Homework 5

1.1] Ch. 5

① a) high support and high confidence

milk → bread

=> Everybody buys these 2 items a lot and usually together

↳ it could be interesting but it's usually not

b) toothpaste → beer

↳ People buy these 2 items a lot but very rarely together.

↳ not interesting; doesn't show us much

c) light bulb → super glue

↳ not many sales and not together

↳ doesn't tell us much => not interesting

d) cake → candles

↳ people don't buy these often but when they do, they buy them together

=> this is interesting

⇒ high confidence tends to be interesting in general;

(2) a) sup({e}) = 8/10 = 0.8 (80%)
sup({b,d}) = 2/10 = 0.2 (20%)
sup({b,d,e}) = 2/10 = 0.2 (20%)

b) $c = \text{conf}(X \to Y) = \frac{\sup(XY)}{\sup(X)}$

$\Rightarrow \text{conf}(\{b,d\} \to \{e\}) = \frac{0.2}{0.2} = 1$ (100%)

$\text{conf}(\{e\} \to \{b,d\}) = \frac{0.2}{0.8} = 0.25$ (25%)

→ confidence is not symmetric

c)
sup({e}) = 4/5 = 0.8 (80%)
sup({b,d}) = 5/5 = 1 (100%)
sup({b,d,e}) = 4/5 = 0.8 (80%)

d) $\text{conf}(\{bd\} \to \{e\}) = \frac{0.8}{1} = 0.8$ (80%)

$\text{conf}(\{e\} \to \{bd\}) = \frac{0.8}{0.8} = 1$ (100%)

e) Here, we already have s2 and c2.
Now, let's find s1 and c1 to compare.

$s_1(\{e\}) = \frac{8}{10} = 0.8$ (80%)

$sup(\{b,d\}) = \frac{2}{10} = 0.2$ (20%)

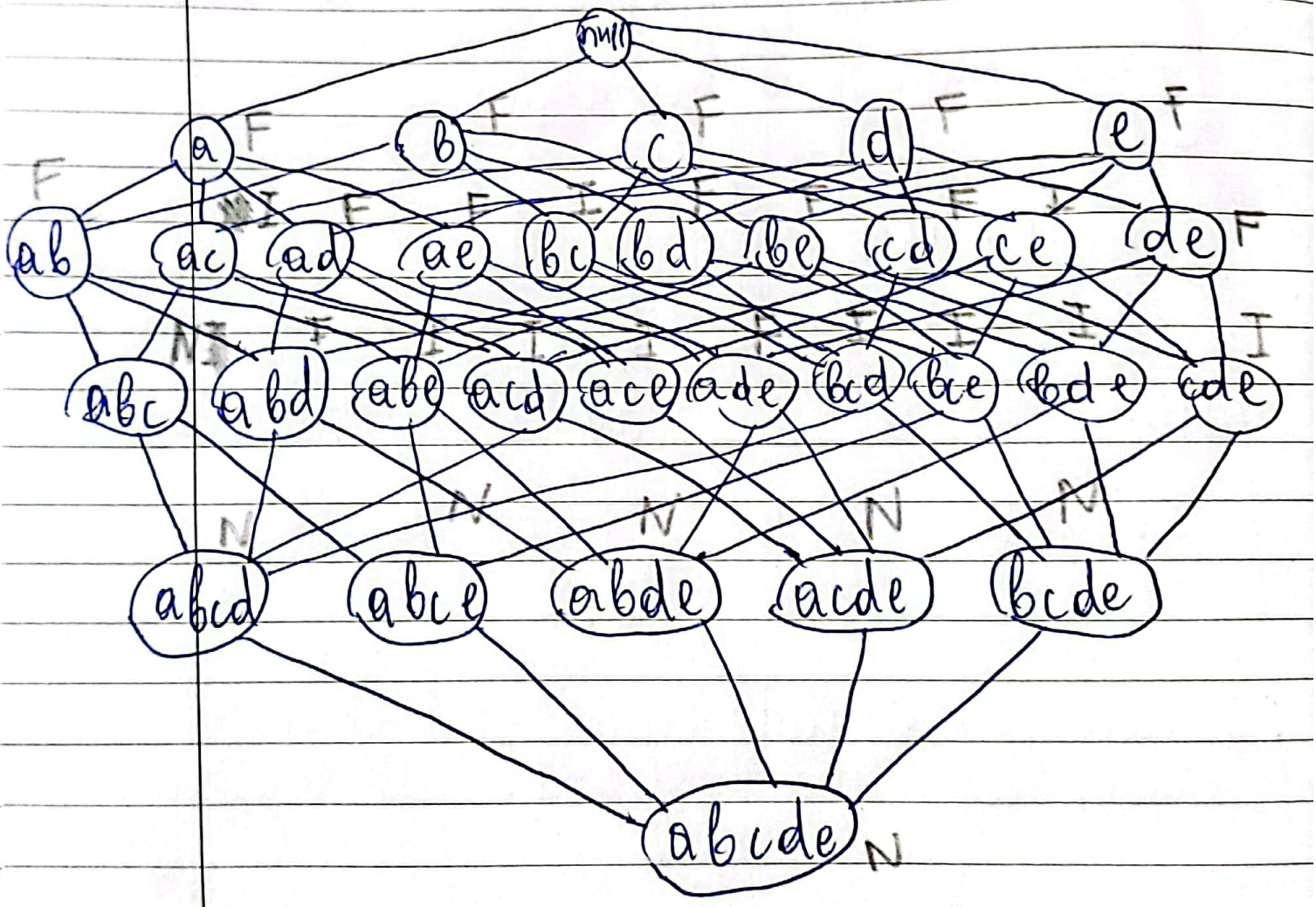
$sup(\{b,d,e\}) = \frac{2}{10} = 0.2$ (20%)

$c(\{b,d\} \rightarrow \{e\}) = 1$ (100%)

$c(\{b,c\} \rightarrow \{b,d\}) = 0.25$ (25%)

The only relationship that could be noticed is that the support of $\{e\}$ is the same in both cases. Everything else is different.

(9) a) minsup = 30%



candidate

pruned

(9) b)

Itemset	Support
{A}	5
{B}	7
{C}	5
{D}	9
{E}	6

Itemset	support
{A}	5
{B}	7
{C}	5
{D}	9
{E}	6

C

Itemset	Support
{AB}	3
{AC}	2
{AD}	4
{AE}	4
{BC}	2
{BD}	6
{BE}	4
{CD}	4
{CE}	2
{DE}	6

P

Itemset	Support
{AB}	3
{AD}	4
{AE}	4
{BD}	6
{BE}	4
{CD}	4
{DE}	6

C

Itemset	support
{ABC}	1
{ABD}	2
{ABE}	2
{ACD}	1
{ACE}	1
{ADE}	4
{BCD}	2
{BCE}	1
{BDE}	4
{CDE}	2

p

Itemset	support
{ADE}	4
{BDE}	4

c

Itemset	support
{ABDE}	2
{ACDE}	1
{BCDE}	1

14 frequent items

2^5 → 32 total items

$$\Rightarrow \frac{14}{32} = 0.4375 \Rightarrow \boxed{43.75\ \%}$$