1. Explain why the following algorithm is more efficient than the method for generating association rules from frequent itemsets in section 6.2.2 (section 5.2.2 in 2ed).

Algorithm: Rule_Generator. Given a set of frequent itemsets, output all of its strong rules.

Input:
- \(ls\), set of frequent itemsets;
- \(min\_conf\), the minimum confidence threshold.

Output: Strong rules of itemsets in \(ls\).

Method:
1) for each frequent itemset \(l\) of \(ls\)
2) rule_generator_helper(\(l\), \(l\), \(min\_conf\));

procedure rule_generator_helper
  \((s:\) current subset of \(l\); \(l:\) original frequent itemset; \(min\_conf\))
    (1) \(k = length(s)\);
    (2) if \((k>1)\) then {
        (3) Generate all the \((k-1)\)-subsets of \(s\);
        (4) for each \((k-1)\)-subset \(x\) of \(s\)
            (5) if \((support\_count(l)/support\_count(x) >= min\_conf)\) then {
                (6) output the rule "\(x => (l-x)\)"
                (7) rule_generator_helper(x, l, min\_conf);
                (8) }
        (9) //else do nothing because each subsets of \(x\) will have at //least as much support as \(x\), and hence can never have //greater confidence than \(x\)
    (10)}

2. A database has five transactions. Let \(min\_sup = 60\%\) and \(min\_conf = 80\%\).

<table>
<thead>
<tr>
<th>TID</th>
<th>items_sold</th>
</tr>
</thead>
<tbody>
<tr>
<td>T100</td>
<td>A, B, C, D, E, F</td>
</tr>
<tr>
<td>T200</td>
<td>A, H, S, D, E, T</td>
</tr>
<tr>
<td>T300</td>
<td>A, U, V, D, W, F</td>
</tr>
<tr>
<td>T400</td>
<td>V, B, B, D, E, X</td>
</tr>
<tr>
<td>T500</td>
<td>G, B, C, D, E, F</td>
</tr>
</tbody>
</table>

(a) Find all frequent itemsets using Apriori and FP-growth, respectively. Compare the efficiency of the two mining process.

(b) List all of the strong association rules (with support \(s\) and confidence \(c\)) matching the following metarule, where \(X\) is a variable representing customers, and \(item_i\) denotes variables representing items (e.g., A, B, C):

\[
\forall x \in \text{transaction}, \text{buys}(X, item_1) \land \text{buys}(X, item_2) \Rightarrow \text{buys}(X, item_3)[s, c]
\]

(c) Suppose A is “whole milk” and V is “2% milk”. If we consider both of them as “milk” (M), can we get more rules than (b)? If yes, list them.
3. Some students think that Max patterns must be Closed patterns. The others think that Closed patterns must be Max patterns. How about your opinions? Please show your conclusions.

4. Give a contingency table to show that items in a strong association rule may actually be negatively correlated.

5. Why mining with convertible constraints (e.g., \( \text{avg}(X) \leq 30 \)) is efficient?