

1. Consider the following dataset

A	B	C
1		1
1	1	
		1
1	1	1

- Following the Apriori algorithm, identify the frequent sets, assuming  $\text{Sup}_{\min} = 30\%$ .
- Say which rules will be selected if  $\text{Conf}_{\min} = 100\%$ .

2. Consider the following set of purchases represented in the Table

TID	A	B	C	D
1	1			
2	1		1	
3	1	1	1	1
4				1
5		1	1	1

- Indicate the support and confidence of the rule  $\{D\} \Rightarrow \{B, C\}$ .
- Explain the meaning of the measure of support and confidence in terms of probabilities.
- If we know that the rule  $\{D\} \Rightarrow \{B, C\}$  has sufficient confidence (above a minimum established), what can be said about the rules  $\{B, D\} \Rightarrow \{C\}$  and  $\{C, D\} \Rightarrow \{B\}$ ? Justify.
- Suppose the frequent sets of size 2 are:  $\{A, C\}$ ,  $\{B, C\}$ ,  $\{B, D\}$ ,  $\{C, D\}$ . What the candidate sets of size 3 constructed by the APRIORI algorithm and their respective supports?
- Assume that the minimum support is 0.5. Which frequent sets are found by APRIORI.
- Assume that the minimum confidence is 0.4. What rules would be produced from the sets  $\{A, B\}$  and  $\{B, C\}$ ? Justify.

3. Consider the following dataset

A	x	s	w
B	x	s	r
C	y	s	t
D	x	n	r

- What changes should you make to this dataset so that the Apriori algorithm can be applied? Present the new dataset.
- Identify the frequent sets for a minimum support of 0.5.
- Say which rules will be selected if the minimum confidence is 1.

1. a)

A	B	C
1		1
1	1	
		1
1	1	1

Frequent sets Minimum Support = 0,30

Sets of size 1

Item_set	Support
{ A }	3 / 4 (0,75)
{ B }	1 / 2 (0,50)
{ C }	3 / 4 (0,75)

Sets of size 2

Item_set	Support
{ A, B }	1 / 2 (0,50)
{ A, C }	1 / 2 (0,50)
{ B, C }	1 / 4 (0,25) < 0,30 <b>eliminated !</b>

Sets of size 3

Item_set	Support
{ A, B, C }	1 / 4 (0,25) <b>eliminated !</b>

b)

Minimum Confidence = 1

Rules	Confidence	Support
$A \Rightarrow B$	2 / 3 (0,66)	0,50
$A \Rightarrow C$	2 / 3 (0,66)	0,50
<b><math>B \Rightarrow A</math></b>	<b>1</b>	<b>0,50</b>
$C \Rightarrow A$	2 / 3 (0,66)	0,50

2.

a) Rule  $\{ D \} \Rightarrow \{ B, C \}$

$$\text{Support} = 2/5 = 0,4$$

$$\text{Confidence} = 2/3 = 0,67$$

b)

**Support (  $\{P,Q\} \Rightarrow \{Z\}$  )** is the % of "baskets" in which the joint occurrence of items is observed, that is, it is the probability that the transactions contain  $\{P \cap Q \cap Z\}$ .

$$\text{Support} ( \{P, Q\} \Rightarrow \{Z\} ) = \text{Probability} ( P \cap Q \cap Z )$$

**Confidence (  $\{P,Q\} \Rightarrow \{Z\}$  )** is the % of "baskets" in which the occurrence of  $\{P,Q\}$  correctly predicts the occurrence of  $\{Z\}$ , that is, it is the **conditional probability** that a transaction containing  $\{ P \cap Q \}$  also contains  $Z$ .

$$\text{Conf} ( \{P,Q\} \Rightarrow \{Z\} ) = \text{Prob}(Z|P,Q).$$

c)

By the rule  $\{ D \} \Rightarrow \{ B, C \}$

$$\text{Confidence} ( \{ D \} \Rightarrow \{ B, C \} ) = \frac{\text{Sup} ( D, B, C )}{\text{Sup}(D)}$$

Therefore, the Confidence of the rules:

$$\text{Confidence} ( \{ B, D \} \Rightarrow \{ C \} ) = \frac{\text{Sup} ( B, D, C )}{\text{Sup}(B,D)}$$

$$\text{Confidence} ( \{ C, D \} \Rightarrow \{ B \} ) = \frac{\text{Sup} ( C, D, B )}{\text{Sup}(C,D)}$$

The number of records in the database with B and D:

$$\{ \text{records with B} \} \cap \{ \text{records with D} \} \leq \{ \text{records with D} \}$$

the same applies to the number of records with C and D, so we can conclude that

$$\text{Confidence} ( \{ B, D \} \Rightarrow \{ C \} ) \geq \text{Confidence} ( \{ D \} \Rightarrow \{ B, C \} )$$

$$\text{Confidence} ( \{ C, D \} \Rightarrow \{ B \} ) \geq \text{Confidence} ( \{ D \} \Rightarrow \{ B, C \} )$$

**d)** Candidate sets of size 3

{A,C}, {B,C}, {B,D}, {C,D}

{A}, {B}, {C}, {D}

	Support
{A, B, C }	1 / 5
{A, B, D }	1 / 5
{ A, C, D }	1 / 5
{ B, C, D }	2 / 5

**e)**

Sets of size 1

Item_set	Support
{ A }	3/5 ( 0,6)
{ B }	2/5 (0,4) eliminado
{ C }	3/5 (0,6)
{ D }	3/5 (0,6)

Sets of size 2

Item_set	Support
{ A, C }	2/5 (0,4) eliminado
{ A, D }	1/5 (0,2) eliminado
{ C, D }	2/5 (0,4) eliminado

R: No Frequent sets are found.

**f)**

**Conf<sub>min</sub> = 0,4**

Rules	Support	Confidence
<del>{A} ⇒ {B}</del>	<del>1/5 = 0,2</del>	<del>1/3 = 0,33</del>
{ B } ⇒ { A }	1/5 = 0,2	1/2 = 0,5
{ B } ⇒ { C }	2/5 = 0,4	2/2 = 1
{ C } ⇒ { B }	2/5 = 0,4	2/3 = 0,67

R: The rules { B } ⇒ { A }, { B } ⇒ { C } e { C } ⇒ { B } would be produced.

3.

a)

Cli	x	y	s	n	w	t	r
A	1	0	1	0	1	0	0
B	1	0	1	0	0	0	1
C	0	1	1	0	0	1	0
D	1	0	0	1	0	0	1

b) Minimum Support = 50%

Cli	x	y	s	n	w	t	r
A	1	0	1	0	1	0	0
B	1	0	1	0	0	0	1
C	0	1	1	0	0	1	0
D	1	0	0	1	0	0	1
Sup	<b>3/4</b>	1/4	<b>3/4</b>	1/4	1/4	1/4	<b>2/4</b>