In [ ]:

Lab 2 (Classification1)

Question 1

1. Read the Iris dataset from Scikit Learn (a copy of the dataset is provided) using sklearn datasets.load.
2. use scikit-learn train_test_split to randomly split the data into training (75%) and testing (25%) set.
3. Using scikit-learn neighbors.KNeighborsClassifier to build a basic kNN classifier model for this dataset. You can use either metrics.accuracy_score, or knn.score to obtain the prediction accuracy.
4. Explore the impact of adopting various values of k (k=1, 3, 5, 7) on your model, different distance metrics and algorithms.
5. Next normalise the data and re-run the experiment to see the impact on the prediction.

Question 2

1. Read the Cancer dataset, which is available within DataLab2.zip, into a NumPy array.
2. Using Scikit Learn build a basic kNN classifier model for this dataset (start with k=1 to 100, and then adjut the upper k value to see details) and assess its classification accuracy.