



NATIONAL UNIVERSITY - MANILA
COLLEGE OF COMPUTING AND INFORMATION TECHNOLOGY

Introduction to Machine Learning

K-NEAREST

NEIGHBOURS

(PREDICTING PASS OR FAIL)

ASSIGNMENT NO. 1

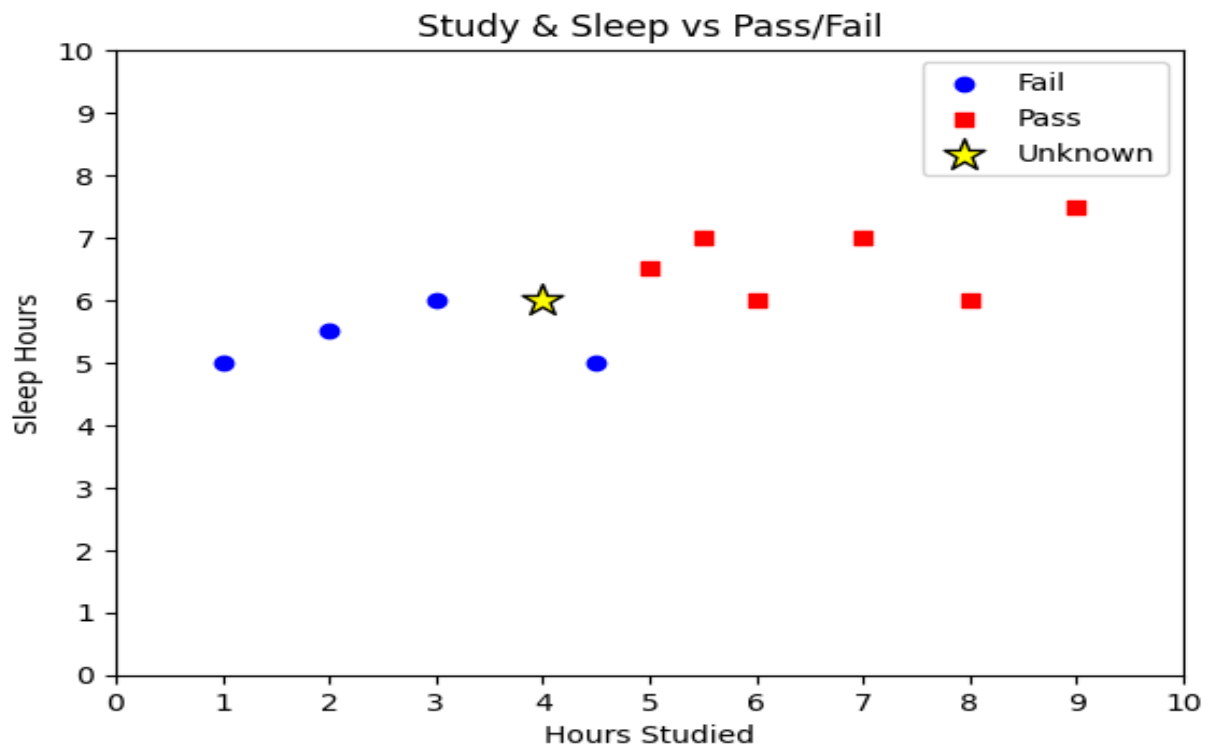
Prepared by:

Rae S. Paulos

TRAINING DATA (10 STUDENTS)

Student	Hours Studied (X1)	Sleep Hours (X2)	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	5.5	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)

1. Plot all data using a scatter plot (10 points)



2. Compute Distances (10 points)

Student	Hours Studied (X1)	Sleep Hours (X2)	Pass/Fail (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	5.5	0 (Fail)	1.11
5	5.0	6.5	1 (Pass)	1.11
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)	4.0
10	9.0	7.5	1 (Pass)	5.22

Solution:

#	Solution	Answer
1	$1. \quad x_1 = 1.0, x_2 = 5.0$ $d = \sqrt{(1.0 - 4)^2 + (5.0 - 6)^2}$ $= \sqrt{9 + 1}$ $= \sqrt{10} = \underline{\underline{3.16}}$	3.16
2	$2. \quad x_1 = 2.0, x_2 = 5.5$ $d = \sqrt{(2.0 - 4)^2 + (5.5 - 6)^2}$ $= \sqrt{4 + 0.25}$ $= \sqrt{4.25} = \underline{\underline{2.06}}$	2.06

3	$ \begin{aligned} 3. \quad x_1 &= 3.0, \quad x_2 = 6.0 \\ d &= \sqrt{(3.0 - 4)^2 + (6.0 - 6)^2} \\ &= \sqrt{1 + 0} \\ &= \sqrt{1} = \underline{\underline{1}} \end{aligned} $	1.0
4	$ \begin{aligned} 4. \quad x_1 &= 4.5, \quad x_2 = 5.0 \\ d &= \sqrt{(4.5 - 4)^2 + (5.0 - 6)^2} \\ &= \sqrt{0.25 + 1} \\ &= \sqrt{1.25} = \underline{\underline{1.11}} \end{aligned} $	1.11
5	$ \begin{aligned} 5. \quad x_1 &= 5.0, \quad x_2 = 6.5 \\ d &= \sqrt{(5.0 - 4)^2 + (6.5 - 6)^2} \\ &= \sqrt{1 + 0.25} \\ &= \sqrt{1.25} = \underline{\underline{1.11}} \end{aligned} $	1.11
6	$ \begin{aligned} 6. \quad x_1 &= 5.5, \quad x_2 = 7.0 \\ d &= \sqrt{(5.5 - 4)^2 + (7.0 - 6)^2} \\ &= \sqrt{2.25 + 1} \\ &= \sqrt{3.25} = \underline{\underline{1.80}} \end{aligned} $	1.80
7	$ \begin{aligned} 7. \quad x_1 &= 6.0, \quad x_2 = 6.0 \\ d &= \sqrt{(6.0 - 4)^2 + (6.0 - 6)^2} \\ &= \sqrt{4 + 0} \\ &= \sqrt{4} = \underline{\underline{2}} \end{aligned} $	2.0

8	$ \begin{aligned} 8. \quad x_1 &= 7.0, \quad x_2 = 7.0 \\ d &= \sqrt{(7.0 - 4)^2 + (7.0 - 6)^2} \\ &= \sqrt{9 + 1} \\ &= \sqrt{10} = \underline{\underline{3.16}} \end{aligned} $	3.16
9	$ \begin{aligned} 9. \quad x_1 &= 8.0, \quad x_2 = 6.0 \\ d &= \sqrt{(8.0 - 4)^2 + (6.0 - 6)^2} \\ &= \sqrt{16 + 0} \\ &= \sqrt{16} = \underline{\underline{4}} \end{aligned} $	4.0
10	$ \begin{aligned} 10. \quad x_1 &= 9.0, \quad x_2 = 7.5 \\ d &= \sqrt{(9.0 - 4)^2 + (7.5 - 6)^2} \\ &= \sqrt{25 + 2.25} \\ &= \sqrt{27.25} = \underline{\underline{5.22}} \end{aligned} $	5.22

3. Find the 3 Nearest Neighbors (3 points)

Highlight the rows with the 3 smallest distance

Student	Hours Studied (X1)	Sleep Hours (X2)	Pass/Fail (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	5.5	0 (Fail)	1.11
5	5.0	6.5	1 (Pass)	1.11
6	5.5	7.0	1 (Pass)	1.80

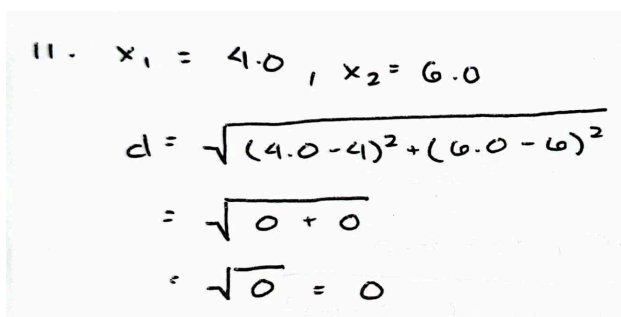
7	6.0	6.0	1 (Pass)	2.0
8	7.0	7.0	1 (Pass)	3.16
9	8.0	6.0	1 (Pass)	4.0
10	9.0	7.5	1 (Pass)	5.22

4. Majority Vote (2 points)

Count how many passed and how many failed. Predict the outcome for the new student who studied 4 hours and slept 6 hours.

Category	Count
Pass	6 students
Fail	4 students

Euclidean Distance

#	Solution	Answer
11		0

Prediction:

Student	Hours Studied (X1)	Sleep Hours (X2)	Pass/Fail (Y)	Euclidean Distance
11	4.0	6.0	0 (Fail)	0

5. Discussion Questions (5 points)

Question	Answer
What was your final prediction?	The final prediction for a student who

	slept for 6 hours and studied for 4 hours failed.
How would the prediction change if we used $k = 5$ instead of $k = 3$?	With $k = 3$, the student is predicted to fail, but with $k = 5$, the student is predicted to pass. This change occurs because increasing k shifts the decision from relying on just a few nearby points to considering a wider group of neighbors. In the smaller neighborhood, the majority outcome was failure, but when two more neighbors were added, the balance tipped toward passing. This illustrates how the choice of k directly influences the classification result.