

CS 422 Homework6

Xingli Li, Illinois Institute of Technology

1. Exercises

1.1

Tan, Ch. 5 (Association Analysis) Question 15

- a. Which data set(s) will produce the most number of frequent itemsets?

Answer:

Data set (e).

- b. Which data set(s) will produce the fewest number of frequent itemsets?

Answer:

Data set (d).

- c. Which data set(s) will produce the longest frequent itemset?

Answer:

Data set (e).

- d. Which data set(s) will produce frequent itemsets with highest maximum support?

Answer:

Data set (b).

- e. Which data set(s) will produce frequent itemsets containing items with wide-varying support levels (i.e., items with mixed support, ranging from less than 20% to more than 70%)?

Answer:

Data set (e).

1.2

Zaki, Chapter 8 (Frequent Pattern Mining) Questions 1(a), 4

1-(a)

Answer:

No.	Date	
tid	itemset	
t ₁	A B C D	
t ₂	A C D F	minsup = $\frac{3}{8}$.
t ₃	A C D E G	→
t ₄	A B D F	minimum support
t ₅	B C G	count = 3
t ₆	D F G	
t ₇	A B G	
t ₈	C D F G	
①		
	Frequent itemsets	
	{A}	support count.
	{B}	5
	{C}	4
	{D}	5
	{E}	6
	{F}	1
	{G}	4
	{A, B}	5.
	{A, C}	✓
	{A, D}	✓
	{A, F}	✓
	{A, G}	✓
	{B, C}	✓
	{B, D}	✓
	{B, F}	✓
	{B, G}	✓
	{C, D}	✓
	{C, F}	✓
	{C, G}	✓
	{D, F}	✓
	{D, G}	✓
	{E, G}	✓
	{D, F, G}	✓
	filter	

② 1-Hemsets.

2-Itemset	minsup = $\frac{3}{8}$.	Frequent itemsets.	support count	filter
{A, B}	→	{A, B}	3	✓
{A, C}	minimum support count: 3	{A, C}	3	✓
{A, D}		{A, D}	4	✓
{A, F}		{A, F}	2	
{A, G}		{A, G}	2	
{B, C}		{B, C}	2	
{B, D}		{B, D}	2	
{B, F}		{B, F}	2	
{B, G}		{B, G}	2	
{C, D}		{C, D}	4	✓
{C, F}		{C, F}	2	
{C, G}		{C, G}	3	✓
{D, F}		{D, F}	4	✓
{D, G}		{D, G}	3	✓
{F, G}		{F, G}	2	
{D, F, G}		{D, F, G}	1	

203 211

3D, 11.

(3)

(4)

No.

Date

3-Itemsets

 $\{A, B, C\}$ minsup = $\frac{3}{8}$. $\{A, B, D\}$

→

 $\{A, C, D\}$

minimum support count = 3

 $\{A, C, G\}$ $\{A, D, F\}$ $\{A, D, G\}$ $\{C, D, G\}$ $\{D, G, F\}$ $\{D, G, F\}$

(5)

Frequent itemsets support count filter

 $\{A, B, C\}$ 1 $\{A, B, D\}$ 2 $\{A, C, D\}$ 3 ✓ $\{A, C, G\}$ 1 $\{A, D, F\}$ 2 $\{A, D, G\}$ 1 $\{C, D, G\}$ 2 $\{C, D, F\}$ 2 $\{D, G, F\}$ 2

(6).

From 1-Itemsets, 2-Itemsets, 3-Itemsets,

Summary:

Itemset count -

 $\{A\}$ 5 $\{B\}$ 4 $\{C\}$ 5 $\{D\}$ 6 $\{F\}$ 4 $\{G\}$ 5 $\{A, B\}$ 3 $\{A, C\}$ 3 $\{A, D\}$ 4 $\{C, D\}$ 3 $\{C, G\}$ 4

$\{D, F\}$	4
$\{D, G\}$	3
$\{A, C, D\}$	3.

- Q4.** Given the database in Table 8.4. Show all rules that one can generate from the set ABE .

Table 8.4. Dataset for Q4

tid	itemset
t_1	ACD
t_2	BCE
t_3	$ABCE$
t_4	BDE
t_5	$ABCE$
t_6	$ABCD$

Rules	support	confidence
$\{A, B\} \rightarrow \{E\}$	0.33	0.66
$\{A, E\} \rightarrow \{B\}$	0.33	0.99
$\{B, E\} \rightarrow \{A\}$	0.33	0.49
$\{E\} \rightarrow \{A, B\}$	0.33	0.49
$\{B\} \rightarrow \{A, E\}$	0.33	0.39
$\{A\} \rightarrow \{B, E\}$	0.33	0.39