Nile University
Communication and Information Technology Program
Fall 2018

CIT690-A – Selected Topics in Machine Learning

Machine Learning - Assignment #1

(Due on: November 17, 2018 at mid-night)

a) You are required to design a Perceptron-based classification algorithm that can recognize scanned images of the 10 digits (0 to 9) provided in the file "Assignment 1 Dataset.zip". The zip file contains three folders: "Train", "Validation" and "Test". The "Train" folder contains 240 images for each digit, while each of the "Validation" and "Test" folders contain 20 images for each digit. The images in the "Train" folder should be used to train a classifier for each digit using the method given at the bottom of slide 9 in Lecture 2.pdf. The folder contains a file named "Training Labels.txt" which includes the labels of the 2400 images in order. You need to train the classifiers using each of the following values for the learning rate $\eta = 1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}, 10^{-9}$. For all Perceptrons, use an initial weight vector that has 1 as the first component (w_1) and the rest are zeros. After the classifiers are trained, test each classifier using the images given in the "Test" folder. The folder also contains a text file named "Test Labels.txt" which include the labels of the 200 images in order.

Deliverables:

- Your code.
- A confusion matrix for each value of η showing the number of images of the "Test" folder of each digit that were classified to belong to different digits (For example: Number of images of 0 that were classified as 0, 1, 2, …, 9, and so on for other digits). Convert the confusion matrices to images and save them as "Confusion-*x*.jpg", where *x* is absolute value of the power of 10 of η.

b) Use the data in the "Validation" folder to find the value of η that achieves the best accuracy for each digit classifier. Use the best classifier of each digit to classify the data in the "Test" folder. The "Validation" folder also contains a text file named "Validation Labels.txt" which include the labels of the 200 images in order.

Deliverables:

- Your code.
- A confusion matrix showing the number of images of the "Test" folder of each digit that were classified to belong to different digits (For example: Number of images of 0 that were classified as 0, 1, 2, …, 9, and so on for other digits). Convert the confusion matrix to an image and save it as "Confusion-b.jpg".

Important Notes:

- Do not use Python built-in functions for the Perceptron. You have to implement your own version of all needed functions. You are only allowed to use functions that load images into Python.
- This is an individual assignment. It is not a team assignment.