Hw 2 Report

Hw 2 - Association Rules

AU7008 Data Mining, SJTU, 2023 Spring

By **Prof. X. He**

Table of Contents

Problem 1

For the given dataset,

事务 ID	购买项
0001	{a,d,e}
0024	{a,b,c,e}
0012	{a,b,d,e}
0031	{a,c,d,e}
0015	{b,c,e}
0022	{b,d,e}
0029	{c,d}
0040	{a,b,c}
0033	{a,d,e}
0038	{a,b,e}

1. calculate the *support* of `{e}`, `{b,e}` and `{b,d,e}`;
2. calculate the *confidence* of association rules `{b,d}->{e}`, `{e}->{b,d}`

[Solution]

1.

- $s(e) = \frac{\sigma(e)}{N} = \frac{8}{10} = 0.8$

- $s(b, e) = \frac{5}{10} = 0.5$
- $s(b, d, e) = \frac{2}{10} = 0.2$

2.

- $c(b, d \to e) = \frac{s(b, d \cup e)}{s(b, d)} = \frac{0.2}{2/10} = 1$
- $c(e \to b, d) = \frac{2}{8} = 0.25$

Problem 2

[Solution]

By definition, we have,

$$\xi(X) = \min_i c(q_i \to X \backslash q_i) = \min_i \frac{s(X)}{s(q_i)} = \frac{s(X)}{\max_i s(q_i)}$$

from which, the optimal solution is $i = argmax_i \; s(q_i)$.

therefore, $\forall X, Y : X \subseteq Y$,

$$s(X) \geq s(Y), max_{x \in X} s(x) \leq max_{y \in Y} s(y) \Rightarrow \frac{s(X)}{\max_{x \in X} s(x)} \geq \frac{s(Y)}{\max_{y \in Y} s(y)}$$

i.e.,

$$\xi(X) \geq \xi(Y)$$

Problem 3

[Solution]

(a) for the mod3 hash function, the hash tree is given by:

```
    *
    |---------------------------
    |              |            |
 {1,2,3}        {2,3,4}      {3,4,6}
 {1,2,6}        {2,4,5}
 {1,3,4}
 {4,5,6}

 =>

    *
    |-------------------------------------
    |                      |            |
    |-------------      {2,3,4}      {3,4,6}
    |           |       {2,4,5}
 {1,2,3}     {1,3,4}
 {1,2,6}
 {4,5,6}
```

(b)

subset using hash tree:

```
1,2,3,4,6
   |
   |-------------------------------------------
   |                          |            |
  (1)                        (2)          (3)
   |                          |            |
   |--------------          {2,3,4}      {3,4,6}
   |             |          {2,4,5}
  (4)           (5)
   |             |
{1,2,3}       {1,3,4}
{1,2,6}
{4,5,6}
```
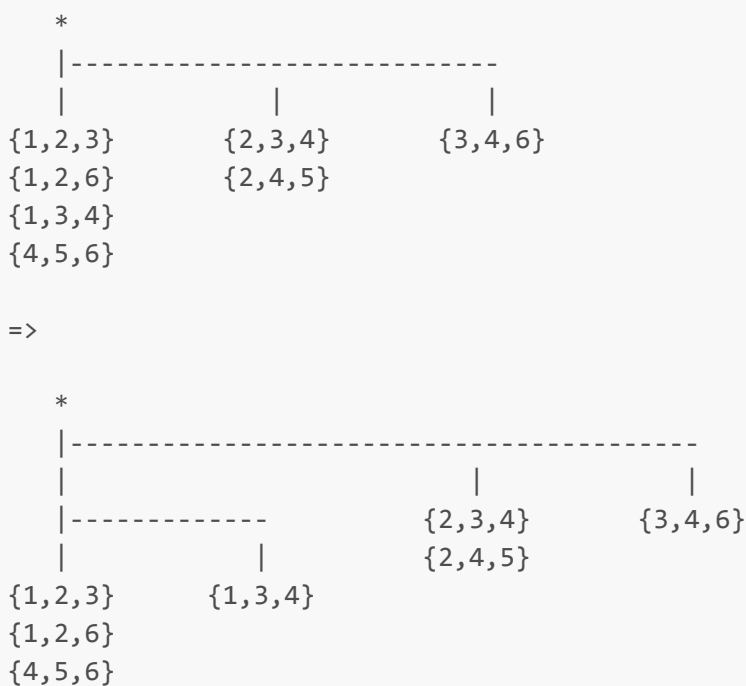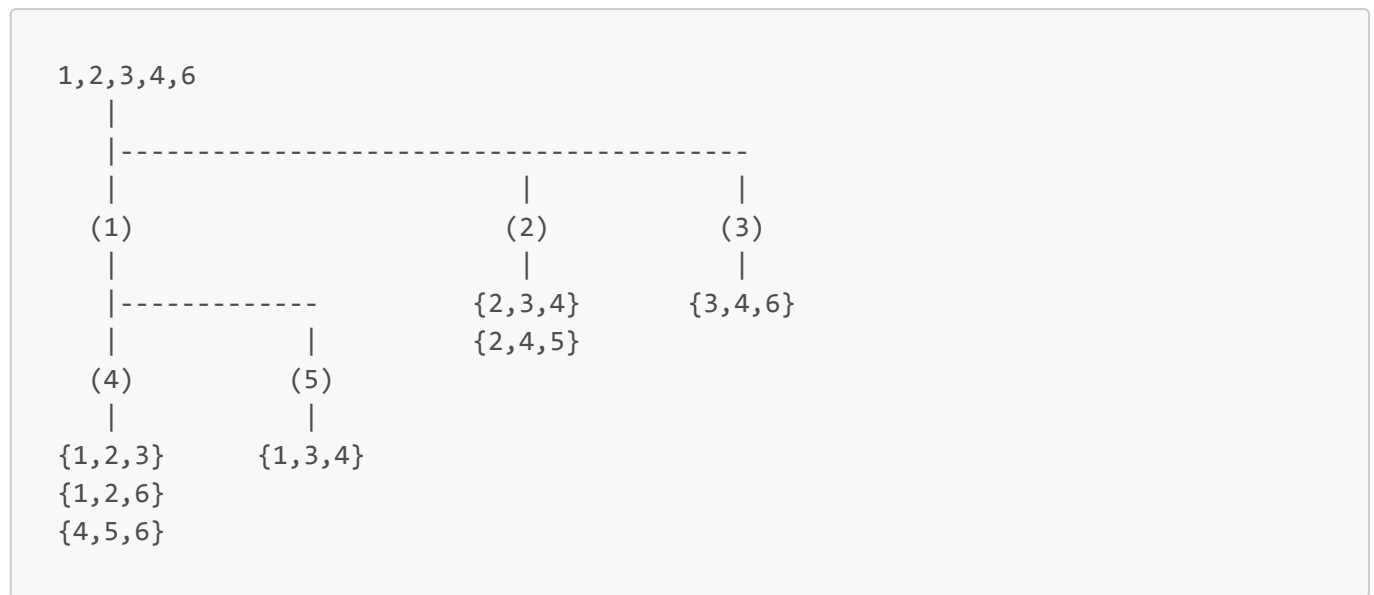
where,

- (1): 1 + 2,3,4,6
- (2): 2 + 3,4,6
- (3): 3 + 4,6
- (4): 1,2 + 3,4,6
- (5): 1,3 + 4,6

subsets of 3:

```
 1,2,3,4,6
    |
    |-----------------------------------------------------------------
    |                                    |                      |
 1 + 2,3,4,6                          2 + 3,4,6              3 + 4,6
    |                                    |                      |
    |----------------------             |-----------           |
    |          |          |             |          |            |
 1,2 + 3,4,6  1,3 + 4,6  1,4 + 6     2,3 + 4,6  2,4 + 6        |
    |          |          |             |          |            |
   1,2,3      1,3,4      1,4,6        2,3,4      2,4,6        3,4,6
   1,2,4      1,3,6                   2,3,6
   1,2,6
```