

Practical-10 (K-Nearest Neighbour)

1) Calculate the Euclidean Distance between the two data points A(5,4),B(2,3)

→ Let A(5,4) = (x₁,y₁)

B(2,3) = (x₂,y₂)

Formula for Euclidean Distance is,

$$\begin{aligned}
 d &= \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \\
 &= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \\
 &= \sqrt{(2 - 5)^2 + (3 - 4)^2} \\
 &= \sqrt{(-3)^2 + (-1)^2} \\
 &= \sqrt{9 + 1} \\
 &= \sqrt{10}
 \end{aligned}$$

$$d = 3.162$$

2) Use KNN to predict given class(Yes or No), X₁={9,9,5,5}; X₂={9,6,6,7}; Y={Yes,Yes,NO,NO}; K=3 (Use Euclidean for distance calculation).What will be Y values for new records X₁=5; X₂=9; Y = ?

→ Given: K=3

X₁={9,9,5,5}

X₂={9,6,6,7}

Y={Yes,Yes,NO,NO}

X ₁	X ₂	Y	Distance	Rank
9	9	Yes	4	3
9	6	Yes	5	4
5	6	No	3	2
5	7	No	2	1

Calculating the Euclidean distance between each point from set and new record, and assigning them their ranks accordingly.

$$d_1 = \sqrt{(9-5)^2 + (9-9)^2} = \sqrt{16} = 4$$

$$d_2 = \sqrt{(9-5)^2 + (6-9)^2} = \sqrt{25} = 5$$

$$d_3 = \sqrt{(5-5)^2 + (6-9)^2} = \sqrt{9} = 3$$

$$d_4 = \sqrt{(5-5)^2 + (7-9)^2} = \sqrt{4} = 2$$

since $K=3$ (given)

:Considering 3 ranks and their respective Y values, No(rank1), NO(rank2), Yes(rank3)

: The Y value for new records, $X_1 = 5$; $X_2 = 9$ is NO

3) Below are the 7 actual values of target variable from the training data [no,no,no,yes,yes,yes,yes].What will be entropy of target variable?

→Probability of positive(yes) and negative(no) are as follows:

$$P(\text{Yes}) = 4/7; P(\text{No})=3/7$$

we calculate entropy by putting probability values in formula :

$$\text{entropy} = -P \log_2 P - N \log_2 N$$

$$= (-4/7) \log_2 (4/7) - 3/7 \log_2 (3/7)$$

$$= (-0.571) \times (-0.808) - 0.428(-1.224)$$

$$= 0.461 - (-0.523)$$

$$\text{Entropy} = 0.984$$