CSG230 Fall '08 Quiz 2 Solution

1. DB
   (abc)
   d c (abe) e a
   b (df) b
   c (g) h (ac)

   min_sup = 2
   frequent item:
   a \checkmark
   b \checkmark
   c \checkmark
   d \checkmark
   e \checkmark
   f x
   g x

   DB with frequent items
   (abc)
   d c (ab) a
   b d b
   c b (ac)

   RESULT
   \langle a \rangle /
   \langle b \rangle /
   \langle c \rangle /
   \langle d \rangle /

   \langle ab \rangle /

   \langle ac \rangle /

   \langle ba \rangle /

   \langle ca \rangle /

   \langle cb \rangle /

   \langle cba \rangle /

   \langle db \rangle /
2. (a) 

<table>
<thead>
<tr>
<th>Age</th>
<th>#buy</th>
<th>#not buy</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;30</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>[30,40]</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>&gt;40</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

\[
P(C \mid Age < 30) = \frac{3}{6} \times \frac{3}{6} + \frac{2}{6} = 1
\]

\[
P(C \mid Age \in [30, 40]) = \frac{2}{3} \times \frac{2}{3} + \frac{1}{3} = \frac{2}{3} \times 0.58 + \frac{1}{3} = 0.91
\]

\[
P(C \mid Age > 40) = 0
\]

So \[ P(C \mid Age) = \frac{6}{12} \cdot 1 + \frac{3}{12} \cdot 0.91 = 0.73 \]

<table>
<thead>
<tr>
<th>Credit</th>
<th>#buy</th>
<th>#not buy</th>
</tr>
</thead>
<tbody>
<tr>
<td>good</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>bad</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

\[
P(C \mid credit = good) = \frac{6}{8} \times \frac{4}{3} + \frac{2}{8} = \frac{3}{4} \times (2 - 1.58) + \frac{1}{4} \times 2 = 0.82
\]

\[
P(C \mid credit = bad) = 1
\]

So \[ P(C \mid credit) = \frac{8}{12} \cdot 0.82 + \frac{4}{12} \cdot 1 = 0.88 \]

(b) 

\[
P(buy) = \frac{8}{12} = 0.67
\]

\[
P(age < 30 \mid buy) = \frac{3}{8} = 0.375
\]

\[
P(age < 30 \mid not buy) = \frac{3}{4} = 0.75
\]

\[
P(age > 40 \mid buy) = \frac{3}{8} = 0.375
\]

\[
P(age > 40 \mid not buy) = \frac{0}{4} = 0
\]

\[
P(credit = good \mid buy) = \frac{6}{8} = 0.75
\]

\[
P(credit = good \mid not buy) = \frac{2}{4} = 0.5
\]

\[
P(credit = bad \mid bad) = \frac{2}{8} = 0.25
\]

\[
P(buy) = 0.67
\]

\[
P(age < 30 \mid buy) \times P(credit = good \mid buy) = 0.375 \times 0.75 = 0.28
\]

\[
P(age < 30 \mid not buy) \times P(credit = good \mid not buy) = 0.375 \times 0.5 = 0.19
\]

\[
P(age > 40 \mid buy) \times P(credit = good \mid buy) = 0.375 \times 0.5 = 0.19
\]

\[
P(age > 40 \mid not buy) \times P(credit = good \mid not buy) = 0 \times 0.5 = 0
\]

\[
P(buy) = 0.67
\]

So predict BUY!
4. F = fever, H = headache, C = cough, D = cold

(a) 
\[
\begin{array}{ccc}
\text{F} & \text{C} & 2 \\
\text{H} & \text{D}
\end{array}
\]

(b) We know already \( P(H|F) = 0.25 \), \( P(D|F \land C) = 0.2 \)
we need to find
\[
P(F), \quad P(C), \quad P(H|\sim F), \quad P(D|\sim F \land C), \quad P(D|F \land \sim C), \quad P(D|\sim F \land \sim C)
\]

(c) \[
P(D | H) = \frac{P(D \land H)}{P(H)}
\]
\[
P(D \land H) = P(D \land H \land F \land C) + P(D \land H \land \sim F \land C) + P(D \land H \land F \land \sim C) + P(D \land H \land \sim F \land \sim C)
\]
\[
= P(D | F \land C) \times P(H | F) \times P(F) \times P(C) + P(D | \sim F \land C) \times P(H | \sim F) \times P(\sim F) \times P(C) + P(D | F \land \sim C) \times P(H | F) \times P(F) \times P(\sim C) + P(D | \sim F \land \sim C) \times P(H | \sim F) \times P(\sim F) \times P(\sim C)
\]
\[
P(H) = P(H | F) \times P(F) + P(H | \sim F) \times P(\sim F)
\]

Note: \( P(\sim C) = 1 - P(C) \)
\( P(\sim F) = 1 - P(F) \)
5. (a) $\text{new } CF_1 = \langle 3, (19,3), 286 \rangle$

$$R(\text{new } CF_1) = \frac{1}{2} \sqrt{3 \times 286 - (841 + 9)} = \frac{1}{2} \sqrt{8} \approx 0.94$$

(b) $\text{new } CF_2 = \langle 2, (17,11), 282 \rangle$

$$R(\text{new } CF_2) = \frac{1}{2} \sqrt{2 \times 282 - (171 + 121)} = \frac{1}{2} \sqrt{172}$$

(c) 

| $\langle 3, (19,3), 286 \rangle$ |  
|-----------------------------|---
| $\langle 1, (10,10), 200 \rangle$ |  
