

CS145 HW4

Calvin Chen

004943505

1.

a)

TDB. $Sup_{min} = 2$

Tid	Items		Itemset	Sup		Itemset	Sup		Itemset		Itemset	Sup
1	b, c, f	1st Scan	{a}	6	→	{a}	6	→	{a, b}	2nd Scan	{a, b}	4
2	a, b, d		{b}	8		{b}	8		{a, c}		{a, c}	4
3	a, c		{c}	6		{c}	6		{a, d}		{a, d}	2
4	b, d		{d}	4		{c}	6		{a, e}		{a, e}	2
5	a, b, c, e		{e}	2		{d}	4		{b, c}		{b, c}	4
6	b, c, k		{e}	2		{e}	2		{b, d}		{b, d}	4
7	a, c		{f}	1					{b, e}		{b, e}	2
8	a, b, c, d		{f}	1					{c, d}		{c, d}	1
9	b, d		{k}	1					{c, e}		{c, e}	1
10	a, b, c, d								{d, e}		{d, e}	0

L_2	Itemset	Sup		C_3	Item Set		C_3	Item Set	Sup
→	$\{a, b\}$	4	→	$\{a, b, c\}$	3^{rd} Scan	→	$\{a, b, c\}$	2	
	$\{a, c\}$	4		$\{a, b, d\}$			2		
	$\{a, d\}$	2		$\{a, b, e\}$			2		
	$\{a, e\}$	2							
	$\{b, c\}$	4							
	$\{b, d\}$	4							
	$\{b, e\}$	2							

freq patterns :

{a, b}
 {a, c}
 {a} {a, d}
 {b} {a, e}
 {c} {b, c} {a, b, c}
 {d} {b, d} {a, b, d}
 {e} {b, e} {a, b, e}

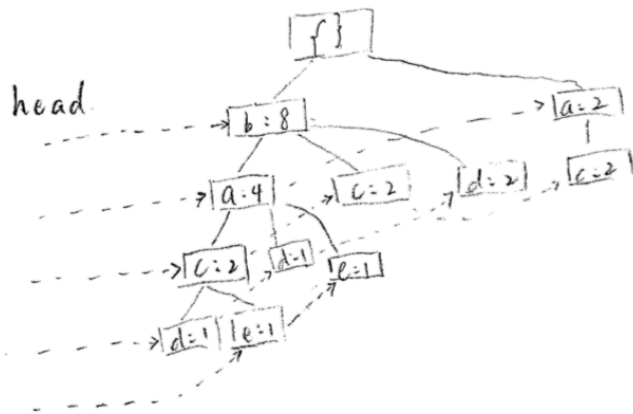
Header	TID	(ordered) freq items
a: 6.	1	{ b, c }
b: 8.	2	{ b, a, d }
c: 6	3	{ a, c }
d: 4	4	{ b, d }
e: 2	5	{ b, a, c, e }
i: 1	6	{ b, c }
j: 1	7	{ a, c }
k: 1	8	{ b, a, e }
	9	{ b, d }
	10.	{ b, a, c, d }

min-support = 2

b).

header Table:

item	freq
b.	8
a.	6
c.	6
d.	4
e.	2



f-list: b-a-c-d-e

c)

d-conditional database:

bac: 1 ba: 1 b: 2

{ }
|
b: 4.
|
a: 2

d)

d.
bd.
ad.
bad.

2.

a)

$$\text{Confidence (Nuts} \Rightarrow \text{Beer)} = \frac{150}{850} = 17.6\%$$

$$\text{Confidence (Beer} \Rightarrow \text{Nuts)} = \frac{150}{500} = 30\%$$

$$\text{Lift (Nuts, Beer)} = \frac{\frac{150}{15000}}{\frac{500}{10000} \cdot \frac{850}{10000}} = 3.529$$

$$\text{all_confidence (Nuts, Beer)} = 17.6\%$$

χ^2

	B		¬B.	
N	150 (42.5)	700 (807.5)	850	
¬N	350 (457.5)	8800 (8692.5)	9150	
	500	9500	10000	

$$\chi^2 = \frac{(150 - 42.5)^2}{42.5} + \frac{(700 - 807.5)^2}{807.5} + \frac{(350 - 457.5)^2}{457.5} + \frac{(8800 - 8692.5)^2}{8692.5}$$

$$= 312.8$$

b) We can conclude from the above measures that those who buy beer are more likely to buy nuts at the same time than those who buy nuts with beer. Also, beer and nuts are positive relative.

3.

a) 4 elements, length = 6, non-empty subsequences = $2^4 * 2 * 2 - 1 = 63$

b)

$$L_3 = \{ \langle (ac)e \rangle, \langle b(c)d \rangle, \langle bce \rangle, \langle a(cd) \rangle, \langle (ab)d \rangle, \langle (abc) \rangle \}$$

Join:

abcd, abce

Pruning:

✓ abcd: abc abd, acd, bcd

~~abce: abc, abe, ace, bce.~~

4.

S ₁ /S ₂	2	3	2	1	7	4	3	0	2	5
1		1	2	1	0	6	3	2	1	4
2		0	1	0	1	5	2	1	2	0
5		3	2	3	4	2	1	2	5	3
3		1	0	1	2	4	1	0	3	1
2		0	1	0	1	5	2	1	2	0
1		1	2	1	0	6	3	2	1	4
7		5	4	5	6	0	3	4	7	5

S ₁ /S ₂	2	3	2	1	7	4	3	0	2	5
1		①	3	4	4	10	13	15	16	21
2		1	②	②	③	8	10	11	13	16
5		4	3	5	6	⑤	6	8	13	16
3		5	3	4	6	9	⑥	⑥	9	10
2		5	4	3	4	9	8	7	⑧	8
1		6	6	4	3	9	11	9	8	⑨
7		11	10	9	9	3	6	10	15	13

distance = 11