

20IT6302:- Machine Learning.Home Assignment - 3

ID	Speed	Agility	Draft
1	2.50	6.00	yes
2	3.75	8.00	no
3	2.25	5.50	yes
4	3.25	8.25	no
5	2.75	7.50	no
6	4.50	5.00	yes
7	3.50	5.25	yes
8	3.00	3.25	no
9	4.00	4.00	no
10	4.25	3.75	yes.

The speed and agility of the 10 college athletes along with whether they have drafted or not is given in the above table. Predict whether an athlete with Speed = 6.75 and agility = 3.00 will be drafted or not.

A) given

Speed = 6.75 and agility = 3.00

Will be drafted or not.

ID	speed	agility	Distance	Draft.
1	2.50	6.00	5.20	yes
2	3.75	8.00	5.83	no
3	2.25	5.50	5.14	yes
4	3.25	8.25	6.30	no
5	2.75	7.50	6.02	no
6	4.50	5.00	3.01	yes
7	3.50	5.25	3.95	yes
8	3.00	3.25	3.75	no
9	4.00	4.00	2.92	no
10	4.25	3.75	2.61	yes

6.75

3.00

Let $K = 3$.

So the 3 nearest neighbors are

~~10~~ ~~4.25~~ ~~3.75~~ 10 \rightarrow 2.61 \rightarrow yes
 9 \rightarrow 2.92 \rightarrow no
 6 \rightarrow 3.01 \rightarrow yes.

Majority Voting Rule:-

\therefore An athlete With speed = 6.75 and agility = 3.00
 will be a drafter.

6) cluster the following ten points (with (x, y) representing locations) into two clusters using k-means clustering.

$A_1(2, 10)$, $A_2(2, 6)$, $A_3(11, 11)$, $A_4(6, 9)$, $A_5(6, 4)$, $A_6(1, 2)$,
 $A_7(5, 10)$, $A_8(4, 9)$, $A_9(10, 12)$, $A_{10}(7, 5)$.

A)

16+9

Let $A_2(2, 6)$, $A_5(6, 4)$ are cluster centers.

data	x	y	Distance b/w centroid (2, 6) & point in cluster 1	Compare	Distance b/w centroid & point in 2 (6, 4) cluster 2	cluster
A_1	2	10	4	<	7.2 7.2	1
A_2	2	6	0	<	4.4	1
A_3	11	11	10.2	>	8.6	2
A_4	6	9	5	=	5	2
A_5	6	4	4.4	>	0	2
A_6	1	2	4.1	<	5.3	1
A_7	5	10	5	<	6.08	1
A_8	4	9	3.6	<	5.3	1
A_9	10	12	10	>	8.9	2
A_{10}	7	5	5.09	>	1.4	2

$$\text{New centroid (cluster 1)} = \left(\frac{2+2+1+5+4}{5}, \frac{10+6+2+10+9}{5} \right)$$

$$= (2.8, 7.4)$$

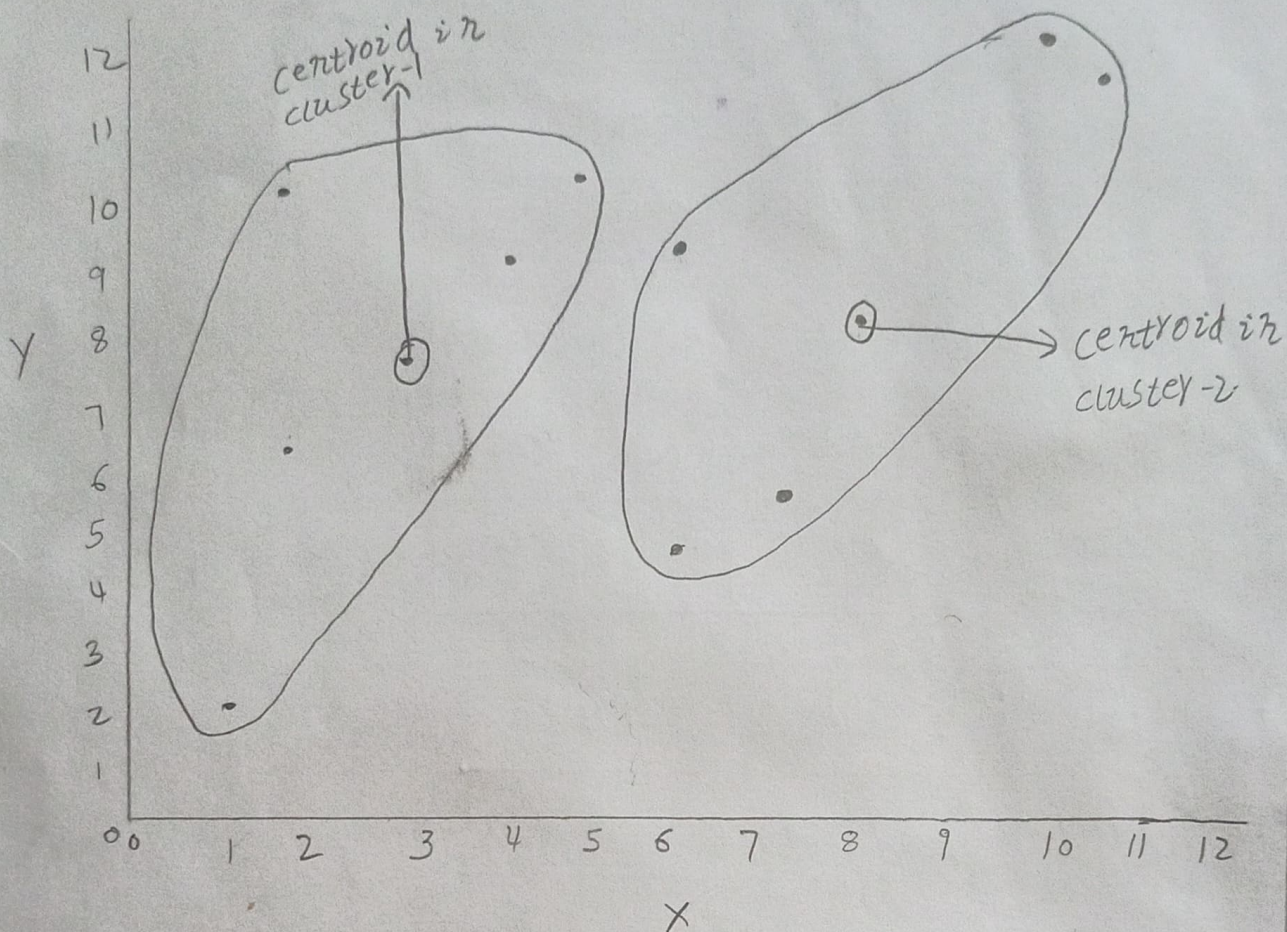
$$\text{new centroid (cluster 2)} = \left(\frac{11+6+6+10+7}{5}, \frac{11+9+4+10+5}{5} \right)$$

$$= (8, 7.8)$$

Date	x	y	Distance b/w centroid (2.8, 7.4) & point in cluster-1	compare	Distance b/w centroid (8, 7.8) & point in cluster-2	cluster
A1	2	10	2.7	<	6.3	1
A2	2	6	1.6	<	6.2	1
A3	11	11	8.9	>	4.3	2
A4	6	9	3.5	>	2.3	2
A5	6	11	4.6	>	4.29	2
A6	1	2	5.6	<	9.09	1
A7	5	10	3.4	<	3.7	1
A8	4	9	2	<	4.1	1
A9	10	12	6.5	>	4.6	2
A10	7	5	4.83	>	2.9	2

$$\text{new centroid (cluster 1)} = \left(\frac{2+2+1+5+4}{5}, \frac{10+6+2+10+9}{5} \right) = (2.8, 7.4)$$

$$\text{new centroid (cluster 2)} = \left(\frac{11+6+6+10+7}{5}, \frac{11+9+4+10+5}{5} \right) = (8, 7.8)$$



12)

	P1	P2	P3	P4	P5	P6
P1	0.0000	0.2357	0.2218	0.3688	0.3421	0.2347
P2	0.2357	0.0000	0.1483	0.2042	0.1388	0.2540
P3	0.2218	0.1483	0.0000	0.1513	0.2843	0.1100
P4	0.3688	0.2042	0.1513	0.0000	0.2932	0.2932
P5	0.3421	0.1388	0.2843	0.2932	0.0000	0.3921
P6	0.2347	0.2540	0.1100	0.2216	0.3921	0.0000

Draw the dendrogram by performing the agglomerative clustering and find the best number of clusters for the 6 objects with the above distance matrix.

A)

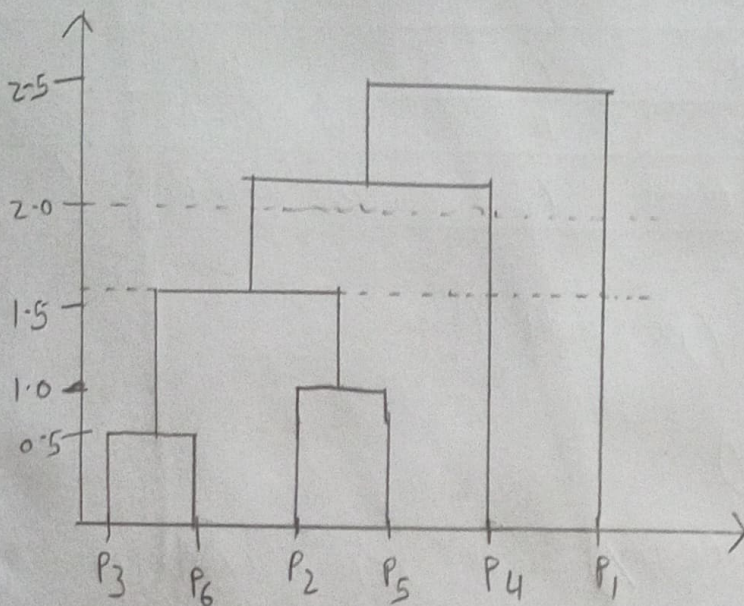
	P1	P2	P3, P6	P4	P5
P1	0.0000	0.2357	0.2218	0.3688	0.3421 0.2357
P2	0.2357	0.0000	0.1483	0.2042	0.1388
P3, P6	0.2218 0.2357	0.1483	0.0000	0.1513	0.2843
P4	0.3688	0.2042	0.1513	0.0000	0.2932
P5	0.2357 0.3421	0.1388	0.2843	0.2932	0.0000

	P1	P2, P5	P3, P6	P4
P1	0.0000	0.2357	0.2218	0.3688
P2, P5	0.2357	0.0000	0.1483	0.2042
P3, P6	0.2218	0.1483	0.0000	0.1513
P4	0.3688	0.2042	0.1513	0.0000

	P1	(P2, P5), (P3, P6)	P4
P1	0.0000	0.2218	0.3688
(P2, P5), (P3, P6)	0.2218	0.0000	0.1513
P4	0.3688	0.1513	0.0000

	p_1	$((p_2, p_5), (p_3, p_6)), p_4$
p_1	0.0000	0.2218
$((p_2, p_5), (p_3, p_6)), p_4$	0.2218	0.0000

$$(((p_2, p_5), (p_3, p_6)), p_4), p_1$$



The best number of clusters for the 6 objects with the above distance matrix are three.

7) use Naive Bayesian classification to predict label of the 8th tuple.

Example	colour	Toughness	Fungus	Appearance	Poisonous
1	green	Hard	N	Smooth	N
2	green	Hard	Y	Smooth	N
3	brown	Soft	N	Wrinkled	N
4	orange	Hard	N	wrinkled	Y
5	green	Soft	Y	smooth	Y
6	green	Hard	Y	wrinkled	Y
7	orange	Hard	N	wrinkled	Y
8	green	soft	Y	wrinkled	?

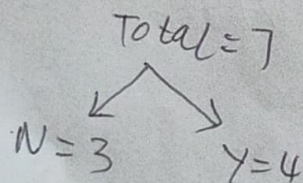
8) $X = \{ \text{greeny, soft, Y, wrinkled} \}$.

$$P(Y/X) = P(Y) * P(X/Y)$$

$$= P(Y) * P(\text{green}/Y) * P(\text{soft}/Y) * P(Y/Y) * P(\text{wrinkled}/Y)$$

$$P(N/X) = P(N) * P(X/N)$$

$$= P(N) * P(\text{green}/N) * P(\text{soft}/N) * P(Y/N) * P(\text{wrinkled}/N)$$



$$P(Y/X) = \frac{4}{7} * \frac{2}{4} * \frac{1}{4} * \frac{2}{4} * \frac{3}{4} = 0.026$$

$$P(N/X) = \frac{3}{7} * \frac{2}{3} * \frac{1}{3} * \frac{1}{3} * \frac{1}{3} = 0.010$$

According to majority class rules - $\max(0.026, 0.010) = 0.026$

The answer is Y.