HOMEWOEK 1

Runlin Hou
ECE, School Of Graduate Studies
Rutgers University
hourunlinxa@gmail.com

PROBLEM 1

According to the data, the three points with the smallest L2 distance to test data is.

$$L2_{A1} = \sqrt{(0-1)^2 + (1-0)^2 + (1-1)^2} = \sqrt{2}$$
$$L2_{C2} = \sqrt{(0-1)^2 + (-1-0)^2 + (1-1)^2} = \sqrt{2}$$
$$L2_{A0} = \sqrt{(0-1)^2 + (1-0)^2 + (0-1)^2} = \sqrt{3}$$

When $K = 1$, the point chose to decide the class of the test data might be A1 or C2. So test data might be classified to be A or C in a same probability.
When $K = 2$, the points chose to dicide the class of the test data would be A1 and C2, which means test data got the same probability to be classified in A and C.
When $K = 3$, all the three points would be chose to decide the class of test data. So the test data would be classified to be A.

PROBLEM 2

In this problem, we are supposed to use the KNN to classify the data into 4 classes – $[0, 1, 2, 3]$

And the following picture shows the distribution of the training data and test data. As we can see, the stars are the test data predicted by the model we established and the dots are the training data.

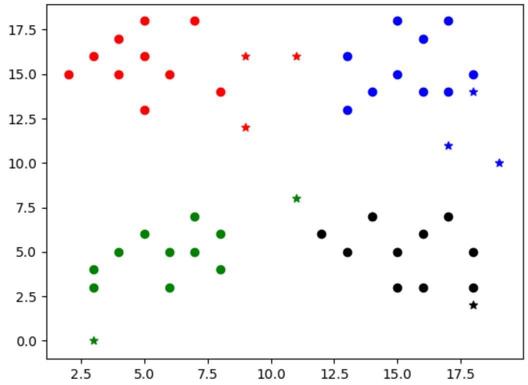


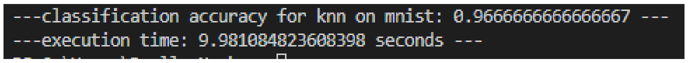
Fig. 1. miniKNN

PROBLEM 3

In this section, the data we accepted is a 28×28 image. To accomplish the training process, we need to do convert the images into one dimensional array. Here I use `append()` to merge the 28 row into an array with 784 attributes.

And the trainning process can be achieved with the same code from problem 2.

As we can see, I choose the data from 600 to 630 and I finally achieve an accuracy of 97%. The whole process took about 10 seconds.



```
---classification accuracy for knn on mnist: 0.9666666666666667 ---
---execution time: 9.981084823608398 seconds ---
```

Fig. 2. KNN