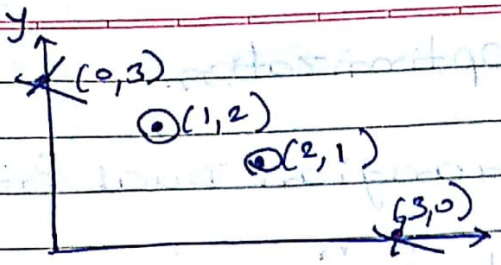
Date : / /



which work good
1) $x+y$
2) $x*y$
3) x^2

	0,3	1,2	2,1	3,0
$x+y$	3	3	3	3
$x*y$	0	2	2	0 ← ✓
x^2	0	1	4	9

* K - mean - Clustering

1) $K = \{2, 3, 4, 10, 11, 12, 20, 25, 30\}$

$K = 2$

$m_1 = 4$ $\quad m_2 = 12$

2)

$K_1 = \{2, 3, 4,\}$ $\quad K_2 = \{10, 11, 12, 20, 25, 30\}$

$m_1 = 3$ (mean 2,3,4) $\quad m_2 = \frac{108}{6} = 18$

3)

$K_1 = \{2, 3, 4, 10\}$ $\quad K_2 = \{11, 12, 20, 25, 30\}$

$m_1 = 4.75 \simeq 5$ $\quad m_2 = 19.6 \simeq 20$

4)

$K_1 = \{2, 3, 4, 10, 11, 12\}$ $\quad K_2 = \{20, 25, 30\}$

$m_1 = 7$ $\quad m_2 = 25$

$K_1 = \{2, 3, 4, 10, 11, 12\}$ $\quad K_2 = \{20, 25, 30\}$

$m_1 = 7$ $\quad m_2 = 25$

Thus we are getting same mean where we have stop.

$\therefore K_1 = \{2, 3, 4, 10, 11, 12\}$ $\quad K_2 = \{20, 25, 30\}$

Page :
Date : / /

Step 1 :- Take mean value (random)
Step 2 :- Find ∴ Nearest number of mean & put in cluster.
Step 3 :- Repeat one & two untill we get same mean.

- K-mean Clustering.

→ The K-means algorithm is an algorithm to cluster n objects based on attributes into K, where K < n

* Unsupervised learning
- Measuring similarity bet^n observation
- Euclidean Distance :- Most common method to measure distance bet^n observation, when observations include continuous variables as ED Euclidean distance

→ The Euclidean distance bet^n observation u & v

$$d_{u,v} = \sqrt{(u_1-v_1)^2+(u_2-v_2)^2+\cdots+(u_q-v_q)^2}$$

→ The Rectilinear distance.

$$d_{u,v} = |u_1-v_1|+|u_2-v_2|+\cdots|u_q-v_q|$$

Ex 1) ⇒ Cluster the following eight points (with (x,y) representing locations) into three clusters
A1 (2,10), A2 (2,5) A3 (8,4), A4 (5,8) A5 (7,5) A6 (6,4), A7 (1,2), A8 (4,9)

→ As K = 3 Need identify 3 centers
→ Select 3 point randomly

∴ Intial cluster center are : A1(2,10), A4 (5,8) & A7 (1,2)

Page :
Date : / /

* The distance fun^n bet^n two points $a = (x_1, y_1)$ & $b = (x_2, y_2)$ is defined as $P(a,b) = |x_2 - x_1| + |y_2 - y_1|$
* Use K-means algorithm to find the three cluster centers after the second iteration

Iteration 1

	Point	x_1, y_1 (2,10) Dist mean1	(5,8) Dist mean2	(1,2) Dist mean3	Cluster
A1	(2,10)	0	5	9	1
A2	(2,5)	5	6	4	3
A3	(8,4)	12	7	9	2
A4	(5,8)	5	0	10	2
A5	(7,5)	10	5	9	2
A6	(6,4)	10	5	7	2
A7	(1,2)	9	10	0	3
A8	(4,9)	3	2	10	2

Cluster 1	Cluster 2	Cluster 3
(2,10)	(8,4)	(2,5)
	(5,8)	(1,2)
	(7,5)	
	(6,4)	
	(4,9)	

Iteration 2 1) for cluster 1, we only have one point A1 (2,10) which was the old mean, so the cluster remain the same.

2) for cluster 2, we have $(8+5+7+6+4)/5$, $((4+8+5+4+9)/5) = 6,6$

3) for cluster 3 $(2+1)/2$, $(5+2)/2 = (1.5, 3.5)$

Page :
Date : / /

2

Iteration Point		(2,10)	(6,6)	(1.5,3.5)	Cluster
A1	(2,10)	0	8	7	1
A2	(2,5)	5	5	2	3
A3	(8,4)	12	4	7	2
A4	(5,8)	2	3	8	2
A5	(7,5)	10	2	7	2
A6	(6,4)	10	2	5	2
A7	(1,2)	9	9	2	3
A8	(4,9)	3	5	8	1

Cluster 1	Cluster 2	Cluster
(2,10)	(8,4)	(2,5)
(4,9)	(5,8)	(1,2)
	(7,5)	
	(6,4)	

1) For Cluster 1 $(2+4)/2, (10+9)/2 = (3, 9.5)$
2) For cluster 2 $(8+5+7+6)/4, (4+8+5+4)/4 = (6.5, 5.25)$
3) For Cluster 3 $(2+1)/2, (5+2)/2 = (1.5, 3.5)$

Iteration 3 Point		(3,9.5) Dist mean1	(6.5,5.25) Dist mean2	(1.5,3.5) Dist mean3	Cluster
A1	(2,10)	1.5	9.25	7	1
A2	(2,5)	5.5	4.75	2	3
A3	(8,4)	10.5	2.75	7	2
A4	(5,8)	3.5	4.25	8	1
A5	(7,5)	8.5	0.75	7	2
A6	(6,4)	8.5	1.75	5	2
A7	(1,2)	9.5	8.75	2	3
A8	(4,9)	1.5	6.25	8	1

Cluster 1	Cluster 2	Cluster 3
(2,10)	(8,4)	(2,5)
(5,8)	(7,5)	(1,2)
(4,9)	(6,4)	
mean (3.67,9)	(7,4.3)	(1.5,3.5)

Page :
Date :

(3.67, 9) (7, 4.3) (1.5, 3.5) ← (Centroid)

	Point	Dist mean1	Dist mean2	Dist mean3	Cluster
A1	(2, 10)	2.67	10.7	7	1
A2	(2, 5)	5.67	5.7	2	3
A3	(8, 4)	9.33	1.3	7	2
A4	(5, 8)	2.33	5.7	8	1
A5	(7, 5)	7.33	0.7	7	2
A6	(6, 4)	7.33	1.3	5	2
A7	(1, 2)	9.67	8.3	2	3
A8	(4, 9)	0.33	7.7	8	1

Same Allocation