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) = (x_1, y_1)

   B(2,3) = (x_2, y_2)

Formula for Euclidean Distance is,

$$d = \sqrt{\sum_{i=1}^{n}(xi - yi)^2}$$

$$= \sqrt{(x2 - x1)^2 + (y2 - y1)^2}$$

$$= \sqrt{(2 - 5)^2 + (3 + 4)^2}$$

$$= \sqrt{(-3)^2 + (-1)^2}$$

$$= \sqrt{9 + 1}$$

$$= \sqrt{10}$$

$$d = 3.162$$

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 by 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}

X1	X2	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 betwn each points from set and new record, and assigning them their ranks accordingly.

d1 = sqrt(9-5)^2 + (9-9)^2 = sqrt16 = 4

d2 = sqrt(9-5)^2 + (6-9)^2 = sqrt25 = 5

d3 = sqrt(5-5)^2 + (6-9)^2 = sqrt9 = 3

d4 = sqrt(5-5)^2 + (7-9)^2 = sqrt4 = 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, X1 =5; X2 =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:

$P(Yes) = 4/7; P(No)=3/7$

we calculate entropy by putting probability values in formula :

entropy = $-Plog_2P - Nlog_2N$

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

$= (-0.571)x(-0.808) - 0.428(-1.224)$

$= 0.461 - (-0.523)$

Entropy = 0.984