ADB Sheet 4 Sol.

28.14.

→ Apply Apriori algo to the following

Tid	items
101	M, B, E
102	M, J
103	J, Bu
104	M, B, E
105	C, E
106	C
107	C, J
108	M, B, K, E
109	K, Bu
110	M, B

- Set of items is {milk (M), bread (B), Cookies (K), eggs (E), butter (Bu), Coffee (C), Juice (J)}
- Use 0.2 for minsup

→ minsup Count = 0.2 x 10 = 2 transactions

Sol.

C_1 =

itemset	Supcount	transactions
M	5	1,2,4,8,10
B	4	1,4,8,10
K	2	8,9
E	4	1,4,5,8
Bu	2	3,9
C	3	5,6,7
J	3	2,3,7

- To quick check Σ SupCount = Σ items

- All of them have SupCount ≥ 2 → $L_1 = C_1$

C_2 =

Itemset	Support
M, B	4
M, K	1
M, E	3
M, Bu	0
M, C	0
M, J	1
B, K	1
B, E	3
B, Bu	0
B, C	0
B, J	0
K, E	1
K, Bu	1
K, C	0
K, J	0
E, Bu	0
E, C	1
E, J	0
Bu, C	0
Bu, J	1
C, J	1

- Has "L,1" C_2 rows as expected (21)

L_2 =

Itemset	Support
M, B	4
M, E	3
B, E	3

- Nothing else qualifies

$$C_3 = \begin{pmatrix} \overset{L_2}{M,B} \\ M,E \\ B,E \end{pmatrix} \bowtie \begin{pmatrix} \overset{L_2}{M,B} \\ M,E \\ B,E \end{pmatrix} = $$

itemset	supCount
M,B,E	3

Join any two that differ in one item
• Keep unique

• No Pruning was needed (all subsets of M,B,E $^{(2)}$ are in L_2)

$L_3 = C_3$ (all items beyond minSup)

$C_4 = \emptyset$

• $L_3 \bowtie L_3$ is empty (no elements diff. by one item)
→ algorithm terminates

• Now $\overbrace{\text{qualifying}}^{\text{large}}$ itemsets are all those from each L_i:

$\underbrace{M, B, K, E, Bu, C, J}_{\text{Can't be used to infer associations}}$, $\underbrace{\{M,B\}, \{M,E\}, \{B,E\}}_{\text{each can give us up to two associations } (2^2-2)\ X\to Y,\ Y\to X}$, $\underbrace{\{M,B,E\}}_{\text{up to 6 assoc. } (2^3-2)}$

28.15. Find two rules with confidence 0.7 or more for an itemset of 3 items.

→ let's consider all possible rules from {M,B,E}

Subsets	{M,B}	{M,E}	{B,E}	{B}	{M}	{E}
rule	E→M,B	B→M,E	M→B,E	M,E→B	B,E→M	M,B→E

• Start with 1-item consequents (right to left) all supps from apriori table

$M, B \rightarrow E$

- has confidence $= \dfrac{Supp(E,M,B)}{Supp(M,B)} = \dfrac{3}{4} = 0.75$

• Rule accepted

$B, E \rightarrow M$

- Conf $= \dfrac{Supp(E,M,B)}{Supp(B,E)} = \dfrac{3}{3} = 1.0$

• Rule accepted

$M, E \rightarrow B$

- Conf $= \dfrac{3}{Supp(M,E)} = 1.0$

• Rule accepted

$M \rightarrow B, E$

Conf $= \dfrac{3}{Supp(M)} = \dfrac{3}{5} = 0.6$

• Rule failed
• any rule with consq. B, E for this set also will fail
→ There's none :)

$B \rightarrow M, E$

Conf $= \dfrac{3}{Supp(B)} = \dfrac{3}{4} = 0.75$

• Rule accepted

$E \rightarrow M, B$

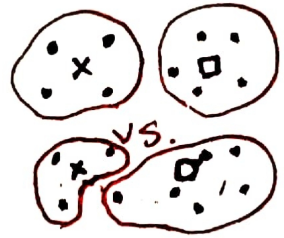
Conf $= \dfrac{3}{Supp(E)} = \dfrac{3}{4} = 0.75$

• Rule accepted

K-Means

Start with initial cluster centers
Repeat until no change:
Assign each example closest center
Recalculate each of the centers as averages

- Given two Clusterings of dataset



→ Better one is that with lowest total Sum of Squared Variations

$J_{Tot} = 0$

For each Cluster C

$$J_c = \sum_{x \text{ in } C} d^2_{x \to c}$$

$J_{Tot} \mathrel{+}= J_c$

$d_{x \to c}$ is the euclid. distance between x and the Center of Cluster C.

only for Points belonging in C

* This is all you need to know for the next two Problems on K-means

Thanks <3