CS5525 - April 24th

Exam II next class

KNN

Deep learning

Associan rule mining + clusting (k means)

↓

Apriori

Suppose this is our dataset of any supermarket, where user id and items are as shown below. Using apriori algorithm find the strongest association rule between items.

User ID	Items
001	1, 3, 4
002	2, 3, 5
003	1, 2, 3, 5
004	2, 5

step 1 : set min support and confidance

min support = 50%

min confidance = 70%

step 2

item	support
1	2/4 = 50%
2	3/4 = 75%
3	75%
4	~~25%~~
5	75%

considering min suppot

⇓

keep {1,2,3,5}

Step3 :

item	support	
1,2	~~1/4~~	25%
1,3	50%	
1,5	~~25%~~	
2,3	50%	
2,5	75%	
3,5	50%	

Min suppot is 50%

so below 50%

support to be removed

remove {1,2}, {1,5}

Step4 {1,3} {2,3} {2,5} {3,5}

---

item: {1,2,3,5}

item	support
{1,2,3}	~~25%~~
{1,3,5}	~~25%~~
{2,3,5}	50%
{1,2,5}	~~25%~~

Min sup. 50%

{2,3,5}

Steps Form association rule {2,3,5}

	rules	support	confidence
1	$\overbrace{\{2,3\}}^{A} \longrightarrow \overbrace{\{5\}}^{B}$		$\frac{2/4}{2/4} = \boxed{100\%}$
2	$\{2,5\} \longrightarrow \{3\}$		$\frac{2/4}{3/4} = \cancel{66.6\%}$
3	$\{3,5\} \longrightarrow \{2\}$		$\frac{2/4}{2/4} = \boxed{100\%}$
4	$\{2\} \longrightarrow \{3,5\}$		$\frac{2/4}{3/4} = \cancel{66\%}$
5	$\{3\} \longrightarrow \{2,5\}$		$\frac{2/4}{3/4} = \cancel{66\%}$
6	$\{5\} \longrightarrow \{2,3\}$		$\frac{2/4}{3/4} = \cancel{66\%}$

$$\text{confidence}\ (A \rightarrow B) = \frac{\text{sup}(A, B)}{\text{sup}(A)}$$

Min conf = 70%

User ID	Items
001	1, 3, 4
002	2, 3, 5
003	1, 2, 3, 5
004	2, 5

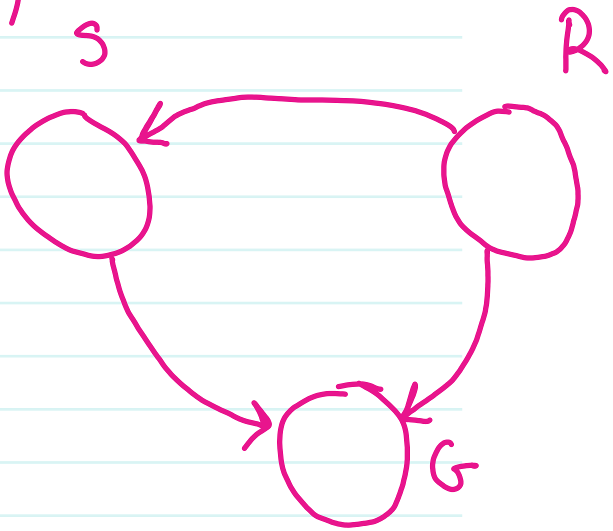
Two finds - strong association rules. by Apriori Algorithm

$$\boxed{\begin{array}{l} \{2,3\} \longrightarrow \{5\} \\ \{3,5\} \longrightarrow \{2\} \end{array}}$$

$$P(R=T \mid G=T) = \frac{P(G,R)}{P(G)} \leftarrow$$

$$P(R=F \mid G=T) = \frac{P(\overline{G},R)}{P(G)}$$

$$\underline{P(G,S,R)} = P(G \mid S,R) * P(S \mid R) * P(R)$$



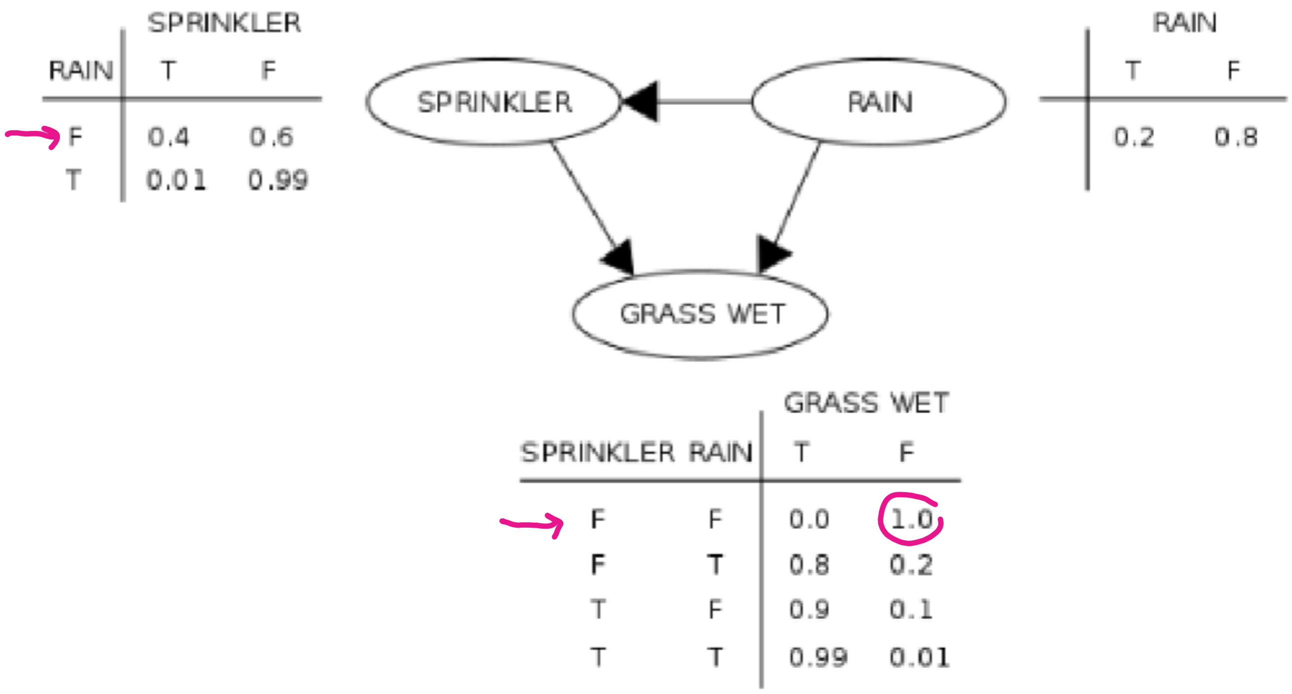
	$\overline{R}$	$\boxed{R}$
$\overline{G}$ $\overline{S}$	a_1	b_1
$\overline{G}$, S	a_2	b_2
G, $\overline{S}$	a_3	b_3
G, S	a_4	b_4

$$P(G,R) = \underline{b_3 + b_4}$$

$$b_3 + b_4 + a_3 + a_4$$

$$P(\bar{G}, \bar{S}, \bar{R}) = P(\bar{G} | \bar{S}, \bar{R}) * P(\bar{S} | \bar{R}) * P(\bar{R})$$

$$= 1 * 0.6 * .8 = .48$$



	SPRINKLER	
RAIN	T	F
F	0.4	0.6
T	0.01	0.99

RAIN	
T	F
0.2	0.8

		GRASS WET	
SPRINKLER	RAIN	T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

Figure 1

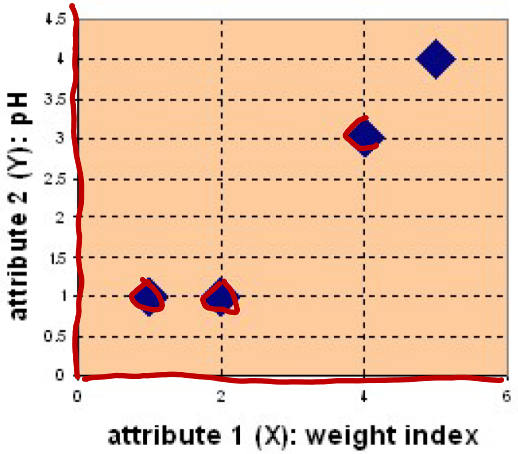
4 - types of Medicines

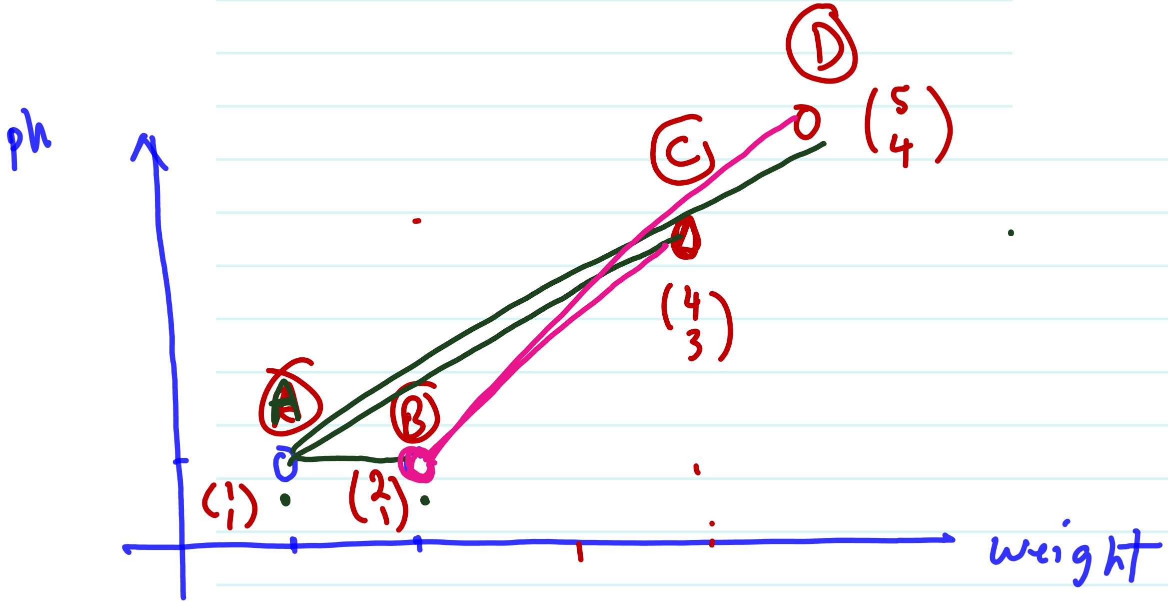
two features

our goal is to group these objects

into k = 2 groups

Object	weight Index	pH
Medicine A	1	1
Medicine B	2	1
Medicine C	4	3
Medicine D	5	4





Step 1) : Initial value of centroids $C_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $C_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

$$\text{Distance} = D^0 = \begin{array}{cccc} A & B & C & D \\ \left[\begin{array}{cccc} 0 & 1 & \sqrt{(4-1)^2+(3-1)^2} & \sqrt{(5-1)^2+(4-1)^2} \\ 1 & 0 & \sqrt{(4-2)^2+(3-1)^2} & \sqrt{(5-2)^2+(4-1)^2} \end{array}\right] \end{array}$$

$$D^0 = \begin{bmatrix} 0 & 1 & \underline{3.6} & 5 \\ 1 & 0 & \underline{2.8} & \underline{4.2} \end{bmatrix} \begin{array}{l} \text{group 1 : } c_1 \\ \text{group 2 : } c_2 \end{array}$$

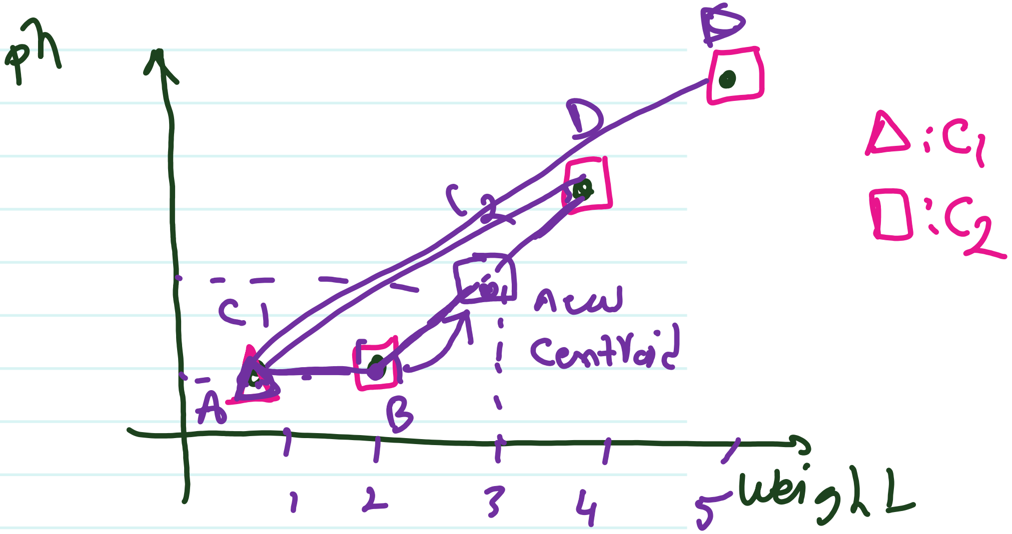
object clustring : we assign each object based on the minimum.

$$G^0 = \begin{array}{cccc} A & B & C & D \\ \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{array}\right] \end{array} \begin{array}{l} \text{group } C_1 \\ \text{group } C_2 \end{array}$$

1 means the object is associated to that cluster.

Step 4) iterations

$A = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad B = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad C = \begin{bmatrix} 4 \\ 3 \end{bmatrix} \quad D = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$



$C_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$$c_2 = \begin{bmatrix} \frac{2+4+5}{3} \\ \frac{1+3+4}{3} \end{bmatrix} = \begin{bmatrix} 11/3 \\ 8/3 \end{bmatrix} = \begin{bmatrix} 3.6 \\ 2.6 \end{bmatrix}$$

$\boxed{\text{step 5}}$: object_centroid distance:

$$D^1 = \begin{array}{c} \\ \\ \end{array}\begin{array}{cccc|c} A & B & C & D & \\ 0 & 1 & 3.6 & 5 & c_1 \\ \sqrt{(\frac{11}{3}-1)^2+(\frac{8}{3}-1)^2} & \sqrt{(\frac{11}{3}-2)^2+(\frac{8}{3}-1)^2} & \sqrt{(\frac{11}{3}-4)^2+(\frac{8}{3}-3)^2} & \sqrt{(\frac{11}{3}-5)^2+(\frac{8}{3}-4)^2} & c_2 \end{array}$$

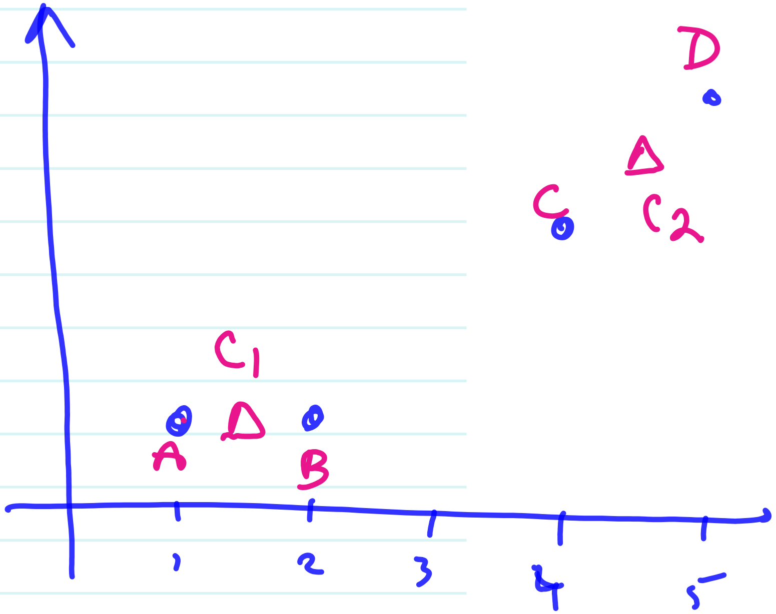
$$D^1 = \begin{array}{c|cccc|c} & A & B & C & D & \\ & 0 & 1 & 3.6 & 5 & C1 \\ & 3.14 & 2.35 & 0.47 & 1.88 & C2 \end{array}$$

$$\boxed{G^1} = \begin{array}{cccc} A & B & C & D \\ 1 & 1 & 0 & 0 \\ 0 & 0 & \underbrace{1 \quad 1}_{} & \end{array}$$

7- Find centroid

$$C_1 = \begin{bmatrix} \frac{1+2}{2} \\ \frac{1+1}{2} \end{bmatrix} = \begin{bmatrix} 1.5 \\ 1 \end{bmatrix}$$

$$C_2 = \begin{bmatrix} \frac{4+5}{2} \\ \frac{3+4}{2} \end{bmatrix} = \begin{bmatrix} 4.5 \\ 3.5 \end{bmatrix}$$



8- Iteration - object centroid distance

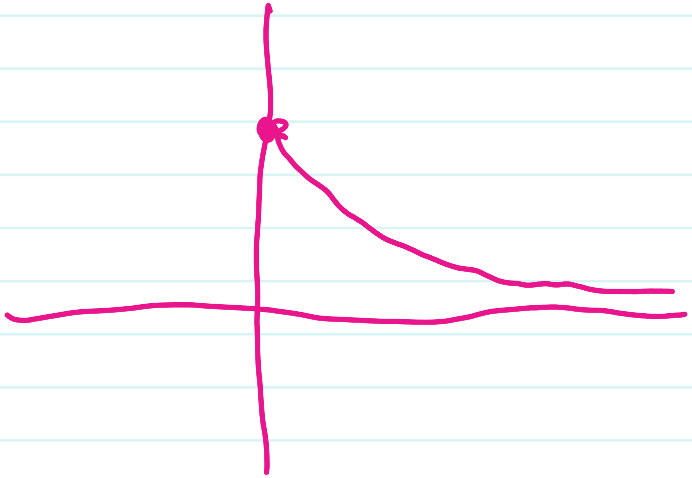
$$D^2 = \begin{array}{cccc} A & & & \\ \sqrt{(1.5-1)^2+(1-1)^2} & \sqrt{(2-1.5)^2+(1-1)^2} & \sqrt{(4-1.5)^2+(1-3)^2} & \sqrt{(5-1.5)^2+(4-1)^2} \\ \sqrt{(4.5-1)^2+(3.5-1)^2} & \sqrt{(2-4.5)^2+(3.5-1)^2} & \sqrt{(4-4.5)^2+(3.5-3)^2} & \sqrt{(5-4.5)^2+(3.5-4)^2} \end{array}$$

$$D^2 = \begin{array}{cc} \begin{bmatrix} 0.5 & .5 & 3.2 & 4.6 \\ 4.3 & 2.9 & 0.7 & 0.7 \end{bmatrix} & \begin{array}{c} c_1 \\ c_2 \end{array} \end{array}$$

$$\boxed{G^2} = \begin{array}{cccc} A & B & C & D \\ \hline 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{array}$$

$$G^2 = G^1$$

$$\frac{1}{1+\lambda}$$



$\lambda = 0$

$x_1 \quad x_2 \quad \dots \quad x_d$

$x_1 \sim x_2 \quad \dots \quad x_d \qquad VIF_1 = \dfrac{1}{1-R_1^2}$

$x_2 \sim x_1, x_3 \,.. \quad x_d, \qquad VIF_2 = \dfrac{1}{1-R_2^2}$

$\vdots$

$x_3 \sim x_1, x_2, x_4 \,.. \quad x_d, \qquad . \qquad \dfrac{1}{1-R_3^2}$

$X_d \sim X_1, X_2 \ldots X_{d-1}$ $\qquad \simeq \dfrac{1}{1-R_d^2}$

VIF