Lab Assignment 4

Part 2

Introduction to Machine Learning

AY 2021-22, Semester - I

Instructions:

1. Prepare one code containing the question and submit it in .py format.
2. Prepare a single report (*pdf file*) containing all the steps that you have followed to obtain the result. There is no need to add theoretical descriptions.
3. Put both the code and the report in a folder named <Lab4_YourRollNo>, create a zip file and upload in google-classroom.
4. Any submission received in another format or after the deadline will not be evaluated.

K-nearest neighbour

1. Download the dataset from the link below. https://drive.google.com/file/d/1s2IhEwbbSAGEtVuPpLwLq_wQ8P3Svf0l/view?usp=sharing
2. Use a distance/similarity metric of your choice, and train-val-test splits of 70:15:15.
3. Choose any 5 different values of K of your choice. Write a code to perform K-nearest neighbor classification from scratch (without using any library). Use cross-validation to identify the best value of K from the five values previously chosen. Plot a graph of error rate vs K, and mention the results briefly in the report. Generate the accuracy and confusion matrix for each value of K as well as the optimal K, and compare the results.
4. Take the optimal value of K and compare the results obtained from the model(from scratch) and the model (from sklearn library).