

# Homework 1

Haocong Wang

## Q1

Based on the given data, the three points with the smallest L2 distance are:

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

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

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

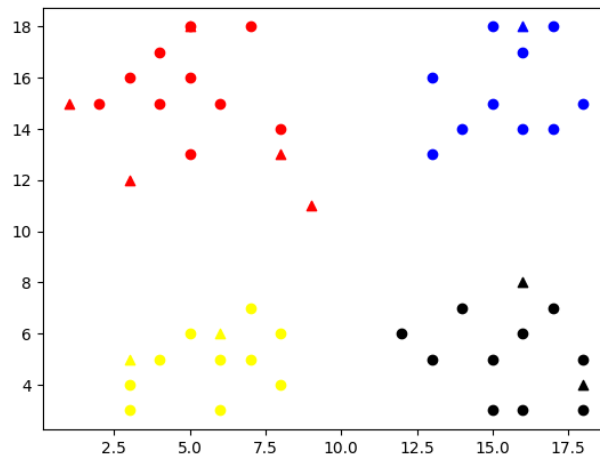
If  $K = 1$ , then the point chosen to decide the class of the test data might be A1 or C2, so the test data might be classified to be A or C under same probabilities.

If  $K = 2$ , the two points will be A1 and C2, which means the test data will be classified in A and C with same probabilities.

If  $K = 3$ , all the three points will be chosen to decide the class, so the test data will be classified to be A.

## Q2

The following picture shows the distribution of the training and test data, in which round and triangle represent training and test data respectively. The value of  $k$  is set to be 10.



## Q3

All the input are  $28 \times 28$  images, so I first convert them into 1-D arrays. Using the code in Q2, I finished the training process. I chose the test data from 500 to 550 as the test set. The result is shown below with accuracy as 0.98 and execution time as 17.4 seconds.

```
---classification accuracy for knn on mnist: 0.98 ---  
---execution time: 17.409408569335938 seconds ---
```