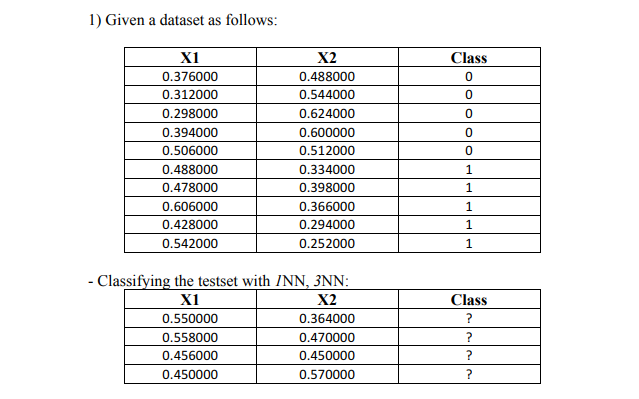
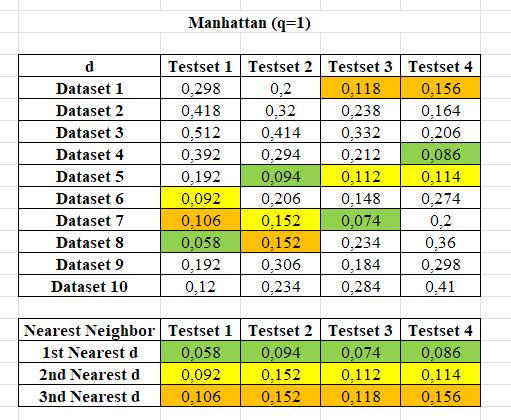
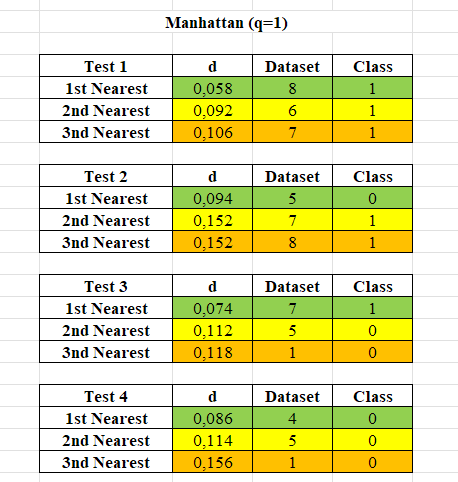
Name: Truong Dang Truc Lam ID: B2111933 Class: CT205H - M04

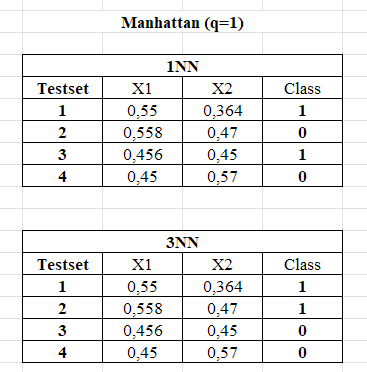
# k nearest neighbors



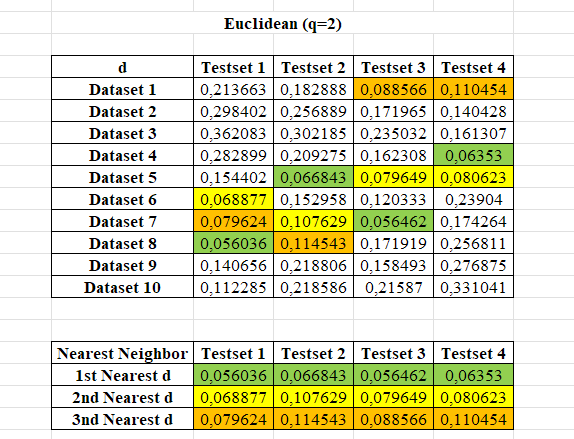
Solution with Manhattan Distance:

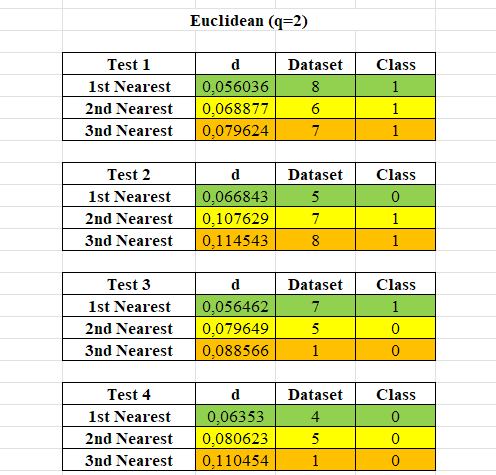


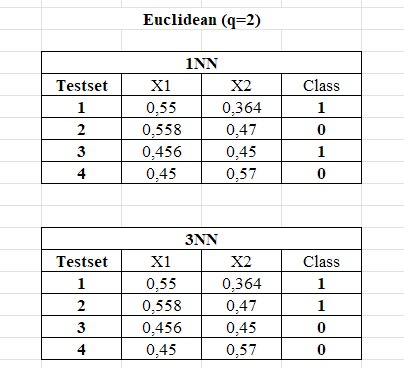


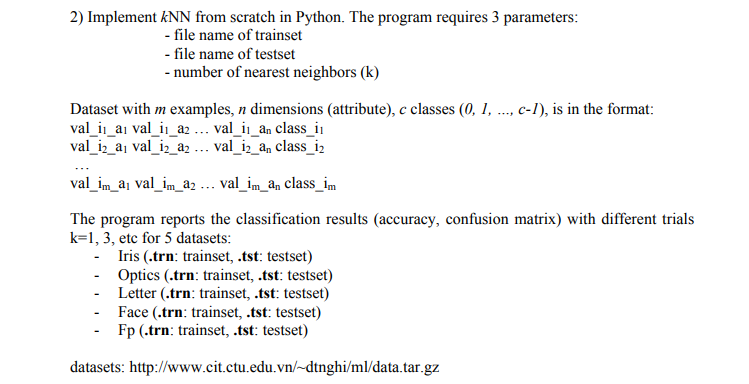


Solution with Euclidean Distance:

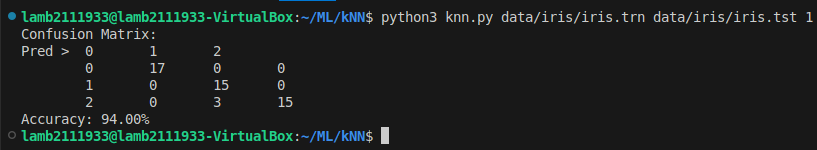




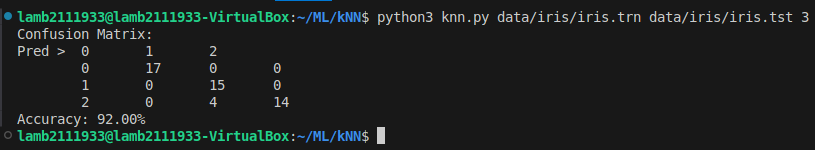




Iris dataset:

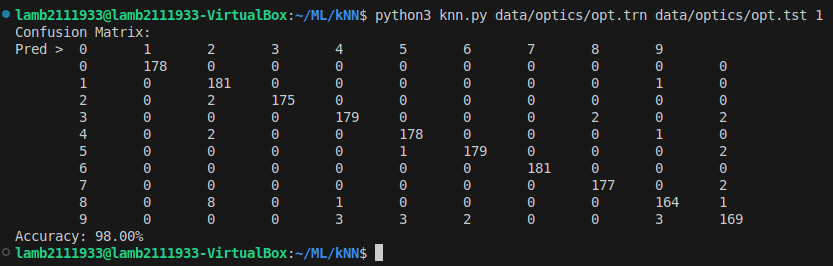


Iris with 1NN

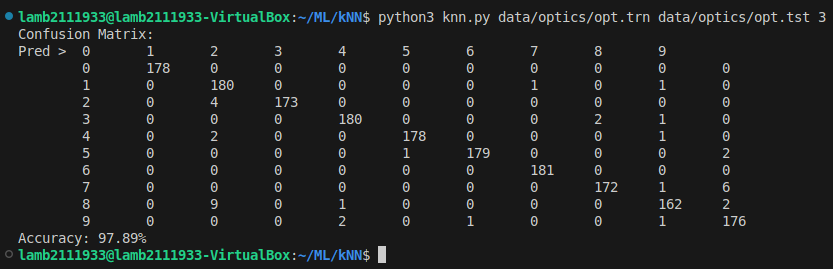


Iris with 3NN

Optics dataset:



Optics with 1NN

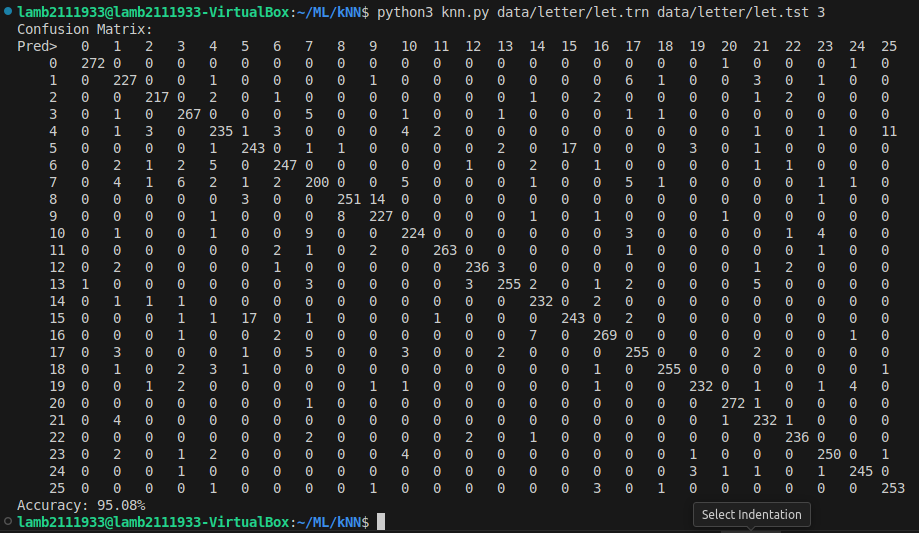


Optics with 3NN

Letter dataset:

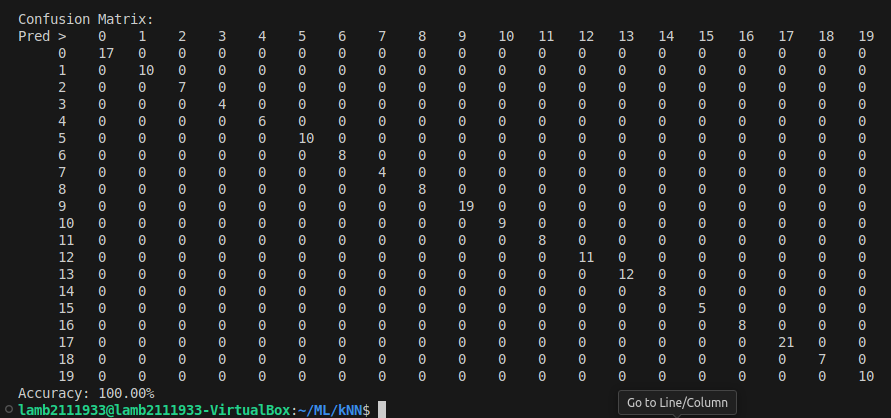


Letter with 1NN

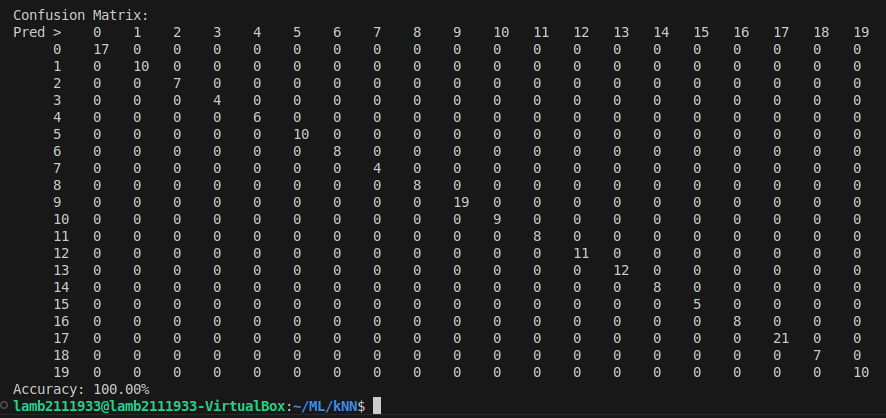


Letter with 3NN

Face dataset:

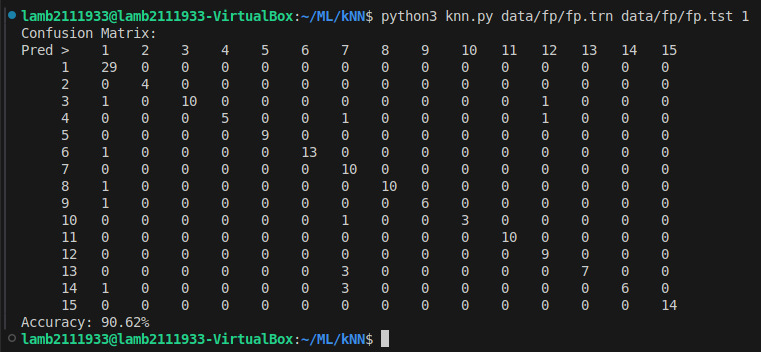


Face with 1NN

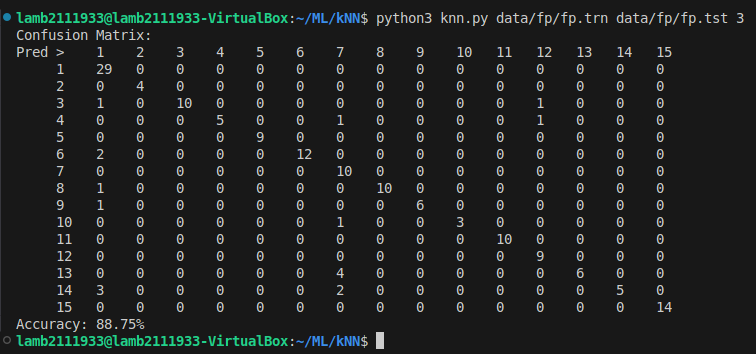


Face with 3NN

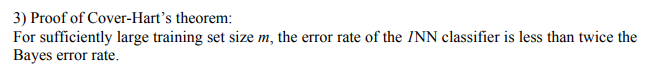
Fp dataset:



Fp with 1NN



Fp with 3NN



The error rate of a classifier is the probability that it makes an incorrect prediction. The Bayes error rate as the minimum possible error rate, which is the error of the optimal classifier that knows the true distribution of the data.

Let ***X*** be the input space which represents the set of all possible input points.

Let ***f*** denote the Bayes decision function which assigns the correct class label to each input ***x*** in ***X*** based on the true conditional probabilities of the data.

***f(x)*** is the class label **y** that maximizes the probability ***P(Y = y | X = x)***, where ***Y*** is the class label and ***P(Y | X)*** is the true conditional distribution of the labels given the inputs.

Consider a training set consisting of ***m*** labeled points drawn independently from the joint distribution ***P(X, Y)***. The 1NN classifier assigns a label to each test point ***x*** in ***X*** by finding the closest point in the training set and using its label as the prediction. The error rate of the 1NN classifier is the probability that the predicted label differs from the true label for a given test point **x**:

***Perror(1NN) = P(1NN classifier's prediction ≠ f(x))***

For sufficiently large ***m***, the 1NN classifier makes predictions that are very close to the predictions of the Bayes decision function ***f***. As ***m*** increases, the training points densely cover the input space ***X***, and the nearest neighbor of a test point ***x*** becomes increasingly close to ***x***. The label assigned by the 1NN classifier approaches the label predicted by ***f(x)***, which minimizes the probability of error. We can see that for a large ***m***, the error rate of the 1NN classifier can satisfiy the inequality:

***Perror(1NN) ≤ 2 \* Perror(Bayes)***

This is true because as ***m*** increases, the probability that a random test point will have its nearest neighbor close to the Bayes decision boundary becomes larger. As the training set grows, the expected nearest neighbor can become a good approximation of the Bayes decision function.

Beside, when the number of training samples increases, the 1NN classifier can have more information, which will lead to a better approximation of the true decision boundary and thus a lower error rate.

In the limit as ***m*** approaches **∞**, the 1NN classifier converges to the optimal Bayes classifier. The error rate of the 1NN classifier approaches the Bayes error rate, and the inequality will hold with equality when the training data is ∞.