# CS5691: Pattern Recognition and Machine Learning

## Programming Assignment 3

Team 31 - CH18B002-Chaitanya Toraskar and CS19B042-Shone Pansambal

May 2, 2022

# 1 KNN

## 1.1 Introduction

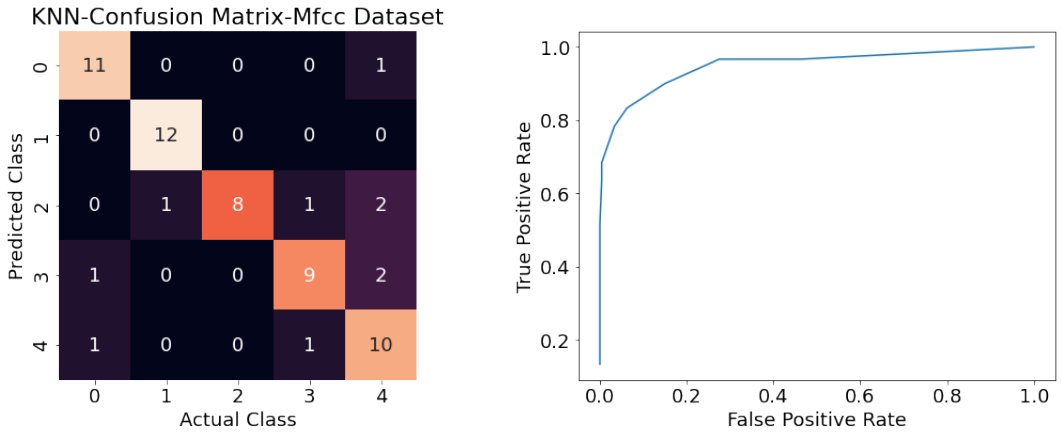
K-nearest neighbors (kNN) is a supervised machine learning algorithm that can be used to solve both classification and regression tasks.kNN classifier determines the class of a data point by majority voting principle.

## 1.2 Algorithm

- KNN does not require any training.We pass the training dataset along with each observation from development dataset.
- We then compute the euclidean distance of all points with the test observation.As we have time series data for isolated digits and handwritten dataset we append zeros to make the vectors of equal length while computing the euclidean distances.
- Then we sort them in ascending order and use the top k distances.
- We find the majority class among them which is the predicted class and also use this for calculating probailtities for ROC curve.

## 1.3 Observations

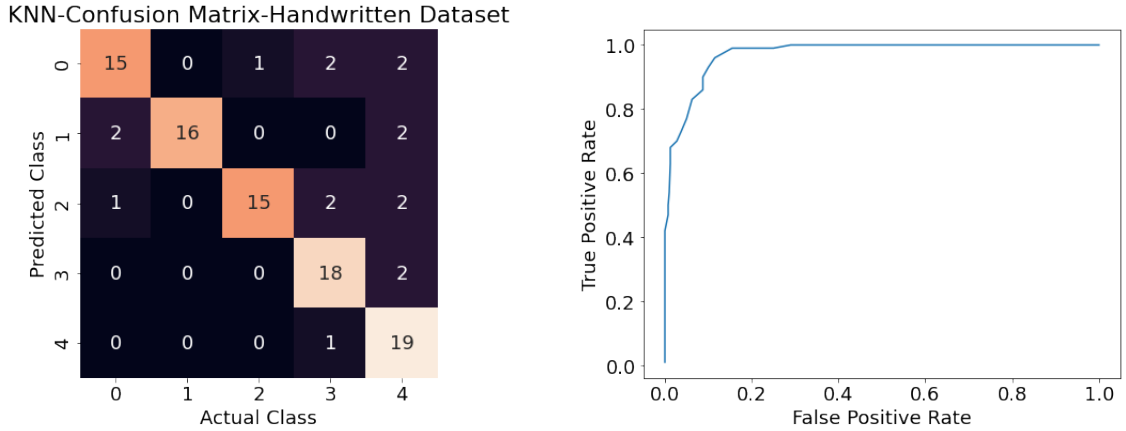
KNN gives very good accuracy for synthetic dataset and a works decently for other 3 datasets where image dataset has the least accuracy.For image dataset opencountry is wrongly classified as highway and coasts many times.This probably can be due to presence of open skies in all 3 types which cannot be distinguished by KNN.For remaining 3 datasets KNN shows very good accuracy.We have also checked the accuracy for various values of k and used the one that gave very good accuracy.Even though k=1 always showed slightly better accuracies we have used higher values of k.Following are the plots for confusion matrices and roc graphs for given 4 datasets.




(a) Confusion matrix for classes-1,2,3,5,o

(b) ROC curve

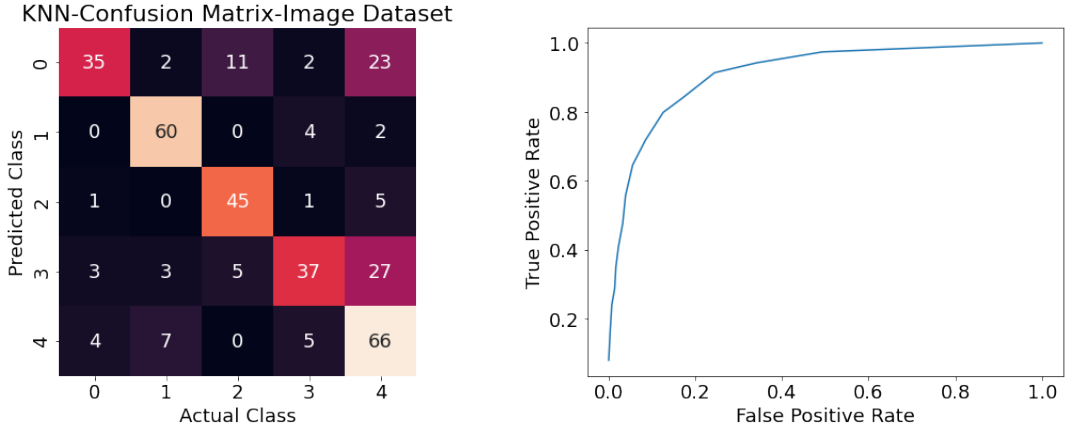
Figure 1: Isolated Digits Dataset,accuracy of 0.83 for k=11




(a) Confusion matrix for classes-a,ai,chA,dA,lA

(b) ROC curve

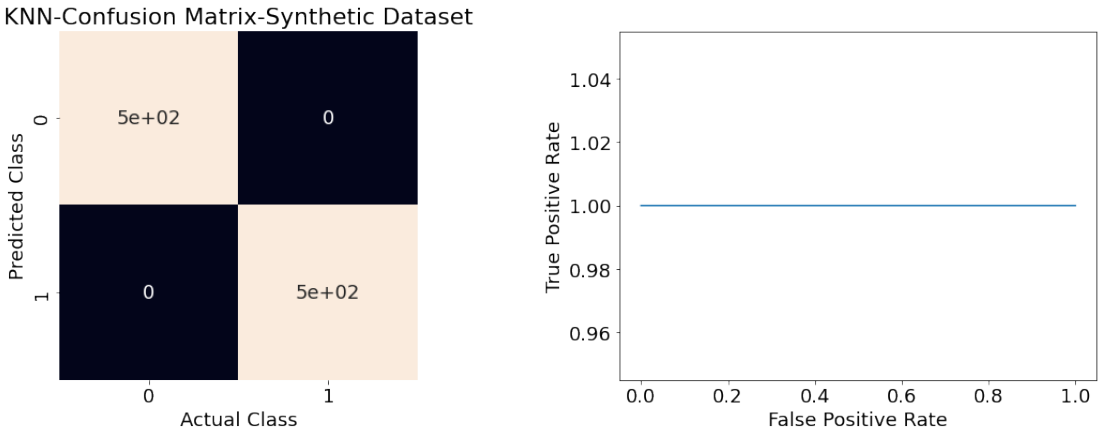
Figure 2: Handwritten Dataset,accuracy of 0.84 for k=9




(a) Confusion matrix for classes-coast,forest,mountain,highway,opencountry

(b) ROC curve

Figure 3: Image Dataset,accuracy of 0.70 for k=15




(a) Confusion matrix for classes-1,2

(b) ROC curve

Figure 4: Synthetic Dataset.Accuracy of 1.0 for k=15

# 2 Logistic Regression

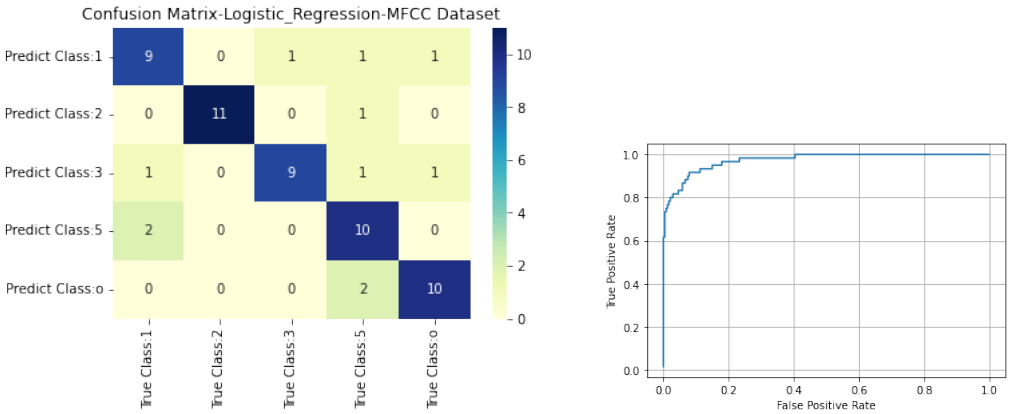
## 2.1 Introduction

Logistic function is similar to logistic regression however it uses a sigmoid function.For multiclass logistic function we have use softmax function from sklearn.

### 2.1.a Algorithm

- We initially take projection of features on weight vector.
- Then apply softmax function on each row to get predicted probabilties.
- We have iterated 1000 times to get a good accuracy
- Finally use argmax to predict the class.

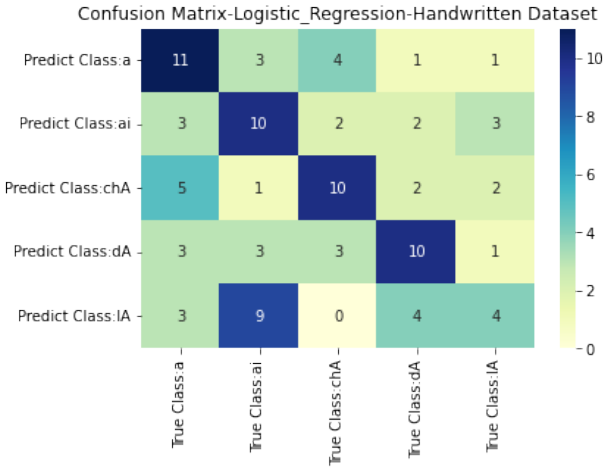
## 2.2 Observations



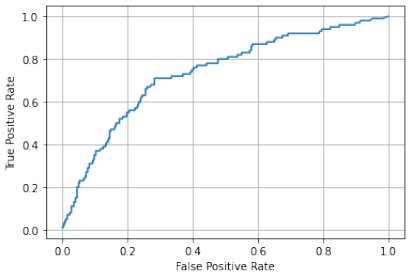

(a) Confusion matrix for classes-1,2,3,5,o

(b) ROC curve

Figure 5: Isolated Digits Dataset,accuracy of 0.81

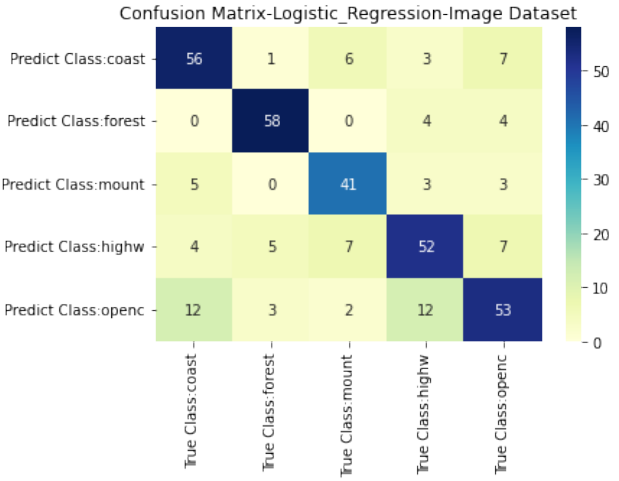



(a) Confusion matrix for classes-a,ai,chA,dA,lA

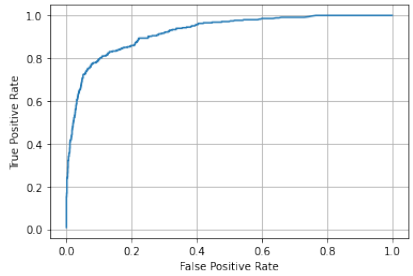



(b) ROC curve

Figure 6: Handwritten Dataset,accuracy of 0.45

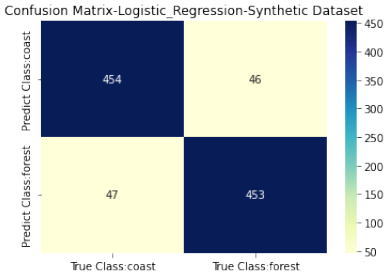



(a) Confusion matrix for classes-coast,forest,mountain,highway,opencountry

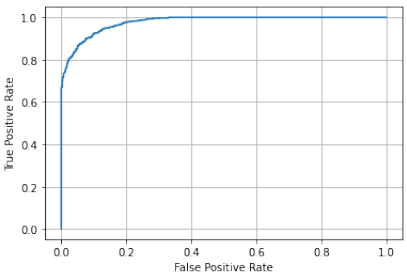



(b) ROC curve

Figure 7: Image Dataset,accuracy of 0.74




(a) Confusion matrix for classes-1,2




(b) ROC curve

Figure 8: Synthetic Dataset.Accuracy of 0.90

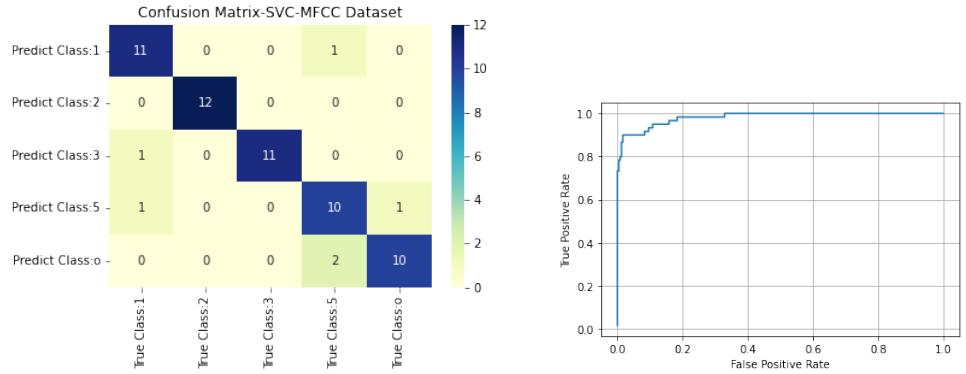
We get good accuracies for all datasets except for the handwritten one where we get an accuracy of only 0.45.In this case the model doesn't work well while predicting the character lA.

# 3 Support Vector Machines

## 3.1 Introduction

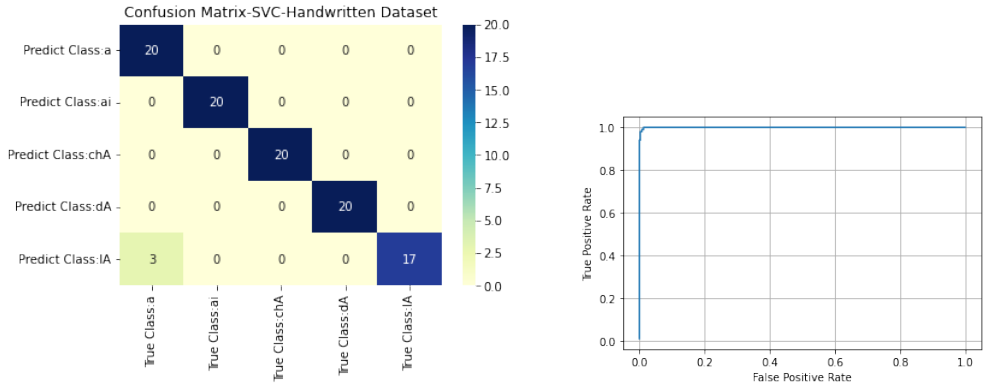
Support vector machine is simply a further extension of the support vector classifier to accommodate non-linear class boundaries. In SVMs we try to find the hyperplane which has maximum distance(margins) from points that are closest to it.Thus to define the plane we require only these closest points which are known as support vectors.We have used the SVC library from the sklearn.

## 3.2 Observations



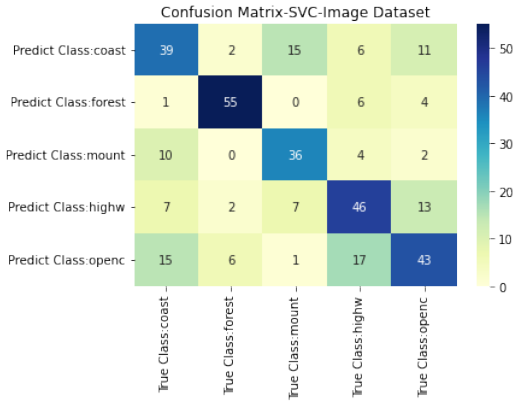

(a) Confusion matrix for classes-1,2,3,5,o (b) ROC curve

Figure 9: Isolated Digits Dataset,accuracy of 0.90

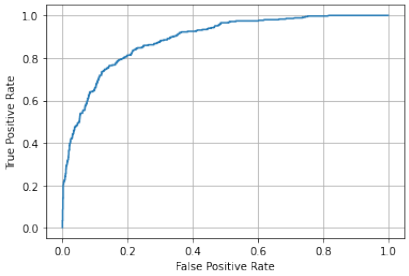



(a) Confusion matrix for classes-a,ai,chA,dA,lA (b) ROC curve

Figure 10: Handwritten Dataset,accuracy of 0.97

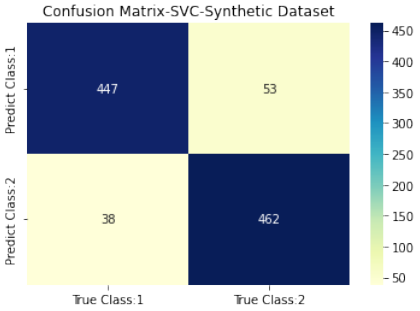



(a) Confusion matrix for classes-coast,forest,mountain,highway,opencountry

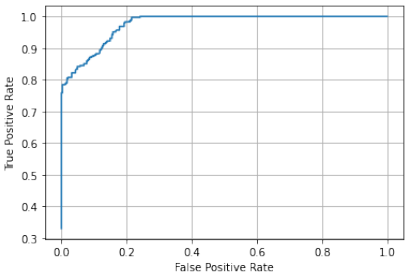



(b) ROC curve

Figure 11: Image Dataset,accuracy of 0.63




(a) Confusion matrix for classes-1,2




(b) ROC curve

Figure 12: Synthetic Dataset.Accuracy of 0.90

We get very good accuracies for all datasets.The least accuracy we get is for image dataset which is 0.63.Here as well opencountry is misclassified as highway and coast many times and similarly highway also misclassifies as open country many times.We have used rbf kernels for mfcc and handwritten datasets and polynomial kernel for synthetic and image dataset all with C=1.

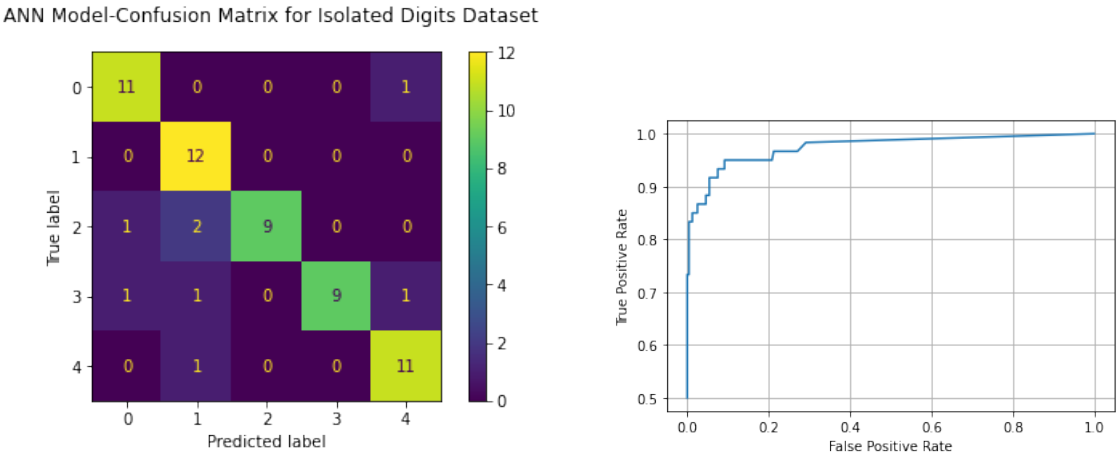
# 4 Artificial Neural Networks

## 4.1 Introduction

We have used a multi layer perceptron for classification.A multilayer perceptron uses feedforward algorithms.It propogates the output during forward pass through the layers and after calculating the errors it sents them back to re-estimate the weights for a better output.

## 4.2 Observations

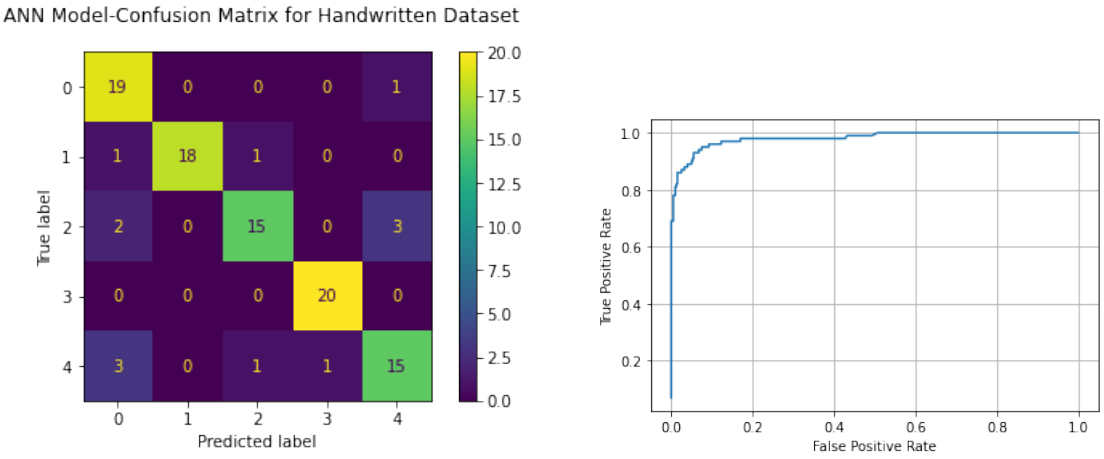
We have used multi layer perceptron control model(MLP Classifier) with 4 hidden layers-256,128,64,32 and activation function 'relu'.We have added zeros to make all the vectors of equal length.We get very good accuracy for all the models with the least for image dataset of 0.74.In this case coast and highways are wrongly classified as open country in many cases.For other images there is good prediction.




(a) Confusion matrix for classes-1,2,3,5,o

(b) ROC curve

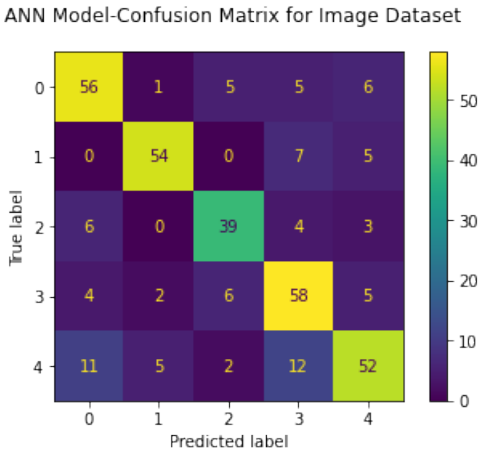
Figure 13: Isolated Digits Dataset,accuracy of 0.866



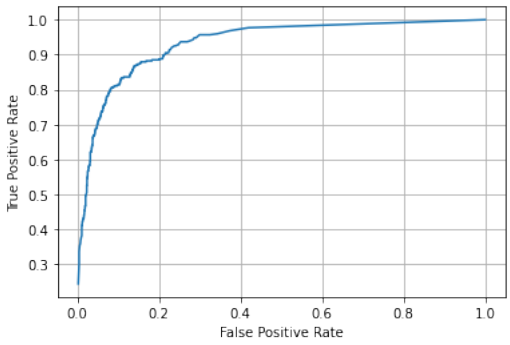

(a) Confusion matrix for classes-a,ai,chA,dA,lA

(b) ROC curve

Figure 14: Handwritten Dataset,accuracy of 0.87

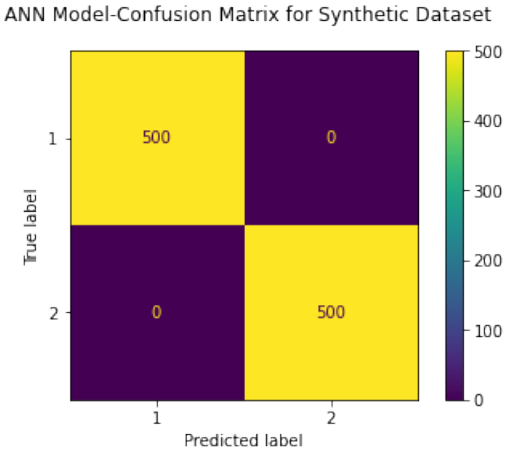



(a) Confusion matrix for classes-coast,forest,mountain,highway,opencountry

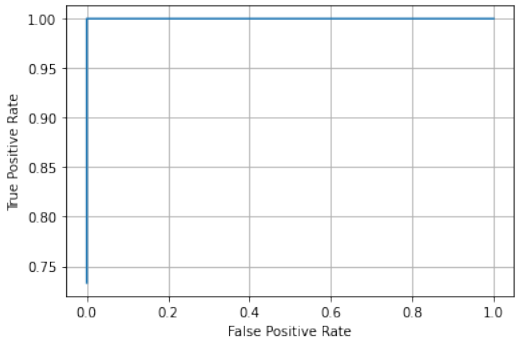



(b) ROC curve

Figure 15: Image Dataset,accuracy of 0.74




(a) Confusion matrix for classes-1,2




(b) ROC curve

Figure 16: Synthetic Dataset.Accuracy of 1.00