

## Apriori Association Rule Generation Example

Minimum support count is set to 2 (or  $2/7=28.5\%$ ), minimum confidence is set to 60%

Data:

T0: I1, I2, I3, I5

T1: I2, I4

T2: I1, I2, I4, I5

T3: I2, I6

T4: I1, I2

T5: I1, I2, I3, I5

T6: I1, I2, I3

### Step1: Frequent itemset generation

Candidate 1-Itemsets:

Item	Support count
I1	5
I2	7
I3	3
I4	2
I5	3
<del>I6</del>	<del>1</del>

Frequent 1-itemsets, L1:

Item
I1
I2
I3
I4
I5

L1 self join L1 → candidate 2-itemsets:

Item	Support count
I1, I2	5
I1, I3	3
<del>I1, I4</del>	<del>1</del>
I1, I5	3
I2, I3	3
I2, I4	2
I2, I5	3
<del>I3, I4</del>	0
I3, I5	2

I4, I5	1
--------	---

Frequent 2-itemsets L2:

Item
I1, I2
I1, I3
I1, I5
I2, I3
I2, I4
I2, I5
I3, I5

L2 self join L2 → candidate 3-itemsets

Item	Support count
I1, I2, I3	3
I1, I2, I5	3
I1, I3, I5	2
I2, I3, I4	pruned
I2, I3, I5	2
I2, I4, I5	pruned

Frequent 3-itemsets L3:

Item
I1, I2, I3
I1, I2, I5
I1, I3, I5
I2, I3, I5

L3 self join L3 → candidate 4-itemsets:

Item	Support count
I1, I2, I3, I5	2

Frequent 4-itemset:

Item
I1, I2, I3, I5

L1, L2, L3, and L4 are frequent 1, 2, 3, and 4 itemsets

### Step two: Rule Generation

For each of the frequent itemset generated from Step 1, the following approach will be applied to generate all the strong association rules. In this example, I am showing the derivation of all the rules generated from the frequent 4-itemset (I1, I2, I3, I5)

1) Singleton item on the consequent side of the rule

I2, I3, I5  $\rightarrow$  I1 (confidence  $2/2=1 > 60\%$ )

I1, I3, I5  $\rightarrow$  I2 (confidence  $2/2=1 > 60\%$ )

I1, I2, I5  $\rightarrow$  I3 (confidence  $2/3=67\% > 60\%$ )

I1, I2, I3  $\rightarrow$  I5 (confidence  $2/3=67\% > 60\%$ )

All I1, I2, I3, I5 are kept as L1

2) L1 self join L2  $\rightarrow \{(I1, I2), (I1, I3), (I1, I5), (I2, I3), (I2, I5), (I3, I5)\}$

I3, I5  $\rightarrow$  I1, I2 (confidence  $2/2=1 > 60\%$ )

I2, I5  $\rightarrow$  I1, I3 (confidence  $2/3=67\% > 60\%$ )

I2, I3  $\rightarrow$  I1, I5 (confidence  $2/3=67\% > 60\%$ )

I1, I5  $\rightarrow$  I2, I3 (confidence  $2/3=67\% > 60\%$ )

I1, I3  $\rightarrow$  I2, I5. (confidence  $2/3=67\% > 60\%$ )

~~I1, I2  $\rightarrow$  I3, I5. (confidence  $2/5=40\% < 60\%$ )~~

3) L2 sets are  $\{(I1, I2), (I1, I3), (I1, I5), (I2, I3), (I2, I5)\}$

L2 self join L2, then perform pruning

L2 self join L2 results in C3:

I1, I2, I3

I1, I2, I5

I1, I3, I5.  $\leftarrow$  pruned because (I3, I5) is not in L2

I2, I3, I5.  $\leftarrow$  pruned because (I3, I5) is not in L2

Now compute the confidence of these two rules:

I5  $\rightarrow$  I1, I2, I3 (confidence  $2/3=67\% > 60\%$ )

I3  $\rightarrow$  I1, I2, I5 (confidence  $2/3=67\% > 60\%$ )