Data Science HW4

R11922066 林庭岑

Problem 1

a)

Database D

TID	Items
1	ACDE
2	CDE
3	AC
4	BE
5	ABDE
6	ABCDE

Scan D →

C_1

Itemset	Sup.
{A}	4
{B}	3
{C}	4
{D}	4
{E}	5

L_1

Itemset	Sup.
{A}	4
{B}	3
{C}	4
{D}	4
{E}	5

C_2

Itemset
{AB}
{AC}
{AD}
{AE}
{BC}
{BD}
{BE}
{CD}
{CE}
{DE}

Scan D →

C_2

Itemset	Sup.
{AB}	2
{AC}	3
{AD}	3
{AE}	3
{BC}	1
{BD}	2
{BE}	3
{CD}	3
{CE}	3
{DE}	4

L_2

Itemset	Sup.
{AC}	3
{AD}	3
{AE}	3
{BE}	3
{CD}	3
{CE}	3
{DE}	4

C_3

Itemset
{ACD}
{ACE}
{ADE}
{CDE}

Scan D →

C_3

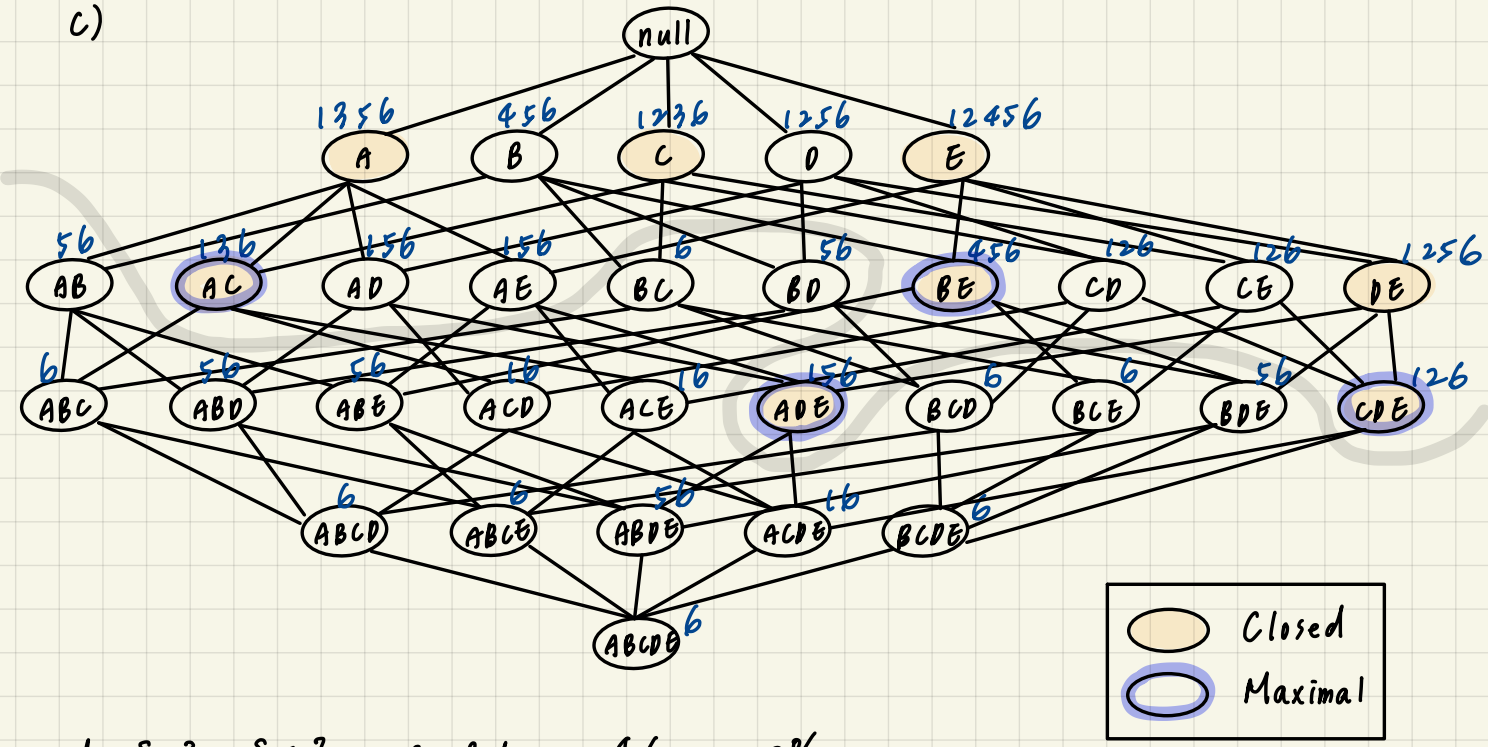
Itemset	Sup.
{ACD}	2
{ACE}	2
{ADE}	3
{CDE}	3

L_3

Itemset	Sup.
{ADE}	3
{CDE}	3

b) {A}, {B}, {C}, {D}, {E}
{AC}, {AD}, {AE}, {BE}, {CD}, {CE}, {DE}
{ADE}, {CDE}

c)



d) $\{D\} \rightarrow \{E\}$ Confidence = $4/4 = 100\%$

$\{AD\} \rightarrow \{E\}$ Confidence = $3/3 = 100\%$

$\{C\} \rightarrow \{DE\}$ Confidence = $3/4 = 75\%$

$\{A\} \rightarrow \{DE\}$ Confidence = $3/4 = 75\%$

$\{A\} \rightarrow \{C\}$ Confidence = $3/4 = 75\%$

e) $\{C\} \rightarrow \{E\}$

Confidence = $\frac{3}{4}$

$p(E) = \frac{5}{6}$

Lift = $\frac{3}{4} \times \frac{6}{5} = \frac{9}{10} < 1$

Problem 2

a) T_1

b) T_1 : $\{ABCDE\}, \{FGHIJ\} \Rightarrow 2$

T_2 : $\{ABG\} \Rightarrow 1$

T_3 : $\{ABC\}, \{CD\}, \{HIJ\} \Rightarrow 3$

c) T_1 : $\{ABCDE\}, \{FGHIJ\} \Rightarrow 2$

T_2 : $\{AB\}, \{ABG\} \Rightarrow 2$

T_3 : $\{ABC\}, \{BC\}, \{CD\}, \{HIJ\}, \{C\}, \{D\} \Rightarrow 6$

Problem 3

a) Earphone → Phone

confidence: $\frac{6}{6} = 1$

support: $\frac{6}{10} = 0.6$

b) Case → Earphone

confidence: $\frac{2}{5} = 0.4$

support: $\frac{2}{10} = 0.2$