Softmax classifier

- A generalization of the binary form of Logistic Regression
- Can be applied for multi-label classification
- Widely used in Deep Learning
Softmax classifier

- In Logistic regression, the score $y = Xb$ is not normalized.
- In Logistic regression, $p(y|X) = \frac{1}{1+e^y}$
- In Software classifier, $p(y = k|X) = \frac{e^{yk}}{\sum_j e^{yj}}$
## Softmax classifier

<table>
<thead>
<tr>
<th>Label</th>
<th>y</th>
<th>Exp(y)</th>
<th>Normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>3.2</td>
<td>24.5</td>
<td>0.13</td>
</tr>
<tr>
<td>C2</td>
<td>5.1</td>
<td>164.0</td>
<td>0.87</td>
</tr>
<tr>
<td>C3</td>
<td>-1.7</td>
<td>0.18</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Likelihood Function

- Training data: \( D(X_i, y_i), 1 \leq i \leq n \)

\[
L(b; X) = \prod_{i=1}^{n} p(y_i | X_i)
\]

Take a logarithmic function

\[
\sum_{i=1}^{n} \log(p(y_i | X_i))
\]