Computer Science & Information Systems

Machine Learning Lab sheet - Module 2

EXERCISE 2 – ML CLASSIFIER

1 Objective

The objective is to

- Build an ML classification model for the MNIST dataset

2 Steps to be performed

Tool	-	Python3, Google Colaboratory
Libraries required	-	numpy, matplotlib, sklearn
Input	-	MNIST dataset, which is a set of 70,000 small images of digits handwritten by high school students and employees of the US Census Bureau.
ML Models	-	SGDClassifier, RandomForestClassifier, SVClassifier, OneVsRestClassifier, KNeighborsClassifier
Implementation	-	02_classification.ipynb

Steps -

- Get the MNIST data.
    - Download the data.
    - Explore the shape of the data and plot samples.
    - Create a Train and Test set.
- Binary Classifier.
    - Fit a classifier model on the training set.
    - Evaluate the cross-validation score on 3 K-folds.
    - Plot the confusion matrix, precision, recall and f1 scores.
- ROC Curves.
    - Plot the ROC Curve for SGD Classifier.
    - Compare the ROC Curves with Random Forest Classifier.

- Multiclass Classification.
    - Train and Evaluate SVClassifier and OneVsRestClassifier.
- Multilabel Classification.
    - Train and Evaluate KNeighborsClassifier.
- Multioutput Classification.
    - Train and Evaluate KNeighborsClassifier with noise added to the data.

3 Expected Results

- Report the training and validation accuracy.
- Report the confusion matrix, precision, recall and f1 scores.

4 Observation

- The ML models were defined, configured, trained and evaluated.
- The results were plotted and displayed.

5 Modifications

- Try to build a classifier for the MNIST dataset that achieves over 97% accuracy on the test set.
- Build a classifier model for the Titanic Dataset.
- Build a spam classifier.