Logistic Regression

1. Let \( Y \) - binary response \( \text{ (given explanatory variable x) } \)

\[ \begin{align*}
1 & \quad - \text{ success} \\
0 & \quad - \text{ failure}
\end{align*} \]

\[ E(Y) = 1 \times P(Y=1) + 0 \times P(Y=0) = P(Y=1) = \pi(x) \]

\[ E(Y^2) = 1^2 \cdot \pi(x) + 0^2 \cdot (1-\pi(x)) = \pi(x) \]

\[ \operatorname{Var}(Y) = E(Y^2) - [E(Y)]^2 = \pi(x) - \pi(x) = \pi(x)(1-\pi(x)) \]

Consider a single explanatory variable \( x \).

\[ E(Y) = \pi(x) = \alpha + \beta x \] - a linear probability model
Logistic Regression model

2. Monotonic relationship \( x \) \& \( \pi(x) \)

\[
\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}
\]

Since

\[
\pi(x) = \frac{e^{\beta x}}{(e^x + e^{-\beta x})^\alpha}
\]

If \( \beta > 0 \), \( \pi(x) = \frac{e^x}{e^x + e^{-\beta x}} \rightarrow \}
\begin{align*}
0, & \quad x \to -\infty \\
1, & \quad x \to +\infty
\end{align*}

\[x \to +\infty = e^{\beta x} \to +\infty\]

\[x \to -\infty = e^{-\beta x} \downarrow 0\]
\[
\frac{\partial \pi(x)}{\partial x} = \beta \pi(x) [1 - \pi(x)]
\]

The steepest slope is at \( x = -\frac{\alpha}{\beta} \), when \( \pi(x) = \frac{1}{2} \).

i.e. when \( \alpha + \beta x = 0 \)

The link function when the Logistic Regression model is GLM.

The odds of making \( Y = 1 \)

\[
\frac{\pi(x)}{1 - \pi(x)} = e^{\alpha + \beta x} = e^{\alpha} (e^{\beta})^x
\]

Remark 1. The odds increase by \( e^{\beta} \) for every unit increase in \( x \).

Log odds has the linear relationship

\[
\log \frac{\pi(x)}{1 - \pi(x)} = \alpha + \beta x.
\]

The appropriate link is logit. The log odds transformation.
It is easy to see that

\[ \frac{d \pi(x)}{dx} = \beta \pi(x) \left[ 1 - \pi(x) \right] \]

and the steepest slope is at \( x = -\frac{2}{\beta} \)

i.e. when \( \pi(x) = \frac{1}{2} \) or \( \alpha + \beta x = 0 \)

Indeed:

\[ \frac{d \pi(x)}{dx} = \frac{e^{\alpha + \beta x} \beta (1 + e^{\alpha + \beta x}) - e^{\alpha + \beta x} \beta}{(1 + e^{\alpha + \beta x})^2} \]

\[ = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}} \cdot \frac{\beta}{1 + e^{\alpha + \beta x}} = \]

\[ = \beta \pi(x) \left[ 1 - \pi(x) \right] \]

\[ \max_x \pi(x) (1 - \pi(x)) = \frac{1}{4} \text{ when } \pi(x) = \frac{1}{2} \]

i.e. \( e^{\alpha + \beta x} = 1 \Rightarrow \alpha + \beta x = 0 \)

\[ x = -\frac{2}{\beta} \]
Binomial GLM for 2x2 table

Let us assume that

\[ Y = 1 \quad \text{or} \quad Y = 2 \]

<table>
<thead>
<tr>
<th></th>
<th>( \pi(0) )</th>
<th>( 1 - \pi(0) )</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x = 0 )</td>
<td>( \pi(0) )</td>
<td>( 1 - \pi(0) )</td>
<td>1</td>
</tr>
<tr>
<td>( x = 1 )</td>
<td>( \pi(1) )</td>
<td>( 1 - \pi(1) )</td>
<td>1</td>
</tr>
</tbody>
</table>

Given link function \( \text{link}(\cdot) = (\cdot) \)

GLM is

\[ \text{link} \left[ \pi(x) \right] = \alpha + \beta x \]

The effect of \( x \) on \( Y \) is

\[ \beta = \text{link} \left[ \pi(1) \right] - \text{link} \left[ \pi(0) \right] \]

Example: a) Identity link: \( \beta = \pi(1) - \pi(0) \)

b) Log link:

\[ \beta = \log \frac{\pi(1)}{\pi(0)} = \log \frac{\pi(1)}{\pi(0)} \]

i.e. \( \beta \) is log relative risk.
c) \( \text{link} = \logit \)

\[
\beta = \logit[\pi(1)] - \logit[\pi(0)] = \\
= \log \frac{\pi(1)}{1-\pi(1)} - \log \frac{\pi(0)}{1-\pi(0)} = \\
= \log \frac{\pi(1)}{1-\pi(1)} \cdot \frac{\pi(0)}{1-\pi(0)}
\]

i.e. \( \beta \) is the log Odds Ratio.

and Measure of association for 2x2 tables are effect parameters in GLMs for binary data.