CS460G – Machine Learning

Assignment # 1

k-NN

**Note:** Please implement these algorithms from scratch. Do not use libraries that provide implementation for the core parts of the ML algorithms.

This problem will test you for the following learning objectives:

* I have acquired or improved my ability to design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of a machine learning problem. (SO2). Questions: 1, 2, and 4.
* I am able to effectively communicate about machine learning models or concepts in a variety of professional contexts. (SO3). Questions: 3 and 5.

1. Implement k-NN from scratch (preferably in python) and test your code on a tiny dataset of your choice. You might want to break down the full algorithm into small functions. 10 pts
2. Write code to compute confusion matrix. Use the confusion matrix to compute accuracy, precision, recall, F1-score (micro/macro/weighted). Test your code on a tiny dataset. 10 pts
3. Use UCI website (<https://archive.ics.uci.edu/ml/datasets/iris>) to download IRIS dataset. Explore webpage to know about the dataset and write a simple paragraph about the dataset. 10 pts
4. Adapt your k-NN algorithm for the above data. 60 pts
   1. Split your dataset into 3 sets (training/validation/test).
   2. Use validation set to find the best k.
   3. Report the results on the test set for the selected k.
      1. Accuracy, Precision, Recall, F1-score (micro/macro/weighted).
5. Try different distance functions, e.g., Manhattan distance, Euclidean distance, and report your findings on what works best in a short paragraph. 10 pts