

Email : LE000288@umn.edu

ID : 5674954

Question 1

(a)

$$3^7 - 2^8 + 1 = 1932$$

1932 association rules.

(b)

4.

(c)

$$C_3^7 = \frac{7!}{3!4!} = 35$$

(d)

$\{Bread\}$: 5

$\{Milk\}$: 5

$\{Bread, Milk\}$: 3

(e)

confidence of the $\{Bread\} \rightarrow \{Milk\}$: $\frac{3}{5}$

confidence of the $\{Milk\} \rightarrow \{Bread\}$: $\frac{3}{5}$

(f)

If the confidence is 0, there is no conclusion about the support of $\{a\}$ and the support of $\{b\}$. If the confidence is not equal to 0, the support of $\{a\}$

would be the same as the support of $\{b\}$.

Question 2

(a)

No. $\{A, E\}$ should be frequent since Apriori principle tells us that "If an itemset is frequent, then all of its subsets must also be frequent".

(b)

True.

(c)

No. $\{A, B, C, D, E, F\}$ should be not frequent since Apriori principle tells us that "If an itemset is not frequent, then all of its supersets must also be not frequent". $\{A, B, D, E\}$ is not frequent, so $\{A, B, C, D, E, F\}$ is not frequent.

Question 3

(a)

True

(b)

False. $\{A, C, H\}$ has the same support count as that of the $\{A, C\}$

(c)

False. Its support count is 10, which is less than the minsup (20). So $\{A, C, D\}$ is not a frequent itemset.

(d)

True

Question 4

(a)

C

(b)

I

(c)

C

(d)

N

(e)

I

Question 5

(a)

$\{p, q, r, s\}, \{p, q, r, t\}, \{p, q, r, w\}, \{p, q, s, t\}, \{p, q, s, w\}, \{p, r, s, t\}, \{p, r, s, w\},$
 $\{p, r, t, w\}, \{p, s, t, w\}, \{q, s, t, r\}, \{q, s, t, w\}, \{r, s, t, w\}$

(b)

$\{p, q, r, s\}, \{p, r, s, t\}, \{p, s, t, w\}$

(c)

$\{p, r, s, t\}$

(d)

No. We only have one frequent 4-itemset. If we want to generate a frequent 5-itemset, we at least need 2 frequent 3-itemsets.

Question 6

(a)

C

(b)

C

(c)

B

(d)

B

(e)

A

Question 7

(a)

1.

No

2.

No

3.

Yes

(b)

cosine measure. Cosine measure is better for measuring the similarity between the documents because the document vectors are sparse and asymmetry. Furthermore, Correlation is non-invariant under null addition.

Question 8

(a)

$$\alpha_1 = 20, \alpha_2 = 30.$$

$$\text{support} = 16.67\%, \text{confidence} = 100\%$$

(b)

No pair of values for α_1 and α_2 satisfy the requirements.

(c)

$$\alpha_1 = 10, \alpha_2 = 13.$$

$$\text{support} = 16.67\%, \text{confidence} = 100\%$$

(d)

If the interval is too wide, some rules may merge several disparate patterns and may lose some of the interesting patterns.

If the interval too narrow, some windows may not meet support threshold.

Question 9

(a)

$$\langle \{r\}, \{s, t\}, \{w\} \rangle: \langle \{r\}, \{s, t\} \rangle + \langle \{s, t\}, \{w\} \rangle$$

$$\langle \{r\}, \{r, s, t\} \rangle: \langle \{r\}, \{s, t\} \rangle + \langle \{r, s, t\} \rangle$$

$$\langle \{p\}, \{s, t\}, \{w\} \rangle: \langle \{p\}, \{s, t\} \rangle + \langle \{s, t\}, \{w\} \rangle$$

$$\langle \{p\}, \{r, s, t\} \rangle: \langle \{p\}, \{s, t\} \rangle + \langle \{r, s, t\} \rangle, \langle \{p\}, \{r, t\} \rangle + \langle \{r, s, t\} \rangle$$

$$\langle \{p\}, \{r\}, \{s, t\} \rangle: \langle \{p\}, \{r\}, \{t\} \rangle + \langle \{r\}, \{s, t\} \rangle, \langle \{p\}, \{r\}, \{s\} \rangle + \langle \{r\}, \{s, t\} \rangle$$

$$\langle \{p, r, s, t\} \rangle: \langle \{p, r, s\} \rangle + \langle \{r, s, t\} \rangle$$

$$\langle \{r, s, t\}, \{w\} \rangle: \langle \{r, s, t\} \rangle + \langle \{s, t\}, \{w\} \rangle$$

(b)

$$\langle \{p\}, \{r\}, \{s, t\} \rangle$$

Question 10

(a)

No. We could generate a 4-sequence if the subsequence obtained by removing an event from the first element in $w1$ is the same as the subsequence obtained by removing an event from the last element in $w2$. But no matter we put any one of the $\langle \{A, B\}, \{E\} \rangle$ and $\langle \{A\}, \{C\}, \{E\} \rangle$ as the $w1$ and the other one as $w2$, we can not get the same subsequence after removing an event from the first element in $w1$ and removing an event from the last element in $w2$.

(b)

Yes. $\langle \{B\}, \{D, E\} \rangle$ is not in the frequent 3-sequence set.