

Machine Learning, Winter 2022
Mini Project 2
Due date is December 13, 2022 at 11:59 PM
Submitted in teams of maximum 2

You are required to design a multiple 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 in Lecture 6 slide 25. 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:

- a) Your code.
- b) 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:

- a) If the inverse of $\tilde{X}^T \tilde{X}$ results in a singular matrix, use the pseudoinverse function.
- b) You have to implement the least squares classifier. Don't use built-in function in Python, Scikit learn or Scipy built-in functions for least squares classifier. You are allowed to use numpy and its functions for transposing, inverting and multiplying matrices.

PLAGIARISM IS NOT TOLERATED AND COPIED WORK WILL BE AWARDED 0 POINTS FOR BOTH TEAMS INVOLVED! **There will be an evaluation for this assignment in the week right after the submission**