

Association Rule

Transaction 1	   
Transaction 2	  
Transaction 3	 
Transaction 4	 
Transaction 5	   
Transaction 6	  
Transaction 7	 
Transaction 8	 

Support {} = 4

RelativeSupport {} = 4/8

$$\text{Confidence} \{ \text{apple} \rightarrow \text{beer} \} = \frac{\text{Support} \{ \text{apple}, \text{beer} \}}{\text{Support} \{ \text{apple} \}} = \frac{3}{4}$$

THUẬT TOÁN K-MEANS

- Bước 1: Đầu tiên chọn random k điểm trong tập n điểm kia làm trọng tâm của k nhóm
- Bước 2: Tính khoảng cách từ mỗi điểm đến trọng tâm (c_x , c_y) của từng nhóm Sử dụng công thức tính khoảng cách:

$$d(a, b)^2 = (a_x - b_x)^2 + (a_y - b_y)^2$$

- Bước 3: Đối với mỗi điểm A, cho vào nhóm có khoảng cách từ trọng tâm nhóm đó đến điểm A là gần nhất.
- Bước 4: Tính lại tọa độ trọng tâm cho mỗi nhóm**

$$c_x = \frac{\text{SUM}(a1.x + a2.x + \dots + a_m.x)}{m}$$

$$c_y = \frac{\text{SUM}(a1.y + a2.y + \dots + a_m.y)}{m}$$

m: số phần tử của nhóm

- Bước 5: Nếu (c_x , c_y) không đổi → đó là giá trị trọng tâm cần tìm. Else làm lại từ B2.

Ứng dụng : Cần chia các điểm dữ liệu sau thành 2 nhóm ⇒ Tìm trọng điểm trọng tâm và phân các điểm trên vào 2 nhóm ?

x	y
127	993
138	741
846	391
876	879
980	580
369	725

Cách giải :

Yêu cầu: Chia 6 điểm trên làm 2 nhóm, sao cho khoảng cách từ mỗi điểm đến trọng tâm của mỗi nhóm là gần nhất.

Yêu cầu: Chia 6 điểm trên làm 2 nhóm, sao cho khoảng cách từ mỗi điểm đến trọng tâm của mỗi nhóm là gần nhất.

x	y	d_c1_lo	d_c2_lo	group	d_c1_l1	d_c2_l1	group_new	d_c1_l2	d_c2_l2	Group_new_2
127	993	937.74	947.72	1	371.12	843.23	1	192.8	860.3	1
138	741	789.79	857.25	1	233.74	790.08	1	107.5	772.7	1
846	391	0.00	231.68	1	574.40	348.29	2	765.9	232.2	2
876	879	488.92	316.57	2	532.69	158.29	2	667.3	263.5	2
980	580	231.68	0.00	2	624.22	158.29	2	805.2	87.4	2
369	725	582.31	627.97	1	12.54	559.02	1	183.9	542.6	1

Ồ đây, có $n = 6$, $k = 2$.

GIẢI:

B1: Chọn random 2 điểm trong tập những điểm đã cho làm trọng tâm:

centroid_point_1 = {x: 846, y: 391}
centroid_point_2 = {x: 980, y: 580}

Dừng thuật toán vì tọa độ trọng tâm không thay đổi

Group 1		Group 2	
x	y	x	y
127	993	980	580
138	741	876	879
846	391	928	729.5
369	725		
370	712.5		

Group_1_moi		Group_2_moi	
x	y	x	y
127	993	846	391
138	741	876	879
369	725	980	580
211.333	819.667	900.667	616.667