

CS412: Data Mining Homework 2

Problem 1

(a)

$$2^{10} = 1024$$

(b)

$$3$$

(c)

$$(2 * 2^{10} - 2) - 8 = 2038$$

(d)

$$1$$

(e)

$$8$$

Problem 2

(a)

True: $\text{sum}(x)$, $\text{sum}(x^2)$ and $\text{count}(x)$, which are distributive measures. $\text{average}(x)$ use above distributive measures from each subset to get the result.

(b)

True: $\text{average}(x) = 1/2[\text{min}(x) + \text{max}(x)]$. $\text{min}(x)$ and $\text{max}(x)$ are distributive measures. $\text{average}(x)$ use above distributive from each subset to get the result.

(c)

True: $\text{max_50}(x)$ and $\text{sum}(x)$ are distributive measures. $\text{max_50}(x)$ to get max 50 numbers from each subset.

(d)

False: If the original dataset has 1000 numbers. Thus, we need to find out the largest number. While if we separate 1000 numbers into 10 groups and each contains 100 numbers. Then we will fail to use the formula to find the result from each subset and get the final result.

(e)

True: Since the data is guaranteed to be binary. We can get the $\text{count}(0)$ and $\text{count}(1)$ of each subset. Finally get the final count of the whole dataset. 0 or 1 which has more count will be the mode.

Problem 3

(a)

{B, C, D}

(b)

$S = AB$

(c)

closed patterns: {D, AD, BC, BCD}

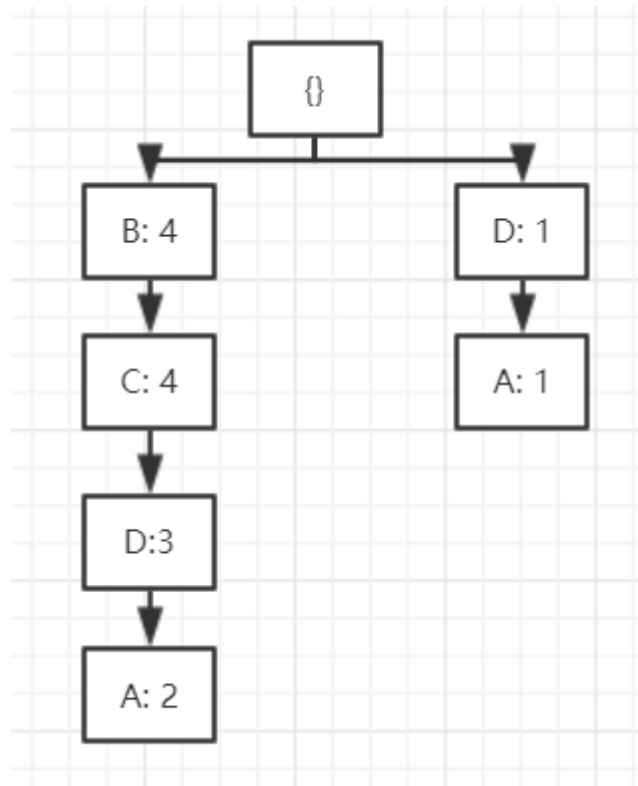
(d)

max-patterns: {AD, BCD}

(e)

$x \in \{001, 002, \dots, 005\}, \text{buys}(x, B) \wedge \text{buys}(x, C) \Rightarrow \text{buys}(x, D). [0.8, 0.75]$ $x \in \{001, 002, \dots, 005\}, \text{buys}(x, B) \wedge \text{buys}(x, D) \Rightarrow \text{buys}(x, C). [0.6, 1.0]$ $x \in \{001, 002, \dots, 005\}, \text{buys}(x, C) \wedge \text{buys}(x, D) \Rightarrow \text{buys}(x, B). [0.6, 1.0]$

(f)



(g)

A D:1, BCD:2

Problem 4

(a)

$\{B, C\}$ $\{B, D\}$ $\{C, D\}$ $\{B, C, D\}$

(b)

$\text{sum}(S.\text{price}) \geq 45$: monotonic

$\text{sum}(S.\text{price}) \leq 45$: anti-monotonic

We can use FP-Tree to mine frequent patterns with $\text{sum}(S.\text{price}) \leq 45$

(c)

Both are convertible because can based on sorted value not sequence order.

$\text{avg}(S.\text{price}) \geq 30$: use value descending order

$\text{avg}(S.\text{price}) \leq 30$: use value ascending order