

Stephen Ace Sy | 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 (10 Students)			
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)

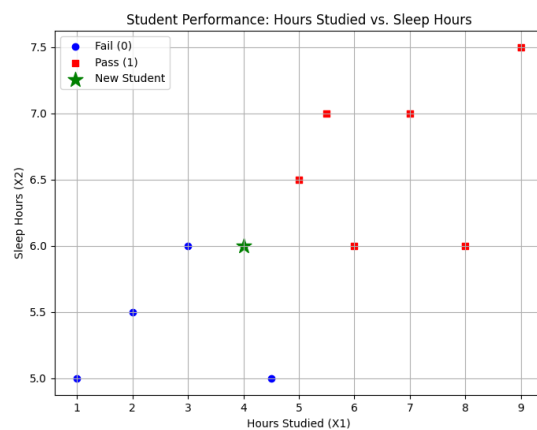
Task

A new student studied **4 hours** and slept **6 hours**.

We want to predict whether they **Pass** or **Fail** using **KNN with $k = 3$** .

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

- 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 (10 points)

- **Note:** Please show all solutions.
- For the new student, compute the **Euclidean distance** from all data points:
- $d = \sqrt{((X_1 - 4)^2 + (X_2 - 6)^2)}$

#	Student	#	Hours Studied (X1)	#	Sleep Hours (X2)	#	Pass/Fail (Y)	#	Euclidean Distance
	1		1		5		0		3.162
	2		2		5.5		0		2.062
	3		3		6		0		1
	4		4.5		5		0		1.118
	5		5		6.5		1		1.118
	6		5.5		7		1		1.803
	7		6		6		1		2
	8		7		7		1		3.162
	9		8		6		1		4
	10		9		7.5		1		5.22

3. Find the 3 Nearest Neighbors (3 points)

3	3	6	0	1
4	4.5	5	0	1.118
5	5	6.5	1	1.118

4. Majority Vote (2 points)

Fail (0): 2 votes

Pass (1): 1 vote

Prediction: So the prediction is the new student will fail the exam

5. Discussion Questions (5 points)

1. What was your final prediction?
 - The final prediction is the new student will fail the exam.
2. How would the prediction change if we used **k = 5** instead of **k = 3**?
 - With k=5, the neighbors would be Students 3, 4, 5, 6, and 7. The vote would be 2 Fails vs. 3 Passes, changing the prediction to **Pass**.