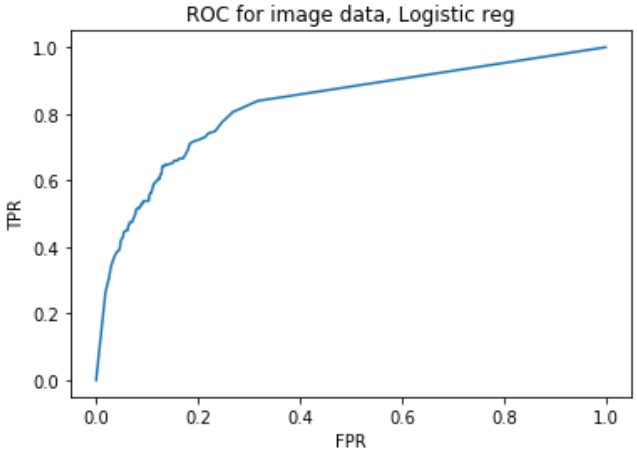
2.ROC curves orange is gaussian kernel and blue is linear

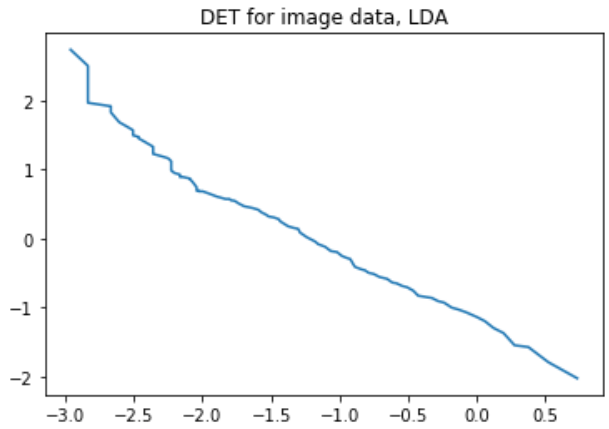
IMAGE DATASET

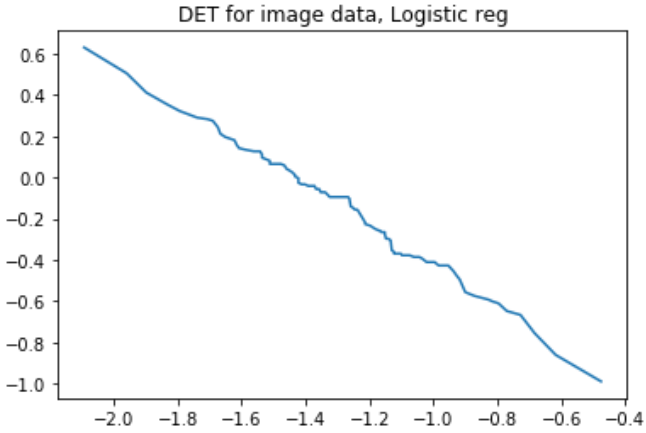


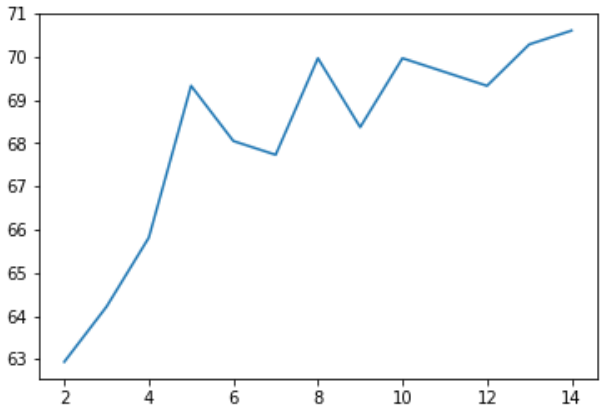


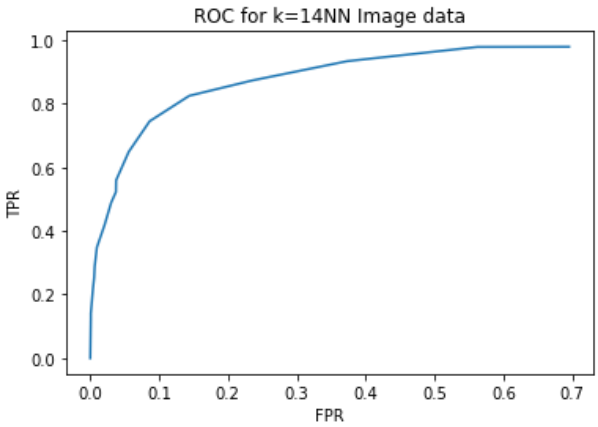


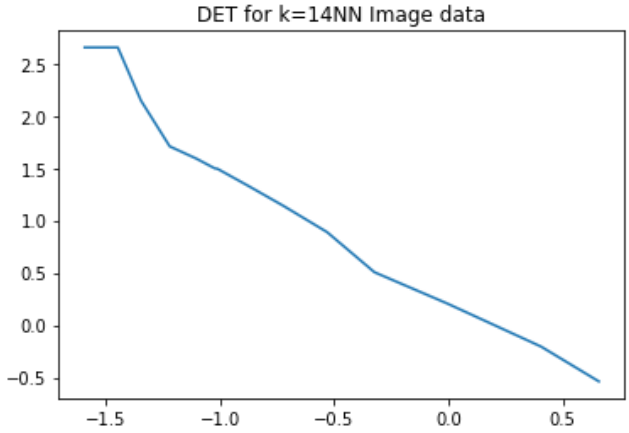


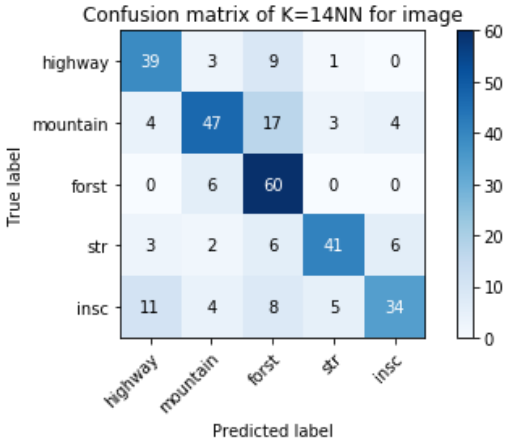


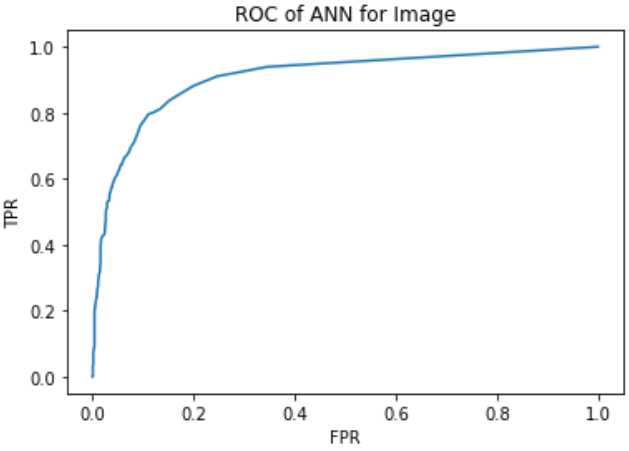


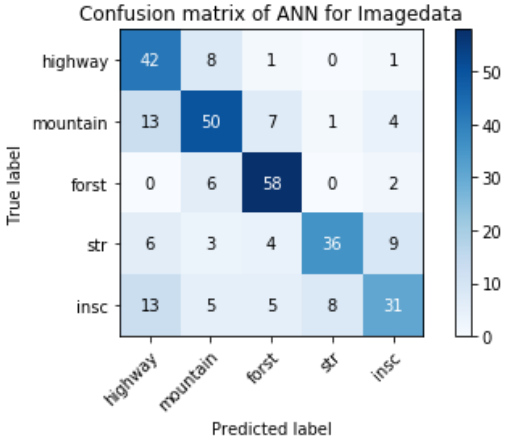


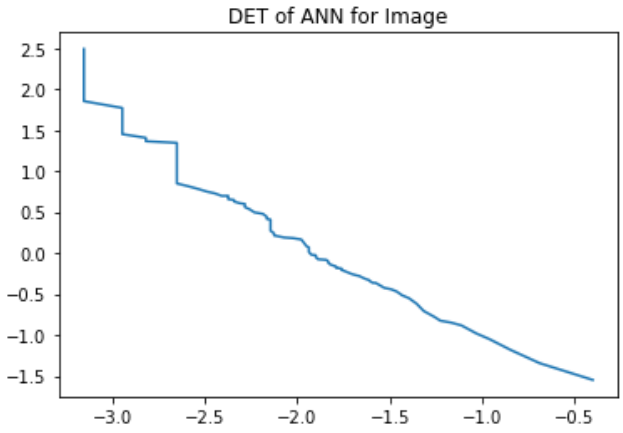






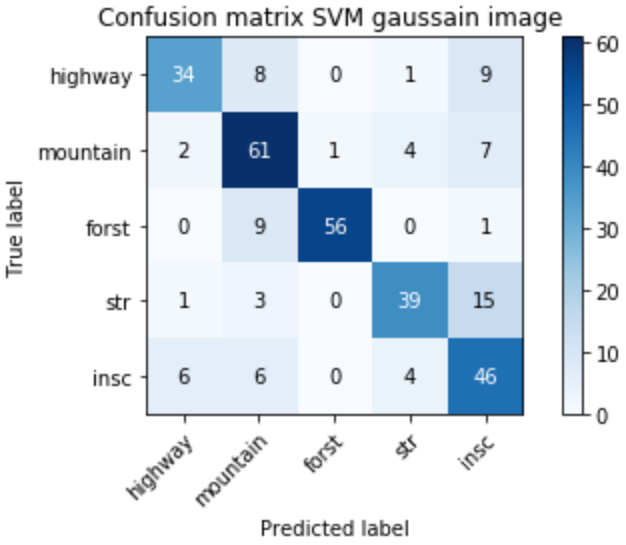


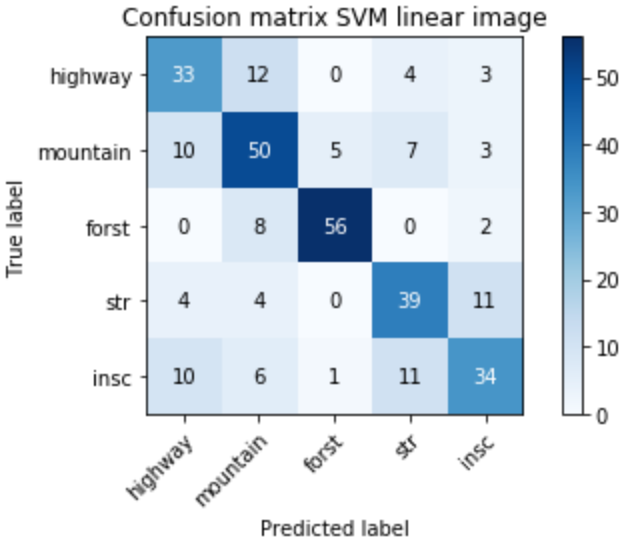


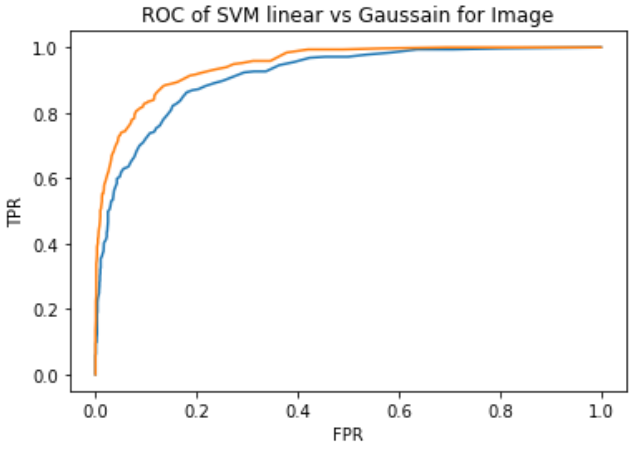


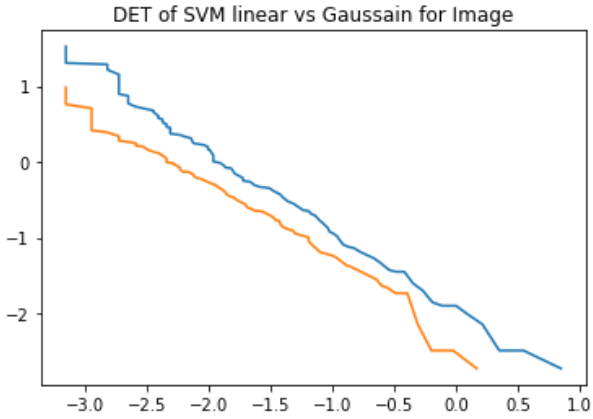
Observations:

1. Here each image is a vector of 428 dimensions PCA to same dimensions yielding poor results so reduced it to 6(best of other) and then did LDA.
2. Though results are poor pertaining to no linear separability of data.
3. k=14 gives maximum accuracy in KNN

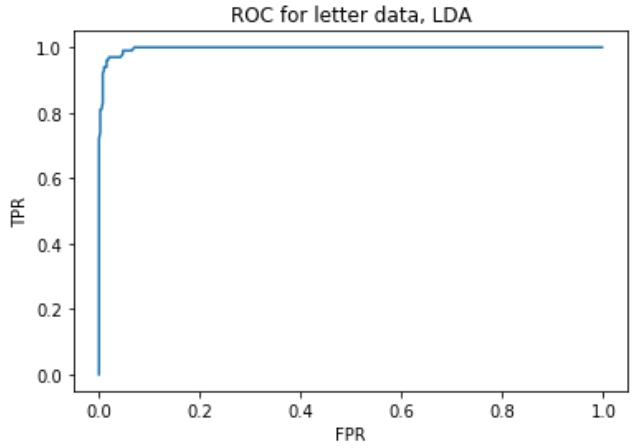


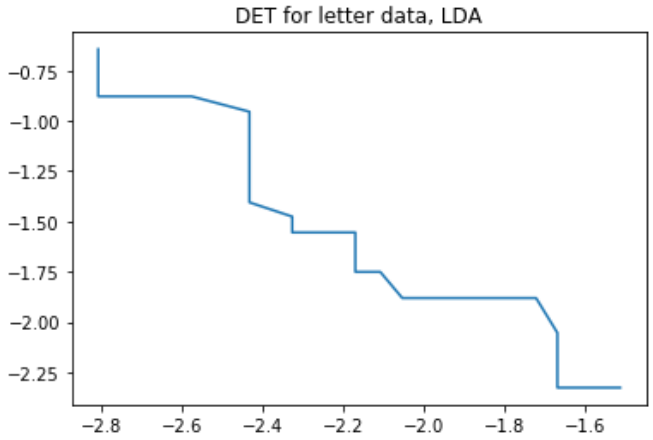


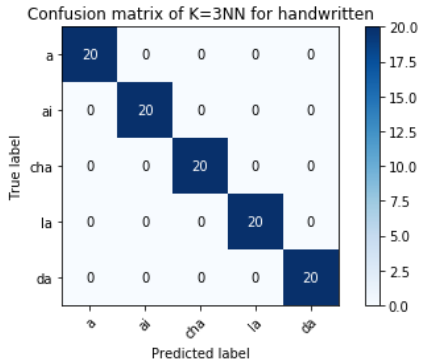


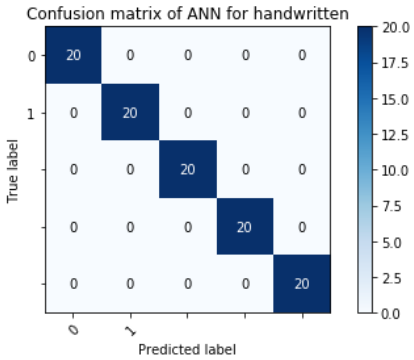


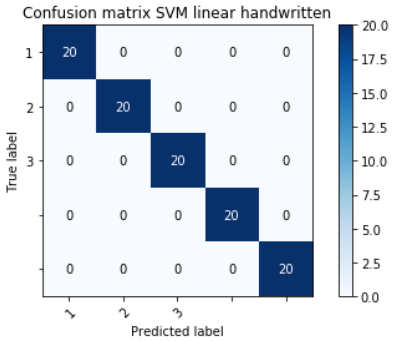
HANDWRITTEN DATA

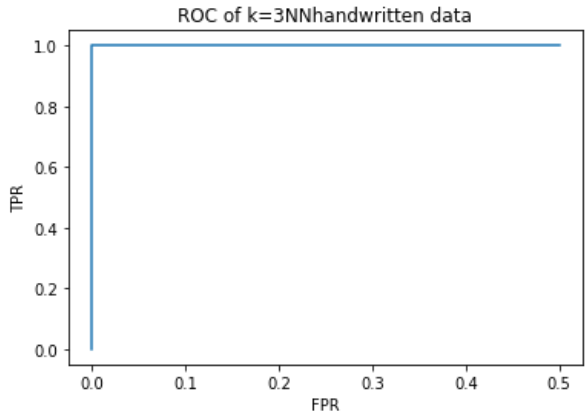

ROC for letter data, LDA
TPR
FPR

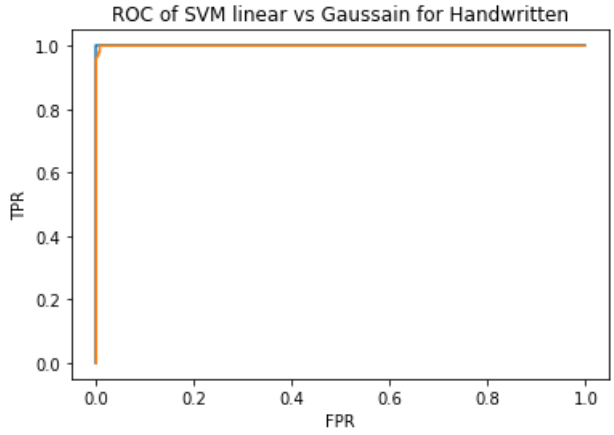

DET for letter data, LDA


Confusion matrix of K=3NN for handwritten
True label
Predicted label
a
ai
cha
la
da

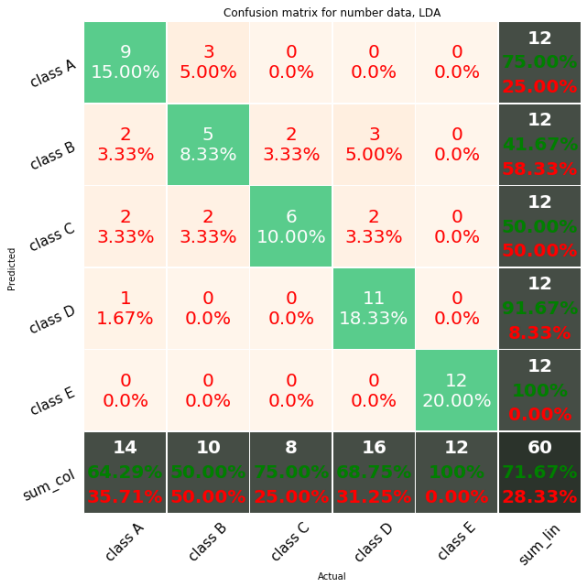

Confusion matrix of ANN for handwritten
True label
Predicted label

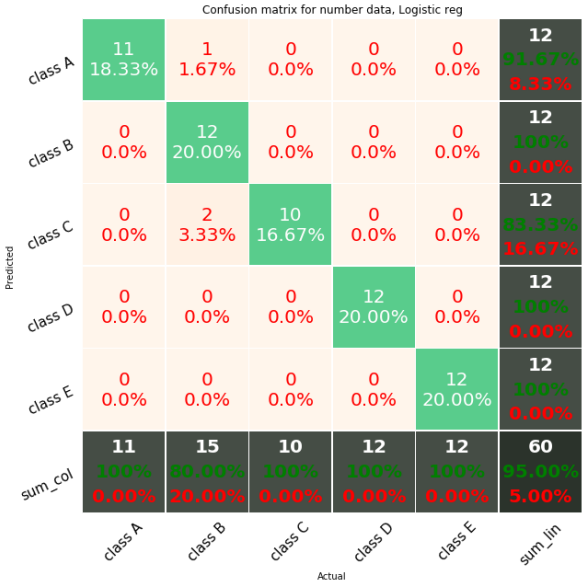

Confusion matrix SVM linear handwritten
True label
Predicted label

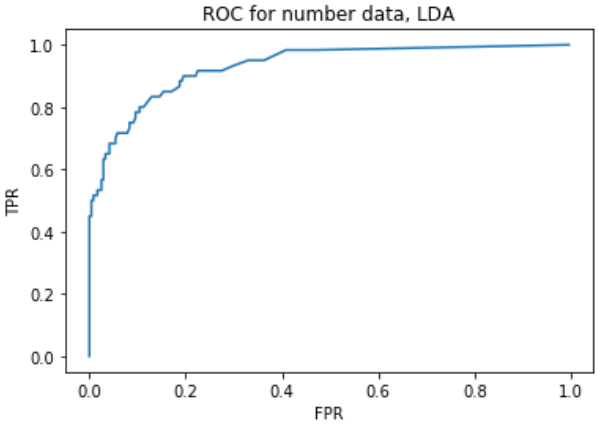

ROC of k=3NNhandwritten data
TPR
FPR

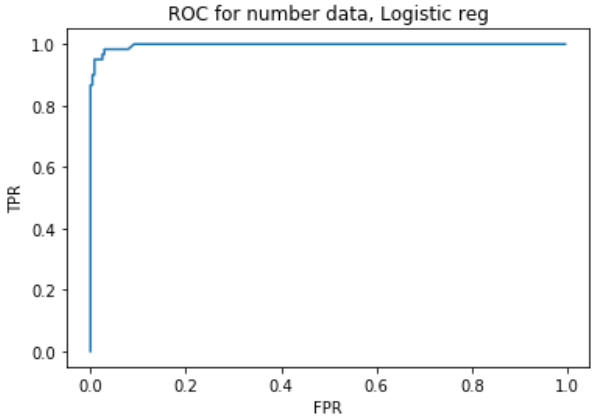

ROC of SVM linear vs Gaussain for Handwritten
TPR
FPR

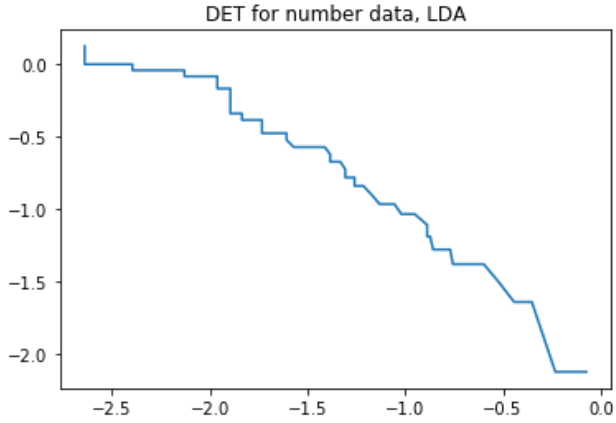
SPEECH DATA

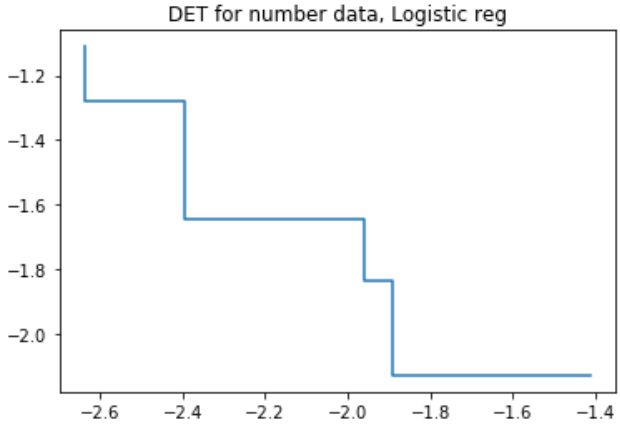

Confusion matrix for number data, LDA
class A
class B
class C
class D
class E
sum_col
sum_lin
Predicted
Actual

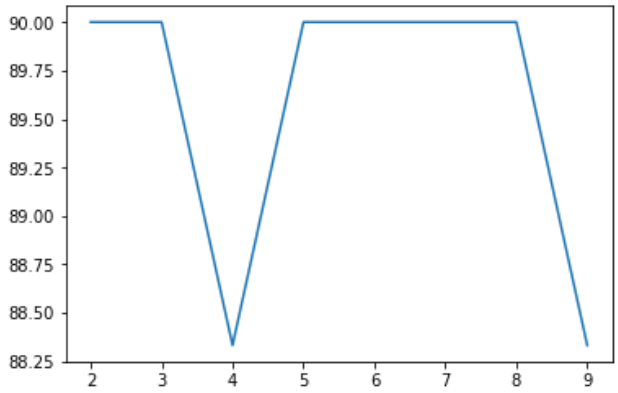

Confusion matrix for number data, Logistic reg
class A
class B
class C
class D
class E
sum_col
sum_lin
Predicted
Actual

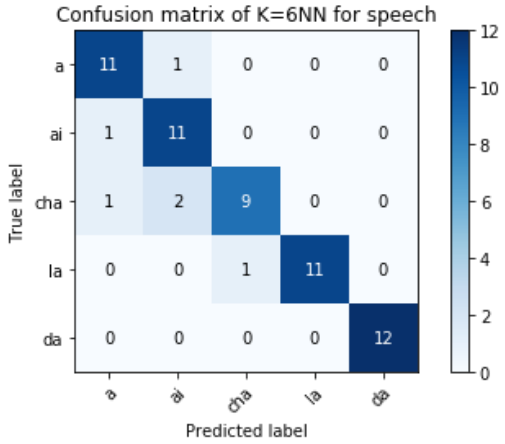

ROC for number data, LDA
TPR
FPR

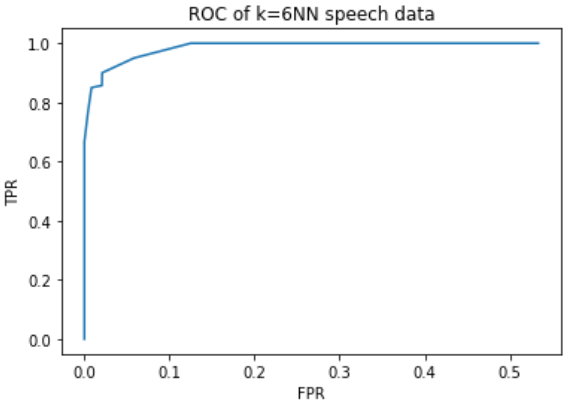

ROC for number data, Logistic reg
TPR
FPR

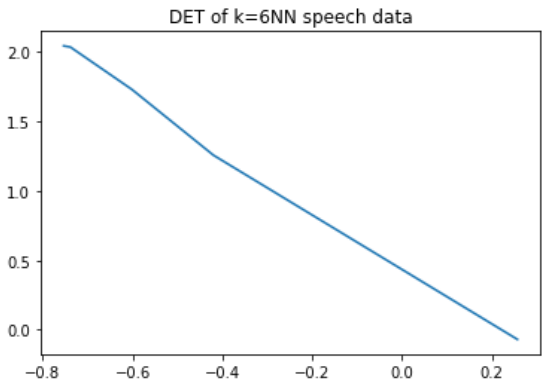

DET for number data, LDA

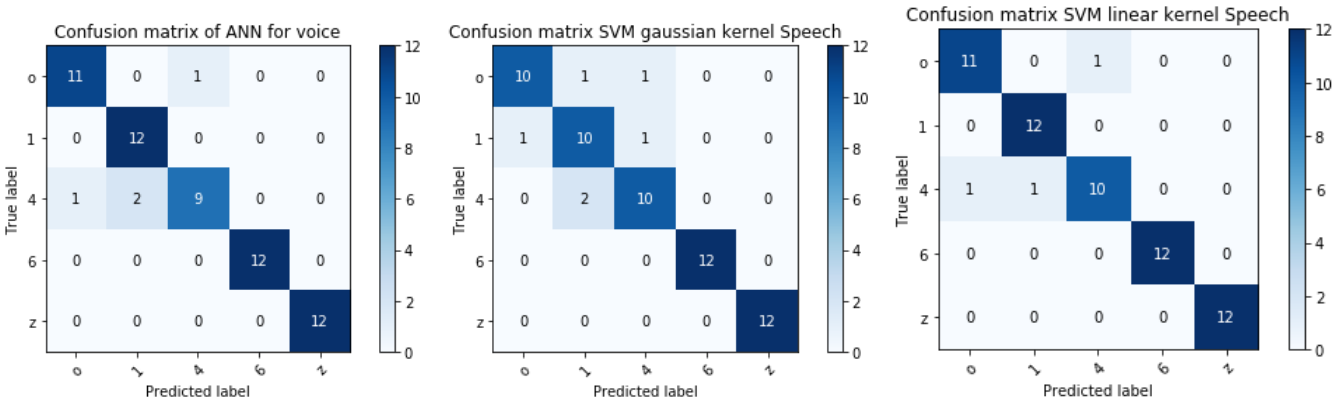

DET for number data, Logistic reg

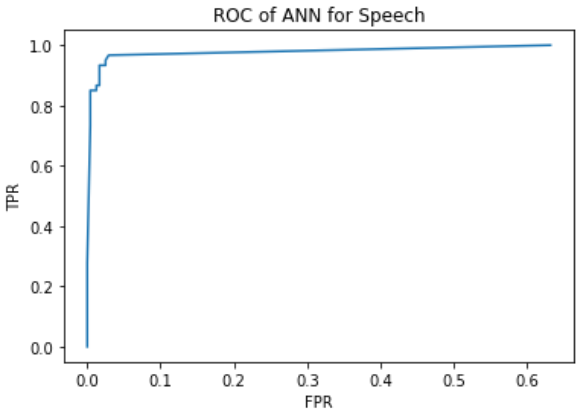


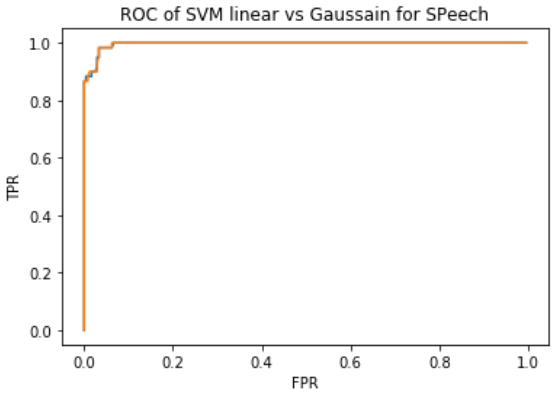

Confusion matrix of K=6NN for speech
True label
Predicted label
a
ai
cha
la
da

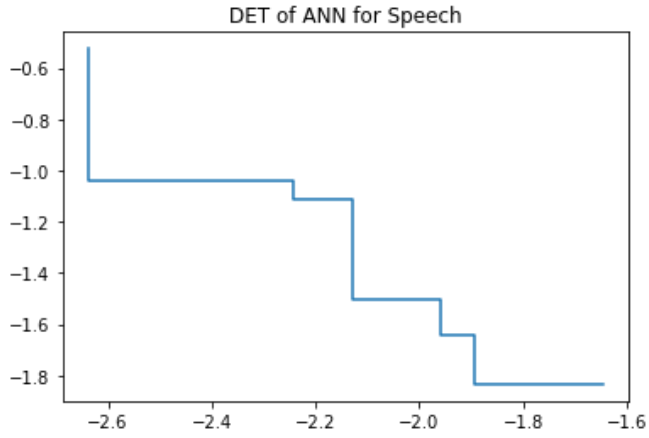

ROC of k=6NN speech data
TPR
FPR

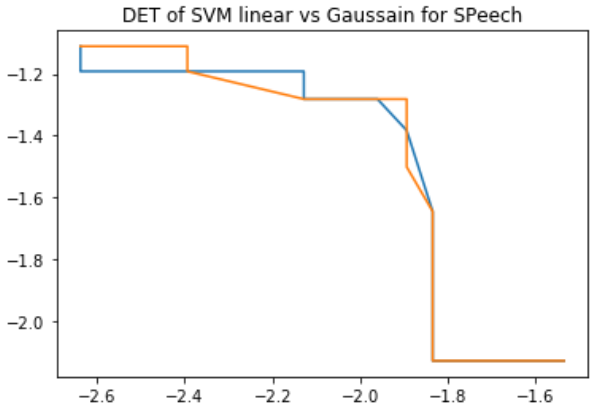

DET of k=6NN speech data











Observations and Implementation details

1.The handwritten data and speech data are sequential.The feature vector length varies so you have to take the average of the length and trim or repeat features to match the average length but while trimming some data is lost, so we decided to add feature vectors uniformly to keep the sequential property to maximum length

2.Now we have feature vectors of the same length

3.Handwritten every classifier gives the best results

4.Speech data LDA performed low when compared to others.