

Sri Sivasubramaniya Nadar College of Engineering, Kalavakkam – 603 110
(An Autonomous Institution, Affiliated to Anna University, Chennai)

UCS2612 Machine Learning Laboratory

Academic Year: 2023-2024 Even

Faculty In-charges: Y.V. Lokeswari & Nilu R Salim

Batch: 2021-2025

VI Semester A & B

A. No. : 3 **Handwritten Character Recognition using Neural Networks**

Download the handwritten character recognition dataset from the link given below:

<https://www.kaggle.com/datasets/dhruvildave/english-handwritten-characters-dataset>

This dataset contains 3,410 images of handwritten characters in English. This is a classification dataset that can be used for Computer Vision tasks. It contains 62 classes with 55 images of each class. The 62 classes are 0-9, A-Z and a-z.

Develop a python program to recognize handwritten characters using Neural Network (NN) Model. Visualize the features from the dataset and interpret the results obtained by the model using Matplotlib library. [CO1, K3]

Use the following steps to do implementation:

1. Loading the dataset.
2. Pre-Processing the data (Image Enhancement techniques)
3. Exploratory Data Analysis.
4. Feature Engineering techniques. (Image segmentation / Image Extraction)
5. Split the data into training, testing and validation sets.
6. Train the model.
7. Test the model.
8. Measure the performance of the trained model.
9. Represent the results using graphs.

Upload the code in GitHub and include the GitHub main branch link in the assignment PDF.

Hints to do the assignment:

Do the following:

1. Load the dataset: If your computing machine could not handle all the images at once, select at least 20 images for each class (62 classes * 20 images = 1240 images). Use these 1240 images as train set.
2. Select at least 5 images for each class (62 classes * 5 images = 310 images). Use these 310 images as test set.
3. Image Pre Processing (refer this Kaggle code)
<https://www.kaggle.com/code/mohammadkumail/handwritten-character-recognition-deep-learning>
Or
<https://www.ris-ai.com/handwritten-character-recognition>
4. Exploratory Data Analysis – Draw Histogram Equalization

Refer to https://docs.opencv.org/4.x/d5/daf/tutorial_py_histogram_equalization.html

5. Build the Simple PLA model using the training set. Evaluate the model using test set. Iterate the algorithm for certain epochs (100 with learning rate 0.01) Obtain the Error.
 6. Build the Multi -Layer PLA model using the training set. Evaluate the model using test set. Iterate the algorithm for certain epochs (100 with learning rate 0.01 – Change this for better performance, conduct different experiments and obtain the best values for number of epochs, learning rate.) Obtain the Error.
 7. Use the Keras python library and construct a simple Convolution Neural Network Architecture. Display the CNN Model Summary.
8. Obtain the Training and Test Accuracy.
 9. Compare all three models (Simple PLA, Multi-Layer PLA and CNN) Write your inference about the results.
 10. Draw ROC curve for training and test accuracy (or error)

Refer to <https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-python/>

Or

<https://www.statology.org/plot-roc-curve-python/>

Upload python project in GitHub and explore all git commands. Git Commands Tutorial : <https://git-scm.com/docs/gittutorial>

Upload VSCode to GitHub

<https://www.youtube.com/watch?v=vRxfnHtCxEO>

Additional Reference:

<https://www.youtube.com/watch?v=ixszqWHYmC0>

or

https://www.youtube.com/watch?v=JrCC66R_EhQ

Optional Question

11. Use Google Co Lab GPU to load all 3,410 images and train the data using Simple PLA, Multi-Layer PLA and CNN. Compare the results obtained from each model on full dataset.