Yohandi - 120040025

CSC3170 Assignment 4

Q1

(i) $\prod_{i=1}^{k} n_i$

(ii) $\prod_{i=1}^{k} (n_i+1)$

(iii) $\sum_{i=1}^{k} \left(\prod_{j=1}^{i} n_j \right) + 1$

Q2(a). The confidence is given by $\frac{\sigma(\text{coffee}, \text{tea})}{\sigma(\text{tea})} = \frac{15}{20} = 0.75$

However, $P(\text{coffee} \mid !\text{tea}) = \frac{75}{80} \approx 0.94$

This implies that the tea → coffee rule doesn't stand even with high confidence, therefore, confidence is not a useful measure in this particular case.

(b). Note that $\text{Lift}(A,B) = \frac{P(A|B)}{P(A)} = \frac{P(A \cap B)}{P(A)P(B)}$

- $\text{Lift}(A,B) = 1$

  $\Rightarrow \frac{P(A \cap B)}{P(A)P(B)} = 1 \Rightarrow P(A \cap B) = P(A)P(B) \Rightarrow$ event A and B are independent

- $\text{Lift}(A,B) > 1$

  $\Rightarrow P(A \cap B) > P(A)P(B) \Rightarrow$ event A and B are positively correlated

- $\text{Lift}(A,B) < 1$

  $\Rightarrow P(A \cap B) < P(A)P(B) \Rightarrow$ event A and B are negatively correlated

- $\text{Lift}(A,B) = 0$

  $\Rightarrow P(A \cap B) = 0 \Rightarrow P(A \cup B) = P(A) + P(B) \Rightarrow$ event A and B are mutually exclusive

(c) $\text{Lift}(\text{tea}, \text{coffee}) = \frac{P(\text{tea}, \text{coffee})}{P(\text{tea})P(\text{coffee})} = \frac{0.15}{0.9 \cdot 0.2} = 0.8\overline{3}$

Q3 $\text{Lift}(\text{compiler}, \text{mining}) = \frac{P(\text{compiler}, \text{mining})}{P(\text{compiler})P(\text{mining})} = \frac{0.1}{0.1 \cdot 0.1} = 10$

$$Lift(data, mining) = \frac{P(data, mining)}{P(data)P(mining)} = \frac{0.9}{0.9 \cdot 0.9} = 1.\bar{1}$$

We notice that the results indicate that terms "compiler" with "mining" correlate better than terms "data" with "mining", which is not quite same to the general expectation. Therefore, the usage of lift is not a good measure.

Q4 (a)

α_1	α_2	support	confidence
9	17	$\frac{2}{12} = 0.1\bar{6}$	$\frac{2}{4} = 0.5$
19	29	$\frac{2}{12} = 0.1\bar{6}$	$\frac{2}{2} = 1$
33	47	$\frac{0}{12} = 0$	0

we notice that $\alpha_1 = 19$ and $\alpha_2 = 29$ satisfy the thresholds requirement

(b)

α_1	α_2	support	confidence
9	14	$\frac{2}{12} = 0.1\bar{6}$	$\frac{2}{3} = 0.\bar{6}$
17	21	$\frac{1}{12} = 0.08\bar{3}$	$\frac{1}{2} = 0.5$
25	33	$\frac{1}{12} = 0.08\bar{3}$	$\frac{1}{1} = 1$
39	47	$\frac{0}{12} = 0$	0

we notice that no (α_1, α_2) pair that satisfies the thresholds requirement. nothing to compare

(c)

α_1	α_2	support	confidence
9	11	$\frac{2}{12} = 0.1\bar{6}$	$\frac{2}{2} = 1$
14	17	$\frac{0}{12} = 0$	$\frac{0}{2} = 0$
19	21	$\frac{1}{12} = 0.08\bar{3}$	$\frac{1}{1} = 1$
25	29	$\frac{1}{12} = 0.08\bar{3}$	$\frac{1}{1} = 1$
33	39	$\frac{0}{12} = 0$	0
41	47	$\frac{0}{12} = 0$	0

we notice that $\alpha_1 = 9$ and $\alpha_2 = 11$ satisfy the thresholds requirement. comparing this result to the one in part (a), we see that the intervals are not the same

Q5(a) There are 5 possible items for the consequent: Apple, Beer, Cake, Diaper, Eggs

$$\sum_{item \in Items} (\underbrace{2^{|Items - \{item\}|} - 1}_{})$$

since we can just select / not select for the rest of items and exclude the empty set

$= 5(2^{5-1} - 1) = 75$ possibilities

(b) As generalized above, for N items, there are $N(2^{N-1} - 1)$ possibilities

Q6(a)

Item	Count	
Apple	5	
Beer	4	
Cake	5	
Diaper	2	×
Eggs	2	×

Item	Count
{Apple, Beer}	3
{Apple, Cake}	4
{Beer, Cake}	3

Item	Count
{Apple, Beer, Cake}	~~3~~ 2

stop here

Apple→Beer, Beer→Apple, Apple→Cake, Cake→Apple, Beer→Cake, Cake→Beer

(b) $c(\text{Apple} \rightarrow \text{Beer}) = \frac{3}{5} = 60\%$ ×

$c(\text{Beer} \rightarrow \text{Apple}) = \frac{3}{4} = 75\%$

$c(\text{Apple} \rightarrow \text{Cake}) = \frac{4}{5} = 80\%$

$c(\text{Cake} \rightarrow \text{Apple}) = \frac{4}{5} = 80\%$

$c(\text{Beer} \rightarrow \text{Cake}) = \frac{3}{4} = 75\%$

these are the ones that we are looking for

$c(\text{Cake} \rightarrow \text{Beer}) = \frac{3}{5} = 60\%$ ×