Machine Learning
Fall 2022
HW1
Due Date: Sept. 6 (Tue), 11:59pm via Blackboard

Please submit 1) A report that summarizes the brief steps and results in pdf file format and 2) well documented Matlab codes

Matlab basic programming

1. In Matlab, create matrices $M = \begin{bmatrix} 1 & 10 & 6 \\ 2 & 3 & 5 \\ 15 & 9 & 4 \end{bmatrix}$ and $N = \begin{bmatrix} 1 & 2 \\ 6 & 9 \\ 3 & 7 \end{bmatrix}$. assign the second row of M to row vector A, and assign the first column of N to column vector B. Please calculate M × N, M^2 ,$\sqrt{N}$and B × A.

2. Multiply all elements of the Matrix M in Problem 1 by 5. Use two methods to solve the problem 1) Use the for loop, and access every elements in M; 2) Use Matlab built-in function ones())

3. Let $S = \begin{bmatrix} 5 & 3 & 8 \\ 2 & 1 & 7 \\ 6 & 4 & 8 \end{bmatrix}$. Use the Matlab built-in function mean() or sum() to solve below operations:

$$G = [5+2+6 \quad 3+1+4 \quad 8+7+8]$$

$$F = \begin{bmatrix} \frac{5+3+8}{3} \\ \frac{2+1+7}{3} \\ \frac{6+4+8}{3} \end{bmatrix}$$

KNN exercise

4. Exploring K nearest neighbor (KNN) for classification (assuming you are the inventor of this method): The following table shows the results of 10 bolts in quality tests 1 and 2, and their status of "pass" or "fail". Based on these results, please judge if a bolt which gets 30 in test 1 and 80 in test 2 will pass or fail. Write a MATLAB code to predict the result using the method of nearest neighbor of k students (k=1, 3, and 5, respectively). In each neighbor of k students, find the status that has more counts, and the new case will get this status. **Please follow the steps below.**

Test 1 (x)	Test 2(y)	Status (c)
67	30	0
63	79	1
78	82	1
51	48	0
80	65	1

90	61	1
96	15	0
83	22	0
72	67	1
50	60	0

0: Fail, 1:Pass

```
ans = 'fail'
```

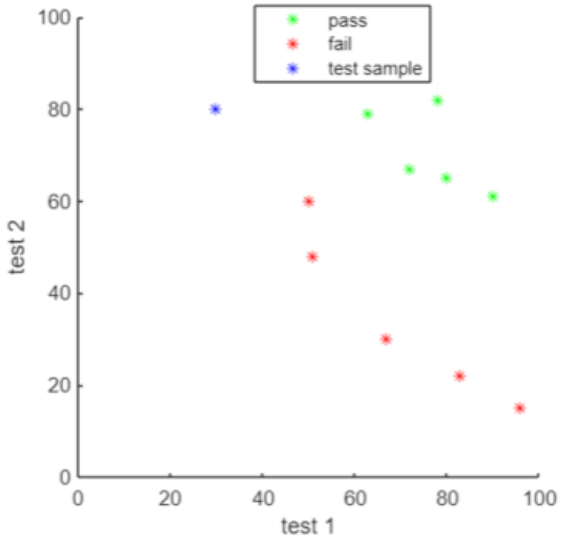

Fig. 1

Steps:

1. Define the above training samples in Matlab. In Matlab, you need to assign each column to a row vector, then we will have 3 vectors **x, y, c**. For example, sample 1(the first row in the above table) is (**x**(1), **y**(1)) and the result of sample 1 is **c**(1)
2. Define the test sample (30, 80), which represents the new case that we need to classify (fail or pass)
3. Compute the distance between all training samples and the test sample (30, 80). If the distance between two samples is small, then they are very similar. Here we choose the Euclidean distance to represent the distance between one training sample and the test sample. The formula for the Euclidean distance between training sample i and the test sample is

$$\text{distance}_i = \sqrt{(x(i) - 30)^2 + (y(i) - 80)^2}.$$

Then you need to use this formula to compute all the distance with Matlab. You can use the built-in function **sqrt()**. At last, you will get a row vector:

$$\text{Distance} = [distance_1, distance_2, \cdots, distance_{10}]$$

Corresponding to

$$\mathbf{c} = [0, 1, \cdots, 0]$$

4. Find the k nearest neighbor (k=1, 3, and 5, respectively). The neighbor when k=1 is shown in the **Fig. 1**. Under each value of k, find the status (fail or pass) that has more counts, and the new case (student) will get this status. Please note the classification results may be different under different values of k.