An Example of Making Quantitative Features for Classification from Qualitative Inputs

As an example of the formalism now discussed in the last part of Module 28, consider the following case where 3 binary variables $x_1, x_2, x_3$ taking values in $\{a, b\}$ make up $x^C$, a qualitative part of the input vector for classification in a $K = 3$ class problem. For sake of illustration we'll use the 3 conditional distributions for $x^C \mid y$ with probability mass functions specified in the tables below. (Note that of course the qualitative vector $x^C$ has $M = 8$ possible values and these could be represented in linear models fashion by $M = 7$ (quantitative) dummy variables.)

For $y = 1$:

$$x_3 = a \quad x_3 = b$$

\[
\begin{array}{c|cc}
   & a & b \\
\hline
x_1 & \frac{3}{16} & \frac{1}{16} \\
   a & \frac{1}{16} & \frac{3}{16} \\
   b & \frac{3}{16} & \frac{1}{16} \\
\end{array}
\]

For $y = 2$:

$$x_3 = a \quad x_3 = b$$

\[
\begin{array}{c|cc}
   & a & b \\
\hline
x_1 & \frac{1}{16} & \frac{3}{16} \\
   a & \frac{3}{16} & \frac{1}{16} \\
   b & \frac{1}{16} & \frac{3}{16} \\
\end{array}
\]

For $y = 3$:

$$x_3 = a \quad x_3 = b$$

\[
\begin{array}{c|cc}
   & a & b \\
\hline
x_1 & \frac{1}{16} & \frac{1}{16} \\
   a & \frac{3}{16} & \frac{3}{16} \\
   b & \frac{3}{16} & \frac{3}{16} \\
\end{array}
\]

These are 3 distributions over the 8 (qualitative) vectors in $\{a, b\}^3$. The development in the last section of Module 28 suggests the creation of $K - 1 = 3 - 1 = 2$ real-valued features/statistics $l_1(x^C)$ and $l_2(x^C)$ that are obtained by making ratios of (class-conditional) probabilities for

1. $y = 1$ and $y = 3$ for $l_1$, and then
2. \( y = 2 \) and \( y = 3 \) for \( l_2 \).

Values for these two features/statistics are given below in two forms.

First, in two tables made by dividing values in corresponding cells of pairs of tables above we have:

**Values of \( l_1 \):**

\[
\begin{array}{c|cc}
 x_3 = a & x_3 = b \\
\hline
 x_2 & a & b \\
 a & 3 & 1 \\
 b & 1 & 3 \\
\end{array}
\]

**Values of \( l_2 \):**

\[
\begin{array}{c|cc}
 x_3 = a & x_3 = b \\
\hline
 x_2 & a & b \\
 a & 1/3 & 1 \\
 b & 1 & 1/3 \\
\end{array}
\]

Then, in a single table listing all 8 values of \( x^c \) and then \( l_1(\mathbf{x}^c) \) and \( l_2(\mathbf{x}^c) \) and their logarithms, we have:

<table>
<thead>
<tr>
<th>( x^c )</th>
<th>( l_1(\mathbf{x}^c) )</th>
<th>( l_2(\mathbf{x}^c) )</th>
<th>( \ln l_1(\mathbf{x}^c) )</th>
<th>( \ln l_2(\mathbf{x}^c) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>((a,a,a))</td>
<td>3</td>
<td>1</td>
<td>1.099</td>
<td>0</td>
</tr>
<tr>
<td>((b,a,a))</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1.099</td>
</tr>
<tr>
<td>((a,b,a))</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1.099</td>
</tr>
<tr>
<td>((b,b,a))</td>
<td>3</td>
<td>1</td>
<td>1.099</td>
<td>0</td>
</tr>
<tr>
<td>((a,a,b))</td>
<td>1/3</td>
<td>1</td>
<td>−1.099</td>
<td>0</td>
</tr>
<tr>
<td>((b,a,b))</td>
<td>1</td>
<td>1/3</td>
<td>0</td>
<td>−1.099</td>
</tr>
<tr>
<td>((a,b,b))</td>
<td>1</td>
<td>1/3</td>
<td>0</td>
<td>−1.099</td>
</tr>
<tr>
<td>((b,b,b))</td>
<td>1/3</td>
<td>1</td>
<td>−1.099</td>
<td>0</td>
</tr>
</tbody>
</table>