

①

Transaction ID	Items Bought
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}

a) (15%) Compute the support for itemsets {e}, {b, d}, and {b, d, e} by treating each transaction ID as a market basket.

support count σ support $\zeta(x)$

$$\sigma_{\{e\}} = 8$$

$$\zeta(\{e\}) = \frac{8}{10} = 0.8$$

$$\sigma_{\{b,d\}} = 2$$

$$\zeta(\{b,d\}) = \frac{2}{10} = 0.2$$

$$\sigma_{\{b,d,e\}} = 2$$

$$\zeta(\{b,d,e\}) = \frac{2}{10} = 0.2$$

b) (13%) Use the results in part (a) to compute the confidence for the association rules {b, d} \rightarrow {e} and {e} \rightarrow {b, d}. Is confidence a symmetric measure?

Confidence $C(X \rightarrow Y)$

$$C(\{b,d\} \rightarrow \{e\}) = \frac{2}{2} = 1$$

$$C(\{e\} \rightarrow \{b,d\}) = \frac{2}{8} = \frac{1}{4} = 0.25$$

②

{1,2,3}, {1,2,4}, {1,2,5}, {1,3,4}, {1,3,5}, {2,3,4}, {2,3,5}, {3,4,5}.

Assume that there are only five items in the data set.

a) (25%) List all candidate 4-itemsets obtained by the candidate generation procedure in Apriori.

Frequent k -itemset		Frequent k -itemset		Candidate Generation
{1, 2, 3}	X	{1, 2, 3}	=	{1, 2, 3, 4}
{1, 2, 4}		{1, 2, 4}		{1, 2, 3, 5}
{1, 2, 5}		{1, 2, 5}		{1, 2, 4, 5}
{1, 3, 4}		{1, 3, 4}		{1, 3, 4, 5}
{1, 3, 5}		{1, 3, 5}		{2, 3, 4, 5}
{2, 3, 4}		{2, 3, 4}		
{2, 3, 5}		{2, 3, 5}		
{3, 4, 5}		{3, 4, 5}		

$$C_4 \Rightarrow \{1, 2, 3, 4\}, \{1, 2, 3, 5\}, \{1, 2, 4, 5\}, \{1, 3, 4, 5\}, \{2, 3, 4, 5\}$$

*

b) (10%) List all candidate 4-itemsets that survive the candidate pruning step of the Apriori algorithm.

$C_4 \Rightarrow \{1,2,3,4\}, \{1,2,3,5\}, \{1,2,4,5\}, \{1,3,4,5\}, \{2,3,4,5\}$

① $\{1,2,3,4\} \Rightarrow \{1,2,3\} \{1,3,4\} \{1,2,4\} \{2,3,4\}$ are all frequent

② $\{1,2,3,5\} \Rightarrow \{1,2,3\} \{1,2,5\} \{1,3,5\} \{2,3,5\}$ are all frequent

③ $\{1,2,4,5\} \Rightarrow \{1,4,5\} \{2,4,5\}$ are infrequent. Prune $\{1,2,4,5\}$

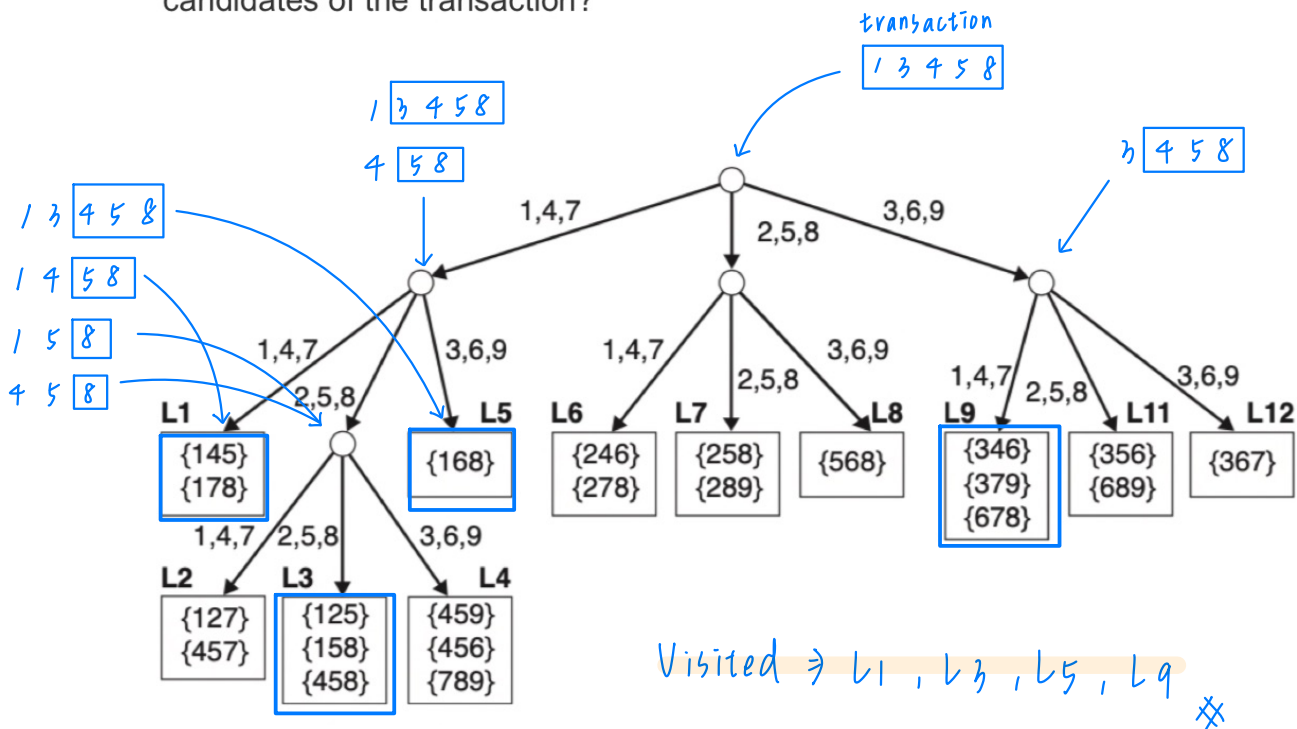
④ $\{1,3,4,5\} \Rightarrow \{1,4,5\}$ is infrequent. Prune $\{1,3,4,5\}$

⑤ $\{2,3,4,5\} \Rightarrow \{2,4,5\}$ is infrequent. Prune $\{2,3,4,5\}$

Update $C_4 \Rightarrow \{1,2,3,4\}, \{1,2,3,5\}$ ✖

3

a) (25%) Given a transaction that contains items $\{1, 3, 4, 5, 8\}$, which of the hash tree leaf nodes will be visited (e.g., L1,...) when finding the candidates of the transaction?



b) (12%) Use the visited leaf nodes in part (a) to determine the candidate itemsets that are contained in the transaction $\{1, 3, 4, 5, 8\}$.

$L1 \Rightarrow \{1,4,5\}$

$L3 \Rightarrow \{1,5,8\} \{4,5,8\}$ ✖