GABRIEL ANGELO VIÑAS

COM232

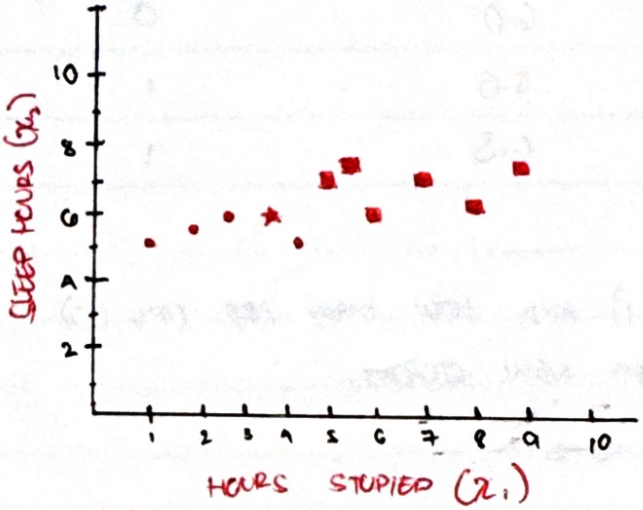
ASSIGNMENT # 1

THEK

A 14Y/. STUDENT STUDIED 4 HOURS AND SLEPT 6 HOURS. WE WANT TO PREDICT WHETHER THEY PASS OR FAIL USING KNN WITH K= 3.

1) PLOT ALL DATA POINTS USING A SCATTER PLOT

- USE A CIRCLE ● FOR STUDENTS WHO FAILED.
- USE A SQUARE ■ FOR STUDENTS WHO PASSED.
- USE A STAR ★ FOR THE UNKNOWN DATA POINT.



2) COMPUTE DISTANCES

FOR THE NEW STUDENT, COMPUTE THE EUCLIDEAN DISTANCE FROM ALL DATA POINTS:

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

STUDENTS	HOURS STUDIED (x_1)	SLEEP HOURS (x_2)	PASS/FAIL (Y)	EUCLIDEAN DISTANCE
1	1.0	5.0	0	3.162
2	2.0	5.5	0	2.062
3	3.0	6.0	0	1.0
4	4.5	5.0	0	1.118
5	5.0	6.5	1	1.118
6	5.5	7.0	1	1.803
7	6.0	6.0	1	2.0
8	7.0	7.0	1	3.162
9	8.0	6.0	1	4.0
10	9.0	7.5	1	5.22

SOL:

1) $d = \sqrt{(1-4)^2 + (5-6)^2} = 3.162$
2) $d = \sqrt{(2-4)^2 + (5.5-6)^2} = 2.062$
3) $d = \sqrt{(3-4)^2 + (6-6)^2} = 1.0$
4) $d = \sqrt{(4.5-4)^2 + (5-6)^2} = 1.118$
5) $d = \sqrt{(5-4)^2 + (6.5-6)^2} = 1.118$
6) $d = \sqrt{(5.5-4)^2 + (7-6)^2} = 1.803$
7) $d = \sqrt{(6-4)^2 + (6-6)^2} = 2.0$
8) $d = \sqrt{(7-4)^2 + (7-6)^2} = 3.162$
9) $d = \sqrt{(8-4)^2 + (6-6)^2} = 4.0$
10) $d = \sqrt{(9-4)^2 + (7.5-6)^2} = 5.22$

2) FIND THE 3 NEAREST NEIGHBORS

- HIGHLIGHT THE ROWS WITH THE 3 SMALLEST DISTANCES.

STUDENT	HOURS STUDIED (x_1)	SLEEP HOURS (x_2)	PASS/FAIL (y)	EUCLIDEAN DISTANCE
3	3.0	6.0	0	1.0
4	4.5	5.6	1	1.12
5	5.0	6.5	1	1.12

3) MAJORITY VOTE

- COUNT HOW MANY ARE PASS (1) AND HOW MANY ARE FAIL (0)
- PREDICT THE OUTCOME FOR THE NEW STUDENT.

FAIL (0): STUDENTS 3 & 4 = 2
PASS (1): STUDENT 5 = 1
PREDICTION: ~~FAIL (0)~~ A NEW STUDENT WHO STUDIED 4 HOURS AND SLEPT 6 HOURS WILL MORE LIKELY TO FAIL THE EXAM.

4) DISCUSSION QUESTIONS

1) WHAT WAS YOUR FINAL PREDICTION?
- THE NEW STUDENT WILL FAIL.

2) HOW WOULD THE PREDICTION CHANGE IF WE USED $K = 5$ INSTEAD OF $K = 3$?
- IF $K = 5$, THEN THE NEAREST NEIGHBORS WOULD BE STUDENTS 3, 4, 5, 6, 7. PASS COUNT = 3 AND FAIL COUNT = 2, SO THE PREDICTION CHANGES TO PASS.