EE527: Machine Learning Laboratory

**Assignment 10**

**Due Date: 10 April 2023**

**[Q1]** Application of the *Perceptron* in Classification of Normal and Shouted Speech using MFCC features. These features are extracted from speech samples of a number of speakers uttering a few sentences normally or by shouting. The features are divided into train-test splits and are made available in two csv files (use dataset of previous assignment). You are tasked to learn a discriminative model to classify normal and shouted speech. This example uses Perceptron as a discriminative model. Consider the .csv file *“Train\_file.csv”* containing *86060* instances of 61- dimensional arrays. The first *60* dimensions of the array contain the feature values for a particular instance and the last dimension contains its label. The label can be either *’0’* or ‘1’.

**(a)** The perceptron input and predicted output are related as follows.

|  |
| --- |
|  |
|  |

Learn the weight vector and bias from the train dataset (*Train\_file.csv*). Do not use any Scikit-Learn functions. Write your own functions for perceptron learning.

**(b)** Read “*Test\_file.csv”* consisting of *21516* instances of 61 dimensional arrays. For each array, the first 60 dimensions contain the feature values for the test data and the last dimension contains its *actual label*. Predict the label of each data instance from the testing set using the learned perceptron and compare the *predicted* and *actual* labels. Report the *class-wise F1-scores* for both classes and the *overall accuracy*.

**[Q2]** Download the [MiniBooNE](https://archive.ics.uci.edu/ml/datasets/MiniBooNE%2Bparticle%2Bidentification) dataset from UIUC Machine Learning Repository. This dataset is imbalanced. First, sample 𝑛 number of samples (say *11000*) from both the classes and set that aside as the test set . Consider the remaining imbalanced data as the training set .

**(a)** Use the perceptron module in Scikit-learn python toolbox to learn a perceptron from . Report the *class-wise F1-scores* of this perceptron on .

**(b)** Balance the dataset using [*KMeansSmote*](https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.KMeansSMOTE.html)algorithm from the imbalanced-learn python tool box to generate the balanced training dataset . Use the perceptron module in Scikit-learn python toolbox to learn a perceptron from . Report the *class-wise F1-scores* of this perceptron on .

**(c)** Report your observations on the two test performances.

Reference: <https://www.jmlr.org/papers/volume18/16-365/16-365.pdf>

**[Q3]** Consider the MNIST Handwritten Digit Recognition dataset used in earlier assignments. Use the perceptron module in Scikit-learn python toolbox to learn a perceptron to perform *10* category classification over the given dataset. Use *SoftMax* as the activation function for the *10* output nodes. Report the *class-wise F1-scores* and the *overall accuracy*. Fold back the weight vectors of the *10* perceptrons as images and visualize the same.