Pattern Recognition and Machine Learning
Assignment 4a & Assignment 4b

RAJ GAURAV TIWARI
EE19S010
Electrical Engineering Department,

January 12, 2020

1   Tasks to be Performed

1. Perform the K-Nearest Neighbour on the data sets given below

2. Perform the Fisher Linear Discriminant analysis on the data sets given below

3. Perform the logistic regression on the data sets given below

4. Perform the Neural Network on the given data sets given below

5. Perform the SVM on the given datasets given below

2   Data Sets Available

2.1   Synthetic data

In this datasets two classes has been given which has been separated in the spiral form

2.2   Image data

There five classes provided coast,highway,insidecity,opencountry and tallbuilding

2.3   Speech data

There have been the five classes of the speech data but the problem with speech data is that it is not of fix length and due to this it cannot be change into fix length vector because of this it cannot be used so I tried to convert into the I-vector form so that it can be used as an input to the neural network,logistic regression and the flda.

2.4   Handwritten data

There are five classes to classify

3   Theory

3.1   K-Nearest Neighbour

In this classifier we tried to find out the nearest neighbour of the test with all the training data and for measuring the nearness of the data we can use simple euclidean distance for the synthetic and the image data and we can use the dtw distance for the speech data and handwritten.After that take the top k values which will behave like the k nearest neighbour to the minimum value after that do the voting rule to find out the class to which class is maximum in the top k values and that will be our resultant class for the given class

3.2   Fisher Linear Discrminant Analysis

It can be called as the way of reducing the dimension of the data as like that of PCA but the difference between the two is that PCA tries to find out the direction of the maximum variance and the FLDA tries to find out the direction of the maximum discrimination but in case of some data set are not very large compared to the dimension of the data.

3.3 Logistic regression

In logistic regression find the optimum weight of each class using the gradient descent algorithm model it for each class

3.4 Neural Network

Use the inbuilt function of the matlab only.

3.5 SVM

Use the inbuilt function of the matlab only.
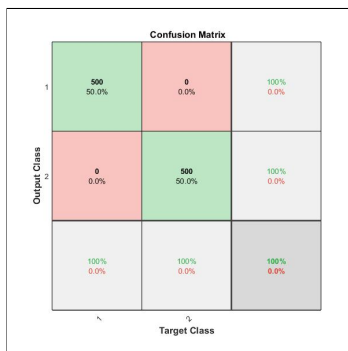
4 Confusion matrices
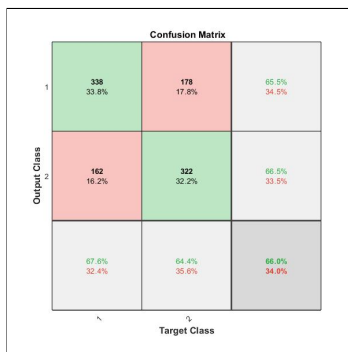
4.1 Synthetic Data

Figure 1: KNN

Figure 2: FLDA

Figure 3: LOGISTIC REGRESSION

Figure 4: NEURAL NETWORK

Figure 5: SVM

4.2 Image data

Figure 6: KNN

Figure 7: FLDA

Figure 8: LOGISTIC REGRESSION

Figure 9: NEURAL NETWORK

Figure 10: SVM

4.3 Speech data

Figure 11: KNN

Figure 12: FLDA

Figure 13: LOGISTIC REGRESSION
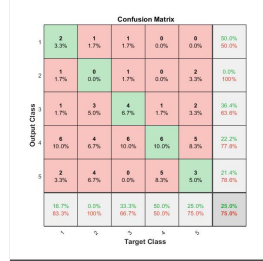

Figure 14: NEURAL NETWORK

Figure 15: SVM

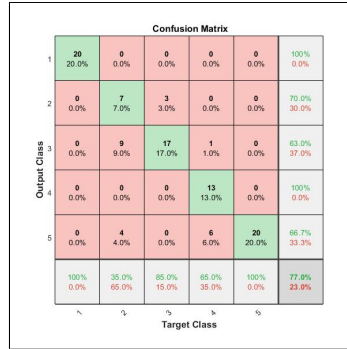## 4.4 Handwritten Data

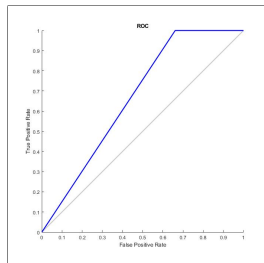Figure 16: SVM

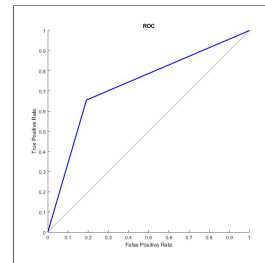# 5 ROC Curves

5.1 synthetic Data

Figure 17: KNN

Figure 18: FLDA
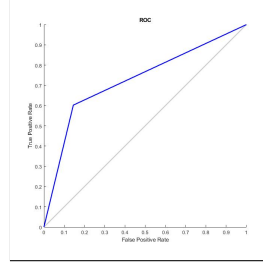
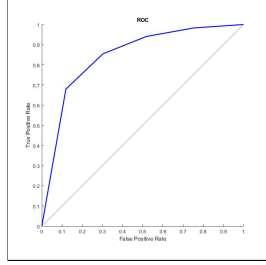# 6 DET Curves

4

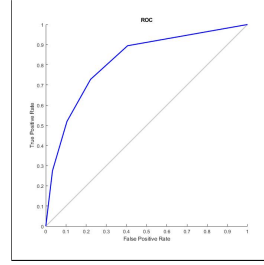Figure 19: LOGISTIC REGRES-SION

Figure 20: KNN IMAGE
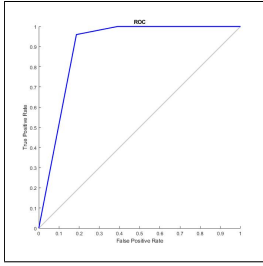

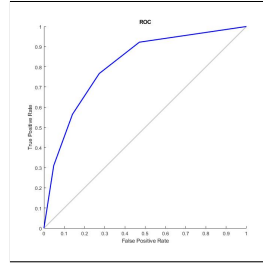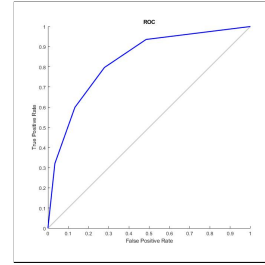Figure 21: FLDA IMAGE

Figure 22: KN SPEECH

Figure 23: FLDA SPEECH
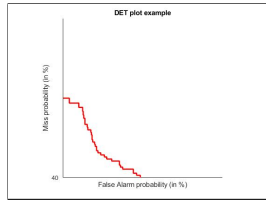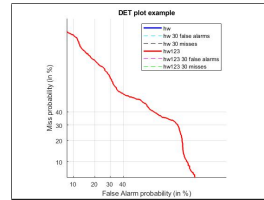

Figure 24: LOGISTIC REGRES-SION

Figure 25: KNN SPEECH

Figure 26: FLDA Synthetic

Table 1: Accuracy of models on different datasets

Datasets	KNN	FLDA	LOGISTIC REGRESSION	NEURAL NET	SVM
SYNTHETIC	100%	52.6%	66%	100%	65.9%
IMAGE	48.5%	38.2%	50.6%	73.2%	59.1%
SPEECH	91.7%	25%	20%	26.7%	25%
HANDWRITTEN	77%	-	-	-	-