Problem 1.

a)

Database DB:

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

Scan DB →

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
{A B}
{AC}
{AD}
{AE}
{BC}
{BD}
{BE}
{CD}
{CE}
{DE}

Scan DB →

C_2:

Itemset	Sup.
{A B}	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 DB →

C_3:

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

L_3

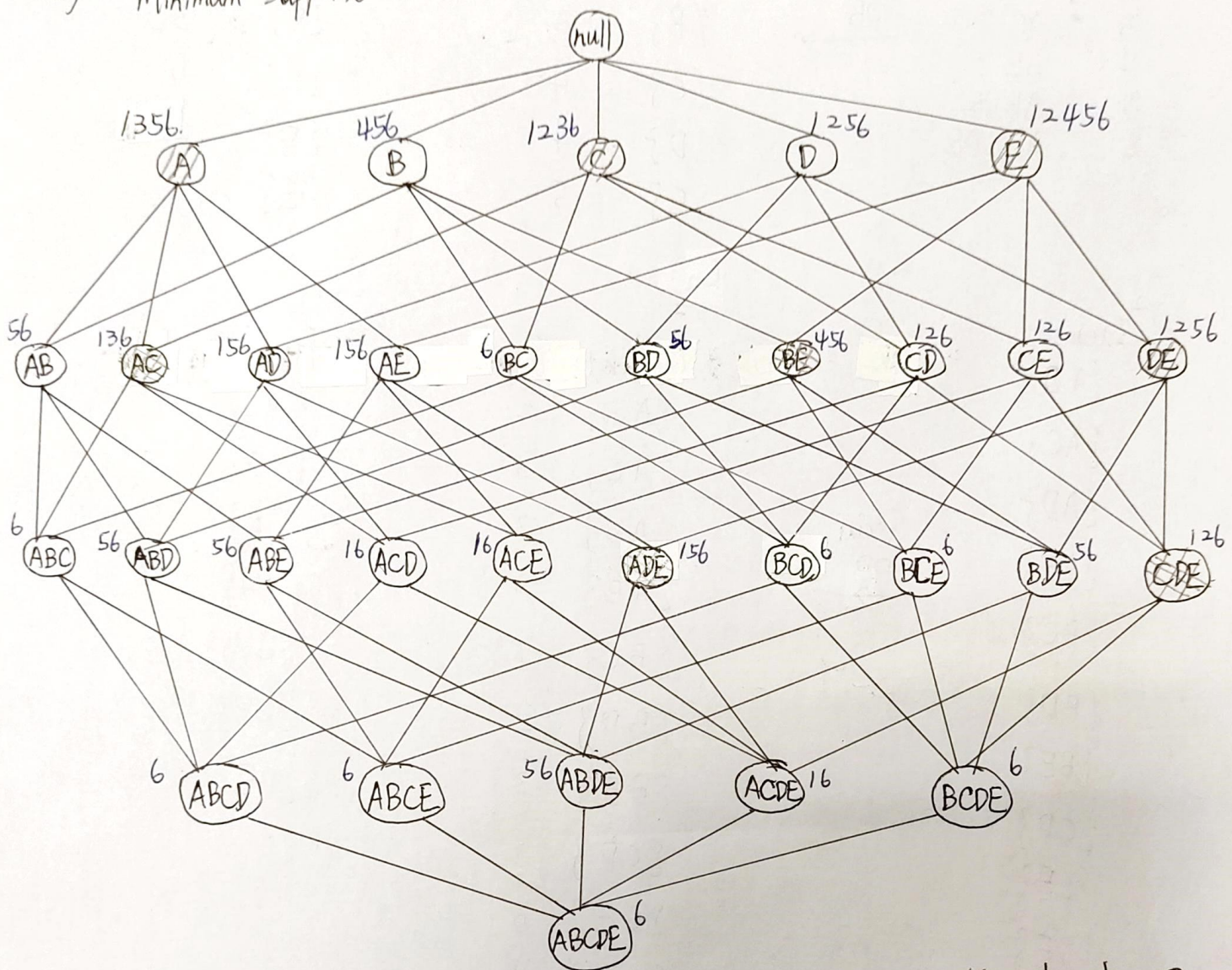| Itemset	Sup.
{ADE}	3
{CDE}	3

DE → C   conf : 75%

⋮

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

共 14 個

c) Minimum Support = 3

null

1356 — A
456 — B
1236 — C
1256 — D
12456 — E

56 — AB
136 — AC
156 — AD
156 — AE
6 — BC
56 — BD
456 — BE
126 — CD
126 — CE
1256 — DE

6 — ABC
56 — ABD
56 — ABE
16 — ACD
16 — ACE
156 — ADE
6 — BCD
6 — BCE
56 — BDE
126 — CDE

6 — ABCD
6 — ABCE
56 — ABDE
16 — ACDE
6 — BCDE

6 — ABCDE

/// Closed = 8
Maximal = 4

d)
CD → E   conf : 100%
AE → D   conf : 100%
AD → E   conf : 100%
CE → D   conf : 100%
DE → A   conf : 75%

e) $C \rightarrow E$

	E	$\bar{E}$	
C	3	1	4
$\bar{C}$	2	0	2
	5	1	

Confidence : $P(E|C) = \frac{3}{4} = 0.75$

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

$\Rightarrow$ Lift $= \frac{3}{4} / (\frac{5}{6}) = \frac{9}{10}$

$= 0.9$

(<1, there is negatively associated.)

Problem 2.

(a)

[T1] Longest frequent itemset : {ABCDE}, {FGHIJ}

[T2] Longest frequent itemset : {ABG}

[T3] Longest frequent itemset : {ABC}

$\Rightarrow$ T1 will produce the longest frequent itemset.

(b)

T1 : Support Item (by count) : 5

Frequent itemsets : All subset of {ABCDE} and

all subset of {FGHIJ}.

Maximal itemsets : {ABCDE}, {FGHIJ}

T2 : Support Item (by count) : 10

Frequent itemsets : {A}, {B}, {AB}

Maximal itemsets : {AB}

Support Item (by count) : 2

Frequent itemsets : {ABG}, {G}

Maximal itemsets : {ABG}

T3: Support Item (by count): 4

Frequent itemsets: {C}

Maximal itemsets: {C}

Support Item (by count): 3

Frequent itemsets: {B}, {D}, {BC}

Maximal itemsets: {D}, {BC}

Support Item (by count): 2

Frequent itemsets: {A}, {H}, {I}, {J}, {CD}, {ABC}, {HI}, {HJ}, {IJ}, {HIJ}, {AB}, {AC}

Maximal itemsets: {HIJ}, {ABC}, {CD}

∴ In T1, the number of maximal frequent itemsets is 2.

In T2, the number of maximal frequent itemsets is 1.

In T3, the number of maximal frequent itemsets is 3.

(c) T1: All support of frequent itemsets is 5,

∴ closed itemsets are {ABCDE} and {FGHIJ}.

T2:

Itemset	Support	"Closed itemsets"
{A}	10	
{B}	10	
{D}	1	
{E}	1	
{G}	2	
{I}	1	
{J}	1	
{A B}	10	✓
{ABD}	1	
{ABE}	1	
{ABG}	2	✓
{ABI}	1	
{ABJ}	1	
{ABGI}	1	
{ABGJ}	1	
{AG}	2	
{BG}	2	

T3:

Itemset	Support	"Closed itemsets"
{A}	2	
{B}	3	
{C}	4	✓
{D}	3	✓
{E}	1	
{F}	1	
{G}	1	
{H}	2	
{I}	2	
{J}	2	
{AB}	2	
{AC}	2	
{BC}	3	✓
{CD}	2	✓
{BD}	1	
{DE}	1	
{DF}	1	
{DG}	1	
{DH}	1	
{DI}	1	
{DJ}	1	
{EH}	1	
{EI}	1	
{EJ}	1	
{FH}	1	
{FI}	1	
{FJ}	1	
{GH}	1	
{GI}	1	
{GJ}	1	

Itemset	Support	"Closed itemsets"
{HI}	2	
{HJ}	2	
{IJ}	2	
{ABC}	2	✓
{BCD}	1	
{HIJ}	2	✓

In T1, the number of closed frequent itemsets is 2.

In T2, the number of closed frequent itemsets is 2.

In T3, the number of closed frequent itemsets is 6.

Problem 3.

(a) Earphone → Phone

$$\text{Support } \{Earphone, Phone\} = 6$$

$$\text{Confidence} = \frac{\text{Support } \{Earphone, Phone\}}{\text{Support } \{Earphone\}} = \frac{6}{6} \times 100\%$$

$$= 100\%$$

(b) Case → Earphone

$$\text{Support } \{Case, Earphone\} = 2$$

$$\text{Confidence} = \frac{\text{Support } \{Case, Earphone\}}{\text{Support } \{Case\}} = \frac{2}{5} \times 100\%$$

$$= 40\%$$