

IT414 Assignment 1

NAME: SUYASH CHINTAWAR

ROLL NO.: 191IT109

TOPIC: APRIORI ALGORITHM

Note:

1) The colab link has been attached below. After opening the link, if it opens in drive, click on “Open with Google Colaboratory” to view the complete code.

Colab notebook link:

https://colab.research.google.com/drive/1Zy6vzLs-FqFmOUaOLmq4I3jyEdTAWI_CV

Q. Generate the frequent itemsets and strong association rules for the below transaction data considering minimum support count=2 and minimum confidence threshold of 70%.

Transactional Data for an *AlI*Electronics Branch

<i>TID</i>	<i>List of item IDs</i>
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

Output Screenshots:

Case 1) min_suport_count = 2 , confidence_threshold = 0.7

Frequent itemsets are:

L1 :

('I2',) : 7
('I5',) : 2
('I1',) : 6
('I4',) : 2
('I3',) : 6

L2 :

('I2', 'I5') : 2
('I1', 'I2') : 4
('I2', 'I4') : 2
('I2', 'I3') : 4
('I1', 'I5') : 2
('I1', 'I3') : 4

L3 :

('I1', 'I2', 'I5') : 2
('I1', 'I2', 'I3') : 2

The frequent itemsets are as shown above and the association rules are as shown in image below,

Association rules which satisfy threshold confidence

```
{'I5'} => {'I2'} 1.0  
{'I4'} => {'I2'} 1.0  
{'I5'} => {'I1'} 1.0  
{'I5'} => {'I2', 'I1'} 1.0  
{'I5', 'I1'} => {'I2'} 1.0  
{'I2', 'I5'} => {'I1'} 1.0
```

Total no. of associations: 6

Case 2) min_support_count = 3 , confidence_threshold = 0.5

Frequent itemsets:

Frequent itemsets are:

L1 :

```
('I2',) : 7  
( 'I1',) : 6  
( 'I3',) : 6
```

L2 :

```
('I1', 'I2') : 4  
( 'I2', 'I3') : 4  
( 'I1', 'I3') : 4
```

Association rules:

Association rules which satisfy threshold confidence

```
{'I1'} => {'I2'} 0.6666666666666666  
{'I2'} => {'I1'} 0.5714285714285714  
{'I2'} => {'I3'} 0.5714285714285714  
{'I3'} => {'I2'} 0.6666666666666666  
{'I1'} => {'I3'} 0.6666666666666666  
{'I3'} => {'I1'} 0.6666666666666666
```

Total no. of associations: 6

Q. Generate the frequent itemsets and strong association rules for the below transaction data considering minimum support = 60% and minimum confidence threshold of 80%.

<i>TID</i>	<i>items_bought</i>
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y}
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I, E}

Output Screenshots:

Case 1) min_support = 0.6 , confidence_threshold = 0.8

Frequent itemsets:

Frequent itemsets are:

L1 :

('K',) : 5

('O',) : 3

('Y',) : 3

('M',) : 3

('E',) : 4

L2 :

('K', 'O') : 3

('K', 'Y') : 3

('K', 'M') : 3

('E', 'K') : 4

('E', 'O') : 3

L3 :

('E', 'K', 'O') : 3

Association rules:

Association rules which satisfy threshold confidence

{'O'} => {'K'} 1.0

{'Y'} => {'K'} 1.0

{'M'} => {'K'} 1.0

{'E'} => {'K'} 1.0

{'K'} => {'E'} 0.8

{'O'} => {'E'} 1.0

{'O'} => {'K', 'E'} 1.0

{'O', 'E'} => {'K'} 1.0

{'K', 'O'} => {'E'} 1.0

Total no. of associations: 9

Case 2) min_support = 0.8 , confidence_threshold = 0.6

Frequent itemsets:

Frequent itemsets are:

L1 :

('K',) : 5

('E',) : 4

L2 :

('E', 'K') : 4

Association rules:

Association rules which satisfy threshold confidence

{'E'} => {'K'} 1.0

{'K'} => {'E'} 0.8

Total no. of associations: 2

Q3. Generate the frequent itemsets and strong association rules for the below transaction data considering minimum support = 60% and minimum confidence threshold of 80%.

Note :

1. Consider the first 20 rows and 5 columns for generating frequent itemsets and association rules of the SPECT dataset.
2. As binary values would be required for generating the data (presence of items), the SPECT data has been used instead of the SPECTF data which has numeric values. All the further analysis has been done on the SPECT.train data.

Data: The csv SPECT data when converted into python dictionary format gives the following result, ('Tx' -> Transaction, 'Cx' -> Column/Attribute/Item)

```

T1 : {'C4'}
T2 : {'C4', 'C3'}
T3 : {'C1', 'C5', 'C3'}
T4 : set()
T5 : set()
T6 : {'C4'}
T7 : {'C1', 'C4', 'C3'}
T8 : {'C3'}
T9 : {'C3'}
T10 : {'C2'}
T11 : {'C1', 'C5', 'C2'}
T12 : {'C1', 'C5', 'C2'}
T13 : {'C1', 'C5'}
T14 : {'C2'}
T15 : {'C1', 'C4', 'C3'}
T16 : {'C2', 'C3'}
T17 : {'C4', 'C3'}
T18 : set()
T19 : {'C1', 'C5', 'C3'}
T20 : {'C1', 'C5'}

```

Output Screenshots:

Case 1) min_support = 0.6 , confidence_threshold = 0.8

Frequent itemsets:

```

Frequent itemsets are:
L1 :

```

Hence, no frequent itemsets are found with the support and confidence threshold values. Due to this, no association rules would be present. But, by varying support and confidence threshold values, we get,

Case 2) min_support = 0.2 , confidence_threshold = 0.5

Frequent itemsets:

```

Frequent itemsets are:
L1 :
('C4',) : 6
('C3',) : 9
('C1',) : 8
('C5',) : 6
('C2',) : 5

L2 :
('C3', 'C4') : 4
('C1', 'C3') : 4
('C1', 'C5') : 6

```

Association rules:

Association rules which satisfy threshold confidence

{'C4'} => {'C3'} 0.6666666666666666

{'C1'} => {'C3'} 0.5

{'C1'} => {'C5'} 0.75

{'C5'} => {'C1'} 1.0

Total no. of associations: 4

Conclusion: Hence, it can be seen that relaxing the support and confidence thresholds results in more no. of frequent itemsets and association rules being generated and similarly stricter support and confidence results in lesser and lesser rules and frequent itemsets (i.e. as support and confidence values increase, the resulting rules or frequent itemsets are subsets of that at the previous values of support and confidence)

THANK YOU