German University in Cairo
Faculty of Media Engineering and Technology
Winter 2019

CSEN 1022 – Machine Learning

Assignment #1

(Due on November 1 at mid-night)

(This assignment can be done in teams of maximum 2 students – Please include a text files with your names and IDs in the submission)

You are required to design a least squares-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 two folders: "Train" and "Test". The "Train" folder contains 240 images for each digit, while the "Test" folder contains 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 18 in Lecture 3.pdf. The folder contains a file named "Training Labels.txt" which includes the labels of the 2400 images in order. 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 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.jpg".

Important Notes:

- If the inverse of $\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}$ results in a singular matrix, use the pseudoinverse function.
- Do not use Scikit learn or Scipy built-in functions for the least squares classifier. You have to implement your own version of all needed functions. You are allowed to use numpy functions.