

02/04/25

## LAB - 5

### KNN classification

Q

consider the following dataset, for  $k=3$  and test data  $(X, 35, 100)$  as (Person, Age, salary) solve using KNN classifier model and predict the target.

Person	Age	salary	Target
A	18	50	N
B	23	55	N
C	24	70	N
D	41	60	Y
E	43	70	Y
F	38	40	Y
X	35	100	?

A

Person	Age	salary	Target	distance	Rank
A	18	50	N	52.8	
B	23	55	N	46.6	
C	24	70	N	31.9	2
D	41	60	Y	40.4	3
E	43	70	Y	31.1	1
F	38	40	Y	60.1	
X	35	100	?		

Step 1:-

$$\text{Euclidean distance}(d) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

$$x_1, y_1 = (25, 100)$$

$$d_1 = \sqrt{(35-18)^2 + (100-50)^2} = 52.8$$

Step-2:-

Identify 3 nearest neighbours, that is

→ E (31.1, Y)

→ C (31.9, N), → D (40.4, Y)

step-3

Predicting Target by Majority Voting  
Person X(35, 100k) is 'Y'.

(Q)

• For Iris dataset

How to choose the K value? Demonstrate using accuracy rate and error rate.

(A)

choosing K for Iris dataset

- Train KNN with different K values
- Measure accuracy (higher is better) and error rate (lower is better).
- The best K is where accuracy is highest and error rate is lowest (typically 5-10).

• For Diabetes dataset

What is the purpose of feature scaling?  
How to perform it?

(A)

Feature Scaling for Diabetes Dataset.

- KNN relies on distance, so large-scale features can dominate smaller ones.
- Standardization (Z-score normalization) ensures all features contribute equally.
- Scaling improves accuracy and prevents biased predictions.

9/4/25