Table 4: Count Example 1

Num	Itemset
1	{A,C}
2	{B,C}
3	{A,B}
4	{C}
5	{A,B,C}

Table 5: Count Example 2

Num	Itemset
1	{A,B}
2	{C}
3	{C}
4	{C}
5	{A,B,C}

3 Task3

3.1 Task3.a

We use the examples in following tables explaining maximal and closed cases.

1. Table 4 Example 1 - Maximal frequent sets. We can see set $\{A, B\}$ are not maximal frequent set, because its superset $\{A, B, C\}$ is frequent. We compute $lift(A \to B) = \frac{2/5}{3/5*3/5} = 10/9 > 1$. So the positive statistical associations $A \to B$ is ignored in this case.

2. Table 5 Example 2 - Closed frequent set. We can see set $\{A, B\}$ are not close frequent set, because its superset $\{A, B, C\}$, $P(ABC) = 1/5 = P(AB)$. But we compute $lift(A \to B) = \frac{2/5}{2/5*2/5} = 5/2 > 1$. So the positive statistical associations $A \to B$ is ignored in this case.

3. Table 6 Example 3 - 0-free frequent set. We can see set $\{A, B\}$ are not 0-free frequent set, because its subset set $P(A) = P(B) = 3/4 = P(AB)$. But we compute $lift(A \to B) = \frac{3/4}{3/4*3/4} = 4/3 > 1$. So the positive statistical associations $A \to B$ is ignored in this case.

3.2 Task3.b

In this task, we need to find all maximal sets in example, whose associations rules are not correct. But all correct associations rules are derived from other sets. The example 4 shows in Table 7.

We can see $\{A, C, D\}$ and $\{B, C, D\}$ are maximal sets and rules derived from them are $A, C \to D$ and $B, D \to C$. Then we compute $lift(A, C \to D) = \frac{1/6}{1/2*2/3} = 1/2 < 1$ and

Table 6: Count Example 3

Num	Itemset
1	{A,B}
2	{A,B}
3	{C}
4	{A,B,C}

Table 7: Count Example 4

Num	Itemset
1	{A,C}
2	{A,C}
3	{B,D}
4	{B,D}
5	{A,C,D}
6	{B,C,D}

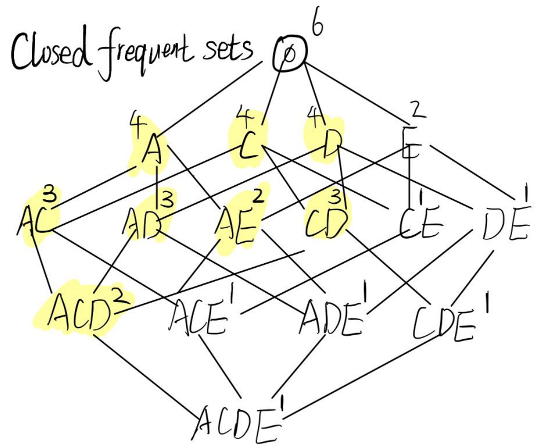
$lift(B, D \to C) = \frac{1/6}{1/2*2/3} = 1/2 < 1$. So rules in maximal sets are not effective. whereas all effective rules actually are $A \to C$ and $B \to D$, $lift(A \to C) = \frac{1/2}{1/2*2/3} = 3/2 > 1$, $lift(B \to D) = \frac{1/2}{1/2*2/3} = 3/2 > 1$, which produced from $\{A, C\}$ and $\{B, C\}$ not belonging to maximal sets.

3.3 Task3.c

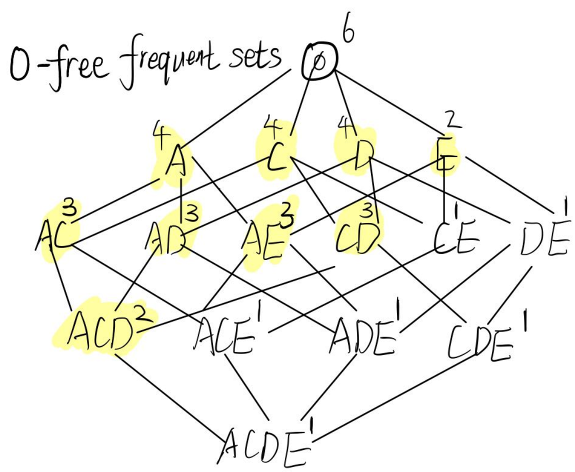
1. Maximal frequent sets. If only given maximal sets, it's hard to detect overfitted rules in these sets, because The subsets of maximal sets can not be maximal sets, there is no redundant in current rules. For example, if $\{A, B, C\}$ is a maximal set, subsets $\{A, B\}, \{A, C\}, \{B, C\}$ will not be considered association rules. There are no redundant rules derived from $\{A, B, C\}$.

2. Closed frequent sets. The example shows in Fig.1a. It an adaptation from P58 lecture 7. The subsets of closed sets $\{A, C, D\}$ are still closed sets. So there are a lot redundant in the rules derived from all closed sets. $\{A, C, D\}$ will most likely produce overfitted rules.

3. 0-free frequent sets. The example shows in Fig.1b. The subsets of 0-free sets $\{A, C, D\}$ are still 0-free sets. So there are a lot redundant in the rules derived from all 0-free sets. $\{A, C, D\}$ will most likely produce overfitted rules.



(a) Closed frequent sets



(b) 0-free frequent sets

Figure 1: Task3.C Examples
(Adaptation from P58 Lecture 7, $min_{fr} = 1/3$)