Homework 6

1.1 Tan, Ch. 5 (Association analysis)

(15) a) In order to get as many frequent itemsets as possible, we need items that apper frequently (dense columns) and these items should overlaps with other frequent items.

Here, the correct answer is a) or e), as both have many frequent itemsets.

---

b) We can notice that the itemsets in dataset d) are not frequent which tells us that there will be the least amount of fr. itemsets

⇒ d)

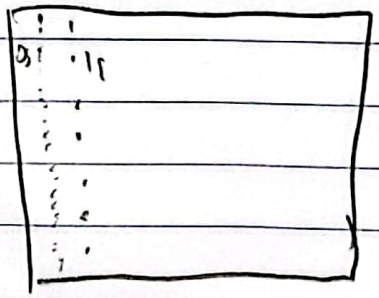
---

c) The longest frequent itemset would be in e).

---

d) We see that the most dense column is in f) and b). This tells us that it has the highest support amongst the frequent itemsets. However, if we talk about the overall highest max support of all of the itemsets in the specific dataset, it would be e).

f) →



90% 0's (only 10% 1's) 12 items?

e) Here, again, the answer is f) and B)

In f), we have item 0 with 90% support and item 1 with 10% support.

In B), we have item ~160 with >70% support and many other items with <20%.

1.2 Zaki, Chapter 8 (Frequent Pattern Mining)

① a) minsup = $\frac{3}{8}$

C1

Itemset	sup. count
A	5
B	4
C	5
D	6
E	1
F	4
G	5

→

L1

Itemsets	sup. count
A	5
B	4
C	5
D	6
F	4
G	5

C2

Itemsets	Sup. count
AB	3
AC	3
AD	4
AF	2
AG	2
BC	2
BD	2
BF	1
BG	2
CD	4
CF	2
CG	3
DF	4
DG	3
FG	1

→

L2

Itemsets	Sup. count
AB	3
AC	3
AD	4
CD	4
CG	3
DF	4
DG	3

C3

Itemsets	Sup. count
ABC	1
ABD	2
ACD	3
ACG	1
ADF	2
ADG	1
CDG	2
CDF	2
DFG	1

→

L3

Itemsets	Sup. count
ACD	3

C1, L1

(4)

Itemset	sup. count
A	4
B	5
E	4

C2

Itemset	Sup. count
AB	3
AE	2
BE	4

C3

Itemsets	Sup. count
ABE	2

$conf(A \rightarrow BE) = \frac{2}{4} = \frac{1}{2} \Rightarrow 50\%$

$conf(B \rightarrow AE) = \frac{2}{2} = 1 \Rightarrow 100\%$

$conf(E \rightarrow AB) = \frac{2}{3} \Rightarrow \sim 67\%$

$conf(AB \rightarrow E) = \frac{2}{4} = \frac{1}{2} \Rightarrow 50\%$

$conf(AE \rightarrow B) = \frac{2}{5} \Rightarrow 40\%$

$conf(BE \rightarrow A) = \frac{2}{4} = \frac{1}{2} \Rightarrow 50\%$