German International University
Faculty of Informatics and Computer Science
Winter 2021

ICS504 – Machine Learning

Assignment #1

(Due on November 29 at mid-night)
(Submit to ics504ml@gmail.com)

You are required to design a least squares-based classification algorithm that can recognize the Iris flower type in the Iris dataset. You will be using the Iris dataset provided in the file "Iris.zip". The zip file contains two csv files: "Train.csv", which represents the training dataset, and "Test.csv", which represents the testing dataset. Each row in the files is represented with three columns: "PetalLengthCm", "PetalWidthCm" and "Species". The first two columns represent the two dimensions of the data, while the third column represents the flower type. The objective is to classify each data point (row) in the file "Test.csv" to one of the three types of flowers (Setosa, Versicolor, or Virginica) using the least squares classifier and the multi-class classification method given in the bottom of slide 18 in Lecture 3.pdf.

Deliverables:

- Your code.
- The classification accuracy for the test data.
- A confusion matrix showing the number of samples in test data of each flower type that were classified to belong to different flowers (For example: Number of flowers of iris_setosa that were classified as iris_setosa, iris_virginica and iris_versicolor, and so on for other types). Convert the confusion matrix to an image and save it as "Confusion.jpg".

Notes:

- This is a group assignment, where each group would be a maximum of 2 students.
- You have to implement the least squares classifier. Don't use built-in function in Python for least squares classifier. You are allowed to use numpy and its functions for transposing, inverting and multiplying matrices.