

25 Jan 2021

# Tut 1 - Apriori Algorithm

IIT2018179 - Mohammed Aakil

Q1) (i) Generating  $C_1$  &  $L_1$  (1-itemset frequent):

Itemset	Sup. Count	Compare with min support count →	Itemset	Sup. Count
$\{l_1\}$	6		$\{l_1\}$	6
$\{l_2\}$	7		$\{l_2\}$	7
$\{l_3\}$	6		$\{l_3\}$	6
$\{l_4\}$	2		$\{l_4\}$	2
$\{l_5\}$	2		$\{l_5\}$	2

$C_1$   $L_1$

(ii) Generating  $C_2$  &  $L_2$  (2-itemset frequent):

Itemset	Sup. Count	Compare with min support count →	Itemset	Sup. Count
$\{l_1, l_2\}$	4		$\{l_1, l_2\}$	4
$\{l_1, l_3\}$	4		$\{l_1, l_3\}$	4
$\{l_1, l_4\}$	1		$\{l_1, l_5\}$	2
$\{l_1, l_5\}$	2		$\{l_2, l_3\}$	4
$\{l_2, l_3\}$	4		$\{l_2, l_4\}$	2
$\{l_2, l_4\}$	2		$\{l_2, l_5\}$	2
$\{l_2, l_5\}$	2			
$\{l_3, l_4\}$	0			
$\{l_3, l_5\}$	1			
$\{l_4, l_5\}$	0			

$C_2$   $L_2$

(iii) Generating  $C_3$  &  $L_3$  (3-itemset frequent):

Itemset	Sup. count
$\{l_1, l_2, l_3\}$	2
$\{l_1, l_2, l_4\}$	1
$\{l_1, l_2, l_5\}$	2
$\{l_1, l_3, l_5\}$	1
$\{l_2, l_3, l_4\}$	0
$\{l_2, l_3, l_5\}$	1
$\{l_2, l_4, l_5\}$	0

Compare  
min support  
count

Itemset	Sup. count
$\{l_1, l_2, l_3\}$	2
$\{l_1, l_2, l_5\}$	2

$L_3$

(iv) Generating  $C_4$  &  $L_4$  (4-itemset frequent):

$$C_4 = \{l_1, l_2, l_3, l_5\} = 1$$

$$L_4 = \emptyset$$

(v) Generating Association Rules using  $L_3$ :

for set  $\{l_1, l_2, l_3\}$

RULES	Confidence
$l_1 \rightarrow l_2 \wedge l_3$	$sc\{l_1, l_2, l_3\} / sc\{l_1\} = 2/6 = 33.3\%$
$l_2 \rightarrow l_1 \wedge l_3$	$sc\{l_1, l_2, l_3\} / sc\{l_2\} = 2/7 \approx 30\%$
$l_3 \rightarrow l_1 \wedge l_2$	$sc\{l_1, l_2, l_3\} / sc\{l_3\} = 2/6 = 33.3\%$
$l_1 \wedge l_2 \rightarrow l_3$	$sc\{l_1, l_2, l_3\} / sc\{l_1, l_2\} = 2/4 = 50\%$
$l_1 \wedge l_3 \rightarrow l_2$	$sc\{l_1, l_2, l_3\} / sc\{l_1, l_3\} = 2/4 = 50\%$
$l_2 \wedge l_3 \rightarrow l_1$	$sc\{l_1, l_2, l_3\} / sc\{l_2, l_3\} = 2/4 = 50\%$

None of the rules have confidence more than 70%.

for set  $\{l_1, l_2, l_5\}$

RULES	Confidence
$l_1 \rightarrow l_2 \wedge l_5$	$sc\{l_1, l_2, l_5\} / sc\{l_1\} = 2/6 = 33.3\%$
$l_2 \rightarrow l_1 \wedge l_5$	$sc\{l_1, l_2, l_5\} / sc\{l_2\} = 2/7 \approx 30\%$
$l_5 \rightarrow l_1 \wedge l_2$	$sc\{l_1, l_2, l_5\} / sc\{l_5\} = 2/2 = 100\%$
$l_1 \wedge l_2 \rightarrow l_5$	$sc\{l_1, l_2, l_5\} / sc\{l_1, l_2\} = 2/4 = 50\%$
$l_1 \wedge l_5 \rightarrow l_2$	$sc\{l_1, l_2, l_5\} / sc\{l_1, l_5\} = 2/2 = 100\%$
$l_2 \wedge l_5 \rightarrow l_1$	$sc\{l_1, l_2, l_5\} / sc\{l_2, l_5\} = 2/2 = 100\%$

3 rules are selected:

$$R_1 = l_5 \rightarrow l_1 \wedge l_2$$

$$R_2 = l_1 \wedge l_5 \rightarrow l_2$$

$$R_3 = l_2 \wedge l_5 \rightarrow l_1$$

Q2)

(i) min-support = 40%

Total TID's = 5

$$\therefore \frac{x}{5} = 40 \Rightarrow \underline{x = 2 = \text{min\_support count}}$$

(ii) Generating  $C_1$  &  $L_1$ :

Itemset	SC		Itemset	SC
{A}	5	Compare min-support count →	{A}	5
{B}	3		{B}	3
{C}	5		{C}	5
{D}	4		{D}	4
{E}	2		{E}	2
$C_1$			$L_1$	

(iii) Generating  $C_2$  &  $L_2$

Itemset	Sup count
{A, B}	3
{A, C}	5
{A, D}	4
{A, E}	2
{B, C}	3
{B, D}	2
{B, E}	1
{C, D}	4
{C, E}	2
{D, E}	2

$C_2$

Compare  
min-support  
count →

Itemset	SC
{A, B}	3
{A, C}	5
{A, D}	4
{A, E}	2
{B, C}	3
{B, D}	2
{C, D}	4
{C, E}	2
{D, E}	2

$L_2$

(iv) Generating  $C_3$  &  $L_3$

Itemset	SC
{A, B, C}	3
{A, B, D}	2
{A, B, E}	1
{A, C, D}	4
{A, C, E}	2
{A, D, E}	2
{B, C, D}	2
{B, C, E}	1
{B, D, E}	1
{C, D, E}	2

$C_3$

Compare  
min-support  
count →

Itemset	SC
{A, B, C}	3
{A, B, D}	2
{A, C, D}	4
{A, C, E}	2
{A, D, E}	2
{B, C, D}	2
{C, D, E}	2

$L_3$

(v) Generating  $C_4 \times L_4$  :

Itemset	SC
$\{A, B, C, D\}$	2
$\{A, B, C, E\}$	1
$\{A, B, D, E\}$	1
$\{A, C, D, E\}$	2
$\{B, C, D, E\}$	1

$C_4$

Compare  
min-support  
count -

Itemset	SC
$\{A, B, C, D\}$	2
$\{A, C, D, E\}$	2

$L_4$

(vi) Generating  $C_5 \times L_5$

$$C_5 = \{A, B, C, D, E\} \times SC = 1$$

$$\therefore L_5 = \underline{\underline{\phi}}$$

(vii) Generating Association Rules using  $L_4$  :

There are 2 sets we will use to generate the rules

$$S1 = \{A, B, C, D\}$$

$$S2 = \{A, C, D, E\}$$


Taking subsets of both sets individually.

for  $S_1 = \{A, B, C, D\}$

RULES	Confidence
$A \rightarrow B \wedge C \wedge D$	$SC\{A, B, C, D\} / SC\{A\} = 2/5 = 40\%$
$B \rightarrow A \wedge C \wedge D$	$SC\{A, B, C, D\} / SC\{B\} = 2/3 = 66.6\%$
$C \rightarrow A \wedge B \wedge D$	$SC\{A, B, C, D\} / SC\{C\} = 2/5 = 40\%$
$D \rightarrow A \wedge B \wedge C$	$SC\{A, B, C, D\} / SC\{D\} = 2/4 = 50\%$
$A \wedge B \rightarrow C \wedge D$	$SC\{A, B, C, D\} / SC\{A, B\} = 2/3 = 66.6\%$
$A \wedge C \rightarrow B \wedge D$	$SC\{A, B, C, D\} / SC\{A, C\} = 2/5 = 40\%$
$A \wedge D \rightarrow B \wedge C$	$SC\{A, B, C, D\} / SC\{A, D\} = 2/4 = 50\%$
$B \wedge C \rightarrow A \wedge D$	$SC\{A, B, C, D\} / SC\{B, C\} = 2/3 = 66.6\%$
$B \wedge D \rightarrow A \wedge C$	$SC\{A, B, C, D\} / SC\{B, D\} = 2/2 = 100\%$
$C \wedge D \rightarrow A \wedge B$	$SC\{A, B, C, D\} / SC\{C, D\} = 2/4 = 50\%$
$A \wedge B \wedge C \rightarrow D$	$SC\{A, B, C, D\} / SC\{A, B, C\} = 2/3 = 66.6\%$
$A \wedge B \wedge D \rightarrow C$	$SC\{A, B, C, D\} / SC\{A, B, D\} = 2/2 = 100\%$
$A \wedge C \wedge D \rightarrow B$	$SC\{A, B, C, D\} / SC\{A, C, D\} = 2/4 = 50\%$
$B \wedge C \wedge D \rightarrow A$	$SC\{A, B, C, D\} / SC\{B, C, D\} = 2/2 = 100\%$

All the rules have more than 40% confidence.

NOTE: We skipped  $A \wedge B \wedge C \wedge D \rightarrow \phi$  and  $\phi \rightarrow A \wedge B \wedge C \wedge D$  as they both have null values.

for  $S_2 = \{A, C, D, E\}$  

RULES	Confidence
$A \rightarrow C \wedge D \wedge E$	$SC\{A, C, D, E\} / SC\{A\} = 2/5 = 40\%$
$C \rightarrow A \wedge D \wedge E$	$SC\{A, C, D, E\} / SC\{C\} = 2/5 = 40\%$
$D \rightarrow A \wedge C \wedge E$	$SC\{A, C, D, E\} / SC\{D\} = 2/4 = 50\%$
$E \rightarrow A \wedge C \wedge D$	$SC\{A, C, D, E\} / SC\{E\} = 2/2 = 100\%$
$A \wedge C \rightarrow D \wedge E$	$SC\{A, C, D, E\} / SC\{A, C\} = 2/5 = 40\%$
$A \wedge D \rightarrow C \wedge E$	$SC\{A, C, D, E\} / SC\{A, D\} = 2/4 = 50\%$
$A \wedge E \rightarrow C \wedge D$	$SC\{A, C, D, E\} / SC\{A, E\} = 2/2 = 100\%$
$C \wedge D \rightarrow A \wedge E$	$SC\{A, C, D, E\} / SC\{C, D\} = 2/4 = 50\%$
$C \wedge E \rightarrow A \wedge D$	$SC\{A, C, D, E\} / SC\{C, E\} = 2/2 = 100\%$
$D \wedge E \rightarrow A \wedge C$	$SC\{A, C, D, E\} / SC\{D, E\} = 2/2 = 100\%$
$A \wedge C \wedge D \rightarrow E$	$SC\{A, C, D, E\} / SC\{A, C, D\} = 2/4 = 50\%$
$A \wedge C \wedge E \rightarrow D$	$SC\{A, C, D, E\} / SC\{A, C, E\} = 2/2 = 100\%$
$A \wedge D \wedge E \rightarrow C$	$SC\{A, C, D, E\} / SC\{A, D, E\} = 2/2 = 100\%$
$C \wedge D \wedge E \rightarrow A$	$SC\{A, C, D, E\} / SC\{C, D, E\} = 2/2 = 100\%$

All the rules have more than 40% confidence.

NOTE: We skipped  $A \wedge C \wedge D \wedge E \rightarrow \phi$  and  $\phi \rightarrow A \wedge C \wedge D \wedge E$  as they both have null values.

Finally, All the rules generated from  $S_1$  &  $S_2$  are valid