LEON MARCO A. DEVELA
COM232

ASSIGNMENT #1

Using k-Nearest Neighbours, predict if a student will Pass (1) or Fail (0) based on:

- x_1 = hours studied
- x_2 = sleep hours

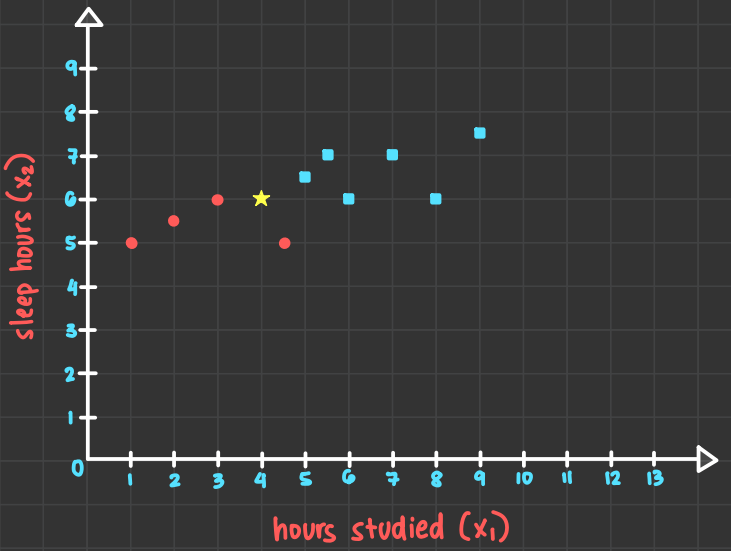
TRAINING DATA

STUDENT	HOURS STUDIED (x_1)	SLEEP HOURS (x_2)	PASS/FAIL (Y)
1	1.0	5.0	0 (FAIL)
2	2.0	5 5	0 (FAIL)
3	3.0	6.0	0 (FAIL)
4	4.5	5.0	0 (FAIL)
5	5.0	6.5	1 (PASS)
6	5.5	7.0	1 (PASS)
7	6.0	6.0	1 (PASS)
8	7.0	7.0	1 (PASS)
9	8.0	6.0	1 (PASS)
10	9.0	7.5	1 (PASS)

TASK

A new student studied 4 hours and slept 6 hours. We want to predict whether they PASS or FAIL using KNN with k=3.

PLOT ALL DATA POINTS USING A SCATTER PLOT



LEON MARCO A. DEVELA
COM232

ASSIGNMENT #1
CCMACLRL

COMPUTE DISTANCES

For the new student, compute the EUCLIDEAN DISTANCE from all data points

$$d=\sqrt{(x_1-4)^2+(x_2-6)^2}$$

STUDENT	(X_1)	(X_2)	(Y)	EUCLIDEAN DISTANCE
1	1.0	5.0	0 (FAIL)	3.16
2	2.0	5.5	0 (FAIL)	2.06
3	3.0	6.0	0 (FAIL)	1.0
4	4.5	5.0	0 (FAIL)	1.12
5	5.0	6.5	1 (PASS)	1.12
6	5.5	7.0	1 (PASS)	1.80
7	6.0	6.0	1 (PASS)	2.0
8	7.0	7.0	1 (PASS)	3.16
9	8.0	6.0	1 (PASS)	4
10	9.0	7.5	1 (PASS)	5.22

SOLUTION:

$d_1=\sqrt{(1-4)^2+(5-6)^2}=3.16$

$d_2=\sqrt{(2-4)^2+(5.5-6)^2}=2.06$

$d_3=\sqrt{(3-4)^2+(6-6)^2}=1.0$

$d_4=\sqrt{(4.5-4)^2+(5-6)^2}=1.12$

$d_5=\sqrt{(5-4)^2+(6.5-6)^2}=1.12$

$d_6=\sqrt{(5.5-4)^2+(7-6)^2}=1.80$

$d_7=\sqrt{(6-4)^2+(6-6)^2}=2.0$

$d_8=\sqrt{(7-4)^2+(7-6)^2}=3.16$

$d_9=\sqrt{(8-4)^2+(6-6)^2}=4$

$d_{10}=\sqrt{(9-4)^2+(7.5-6)^2}=5.22$

FIND THE 3 NEAREST NEIGHBORS

Highlight the rows with the 3 smallest distances

STUDENT	(X_1)	(X_2)	(Y)	EUCLIDEAN DISTANCE
1	1.0	5.0	0 (FAIL)	3.16
2	2.0	5.5	0 (FAIL)	2.06
3	3.0	6.0	0 (FAIL)	1.0
4	4.5	5.0	0 (FAIL)	1.12
5	5.0	6.5	1 (PASS)	1.12
6	5.5	7.0	1 (PASS)	1.80
7	6.0	6.0	1 (PASS)	2.0
8	7.0	7.0	1 (PASS)	3.16
9	8.0	6.0	1 (PASS)	4
10	9.0	7.5	1 (PASS)	5.22

MAJORITY VOTE

- Count how many are PASS (1) and how many are FAIL (0)
  - PASS (1) = 1
  - FAIL (0) = 2
- Predict the outcome for the new student

  A new student who studied 4 hours and slept 6 hours will FAIL (0) the exam

LEON MARCO A. DEVELA
COM232

ASSIGNMENT #1
CCMACLRL

DISCUSSION QUESTION

1. What was your final prediction?

   With k=3, my final prediction is that the new student who studied 4 hours and slept 6 hours will FAIL(0) the exam

2. How would the prediction change if we use k=5 instead of k=3

STUDENT	(X_1)	(X_2)	(Y)	EUCLIDEAN DISTANCE
1	1.0	5.0	0 (FAIL)	3.16
2	2.0	5.5	0 (FAIL)	2.06
3	3.0	6.0	0 (FAIL)	1.0
4	4.5	5.0	0 (FAIL)	1.12
5	5.0	6.5	1 (PASS)	1.12
6	5.5	7.0	1 (PASS)	1.80
7	6.0	6.0	1 (PASS)	2.0
8	7.0	7.0	1 (PASS)	3.16
9	8.0	6.0	1 (PASS)	4
10	9.0	7.5	1 (PASS)	5.22

As you can see on the table, if we change the k valve into k=5, the prediction will change. With PASS(1)=3 and FAIL(0)=2, the new student who studied 4 hours and slept 6 hours will PASS(1) the exam since the majority of nearest neighbors are PASS(1).