Lab-2 (K-Nearest Neighbour)
CSL7670 - Fundamentals of Machine Learning

NOTE:

1. This assignment contains 2 problems. Please go through the references carefully before starting the lab assignment.
2. Objective of this assignment is to get familiarity with K-NN.
3. Submit a brief report. The Lab Report template will be shared separately.
4. **Deadline:** August 22, 2023, 10:30 PM.

---

1. **(Apple vs Orange)** You are given a K-NN code for Apple vs Orange problem. Please read and understand the code. Now perform the following tasks: (a) Synthetically increase the dataset size to 50 samples, (b) Edit the code so that random 80%, 10%, and 10% samples are used for training, testing, and validation respectively. (c) Change the value of K to 3, 5, and 7 and compare the validation set and test set results. (d) Write a code that draws confusion matrices for different K. Use the following link to understand about Confusion Matrix.
2. **(Handwritten Digit Classification)** Use the code provided for classifying the handwritten digits of the MNIST dataset. Read and understand the code. Now, (a) Modify the code so that it uses L1-distance instead of the default L2-distance (Euleadean). (b) Find out the K that gives better performance. (c) Report the Accuracy and (d) Display results by showing the image, actual label, and predicted label. Find out a few samples where the predicted label is incorrect.

References:

1. Sk-learn library for K-NN

2. Confusion Matrix
3. Train-test split

End of Paper