Solution Set: Logistic Regression

1. a) The log-likelihood function $L(\beta)$ is given by

$$L(\beta) = \sum_{i=1}^{N} [y_i \beta^T z_i' - \log(1 + e^{\beta^T z_i'})]$$

$$\begin{aligned} = &-\log\left(1 + e^{\beta_0+\beta_1+2\beta_2}\right) - \log\left(1 + e^{\beta_0+2\beta_1+\beta_2}\right) \\ &+\beta_0 + 2\beta_1 + 3\beta_2 - \log\left(1 + e^{\beta_0+2\beta_1+3\beta_2}\right) \\ &+\beta_0 + 3\beta_1 + 2\beta_2 - \log\left(1 + e^{\beta_0+3\beta_1+2\beta_2}\right) \\ &+\beta_0 + \beta_1 + \beta_2 - \log\left(1 + e^{\beta_0+\beta_1+\beta_2}\right) \end{aligned}$$

b) In iterative reweighted least squares, we pick an initial value $\beta^{(0)}$ and update $\beta^{(t)}$ by

$$\beta^{(t+1)} = (Z^T W Z)^{-1} Z^T W \boldsymbol{v} \qquad \text{where}$$

$$Z = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 2 \\ 1 & 1 & 1 \end{bmatrix}, \boldsymbol{y} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{p} = \begin{bmatrix} p(z_1;\beta^{(t)}) \\ . \\ . \\ . \\ p(z_5;\beta^{(t)}) \end{bmatrix},$$

$$W = \begin{bmatrix} p(z_1;\beta^{(t)})(1-p(z_1;\beta^{(t)})) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & p(z_5;\beta^{(t)})(1-p(z_5;\beta^{(t)})) \end{bmatrix},$$

and $v = Z\beta^{(t)} + W^{-1}(\boldsymbol{y} - \boldsymbol{p})$.

Recall that $p\left(z_i;\beta^{(t)}\right) = \frac{e^{\left(\beta^{(t)}\right)^T z_i'}}{1+e^{\left(\beta^{(t)}\right)^T z_i'}}$.

We'll pick $\mathbf{0}$ as the initial value $\beta^{(0)}$.

$$\text{Then, } \boldsymbol{p} = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}, W = \begin{bmatrix} 1/4 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/4 \end{bmatrix}, \boldsymbol{v} = \begin{bmatrix} -2 \\ -2 \\ 2 \\ 2 \\ 2 \end{bmatrix}$$

$$\Longrightarrow \qquad \beta^{(1)} = \begin{bmatrix} -2 \\ 2/3 \\ 2/3 \end{bmatrix} \approx \begin{bmatrix} -2 \\ 0.667 \\ 0.667 \end{bmatrix}$$

We update $\boldsymbol{p}, W, \boldsymbol{v}$ and calculate $\beta^{(2)}$.

$$\beta^{(2)} \approx \begin{bmatrix} -2.28 \\ 0.77 \\ 0.77 \end{bmatrix}.$$

If we keep iterating,we get

$$\beta^{(3)} \approx \begin{bmatrix} -2.3 \\ 0.778 \\ 0.778 \end{bmatrix}$$

$$\beta^{(4)} \approx \begin{bmatrix} -2.3 \\ 0.778 \\ 0.778 \end{bmatrix}$$

$\beta^{(5)}$ and $\beta^{(6)}$ are nearly the same as $\beta^{(4)}$. So, $\beta^{(t)}$ converges to $\begin{bmatrix} -2.3 \\ 0.778 \\ 0.778 \end{bmatrix}$.

The estimates for $\beta_0, \beta_1, \beta_2$ are $\widehat{\beta_0} = -2.3, \widehat{\beta_1} = 0.778, \widehat{\beta_2} = 0.778$.

c) The estimated probability function $\hat{p}(x)$is given by $\hat{p}(x) = \frac{e^{\widehat{\beta_0}+\widehat{\beta_1}x_1+\widehat{\beta_2}x_2}}{1+e^{\widehat{\beta_0}+\widehat{\beta_1}x_1+\widehat{\beta_2}x_2}}$.

So $\hat{p}(x) = \frac{e^{-2.3+0.778x_1+0.778x_2}}{1+e^{-2.3+0.778x_1+0.778x_2}}$.

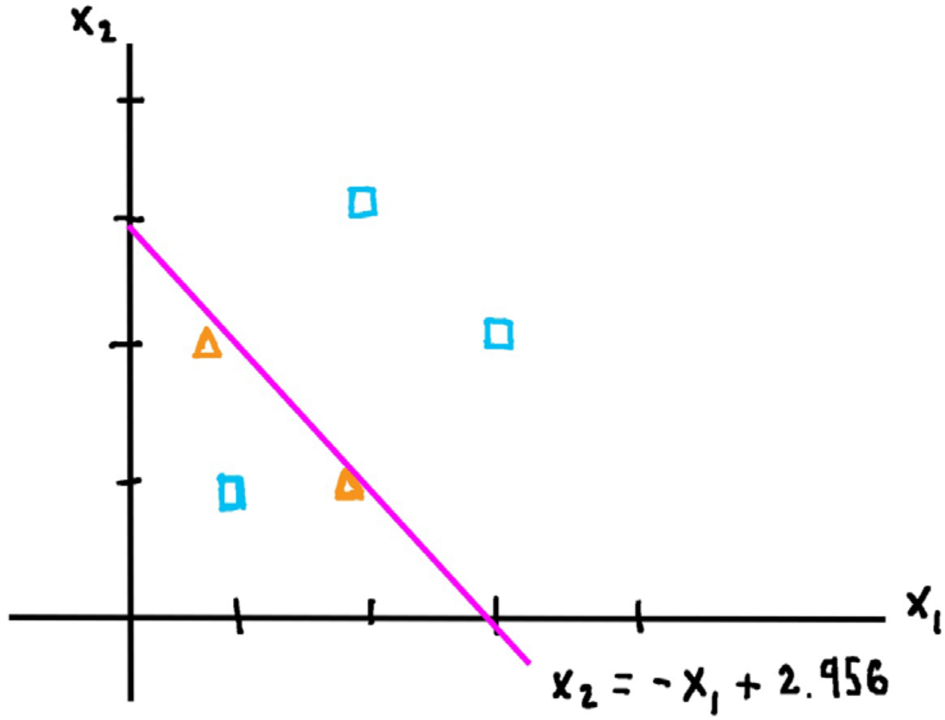
d) $\hat{p}(1.5, 1) = 0.412$. We classify x as of class 1 if $\hat{p}(x) > 1/2$ and as of class 0 if $\hat{p}(x) < 1/2$.

Therefore, we classify $(1.5, 1)$ as of class 0. The decision boundary is given by

$$-2.3 + 0.778x_1 + 0.778x_2 = 0.$$

This is the line $x_2 = -x_1 + 2.956$.

Here is what it looks like with the data points:



$x_2 > -x_1 + 2.956$ corresponds to $\hat{p}(x) > 1/2$, and

$x_2 < -x_1 + 2.956$ corresponds to $\hat{p}(x) < 1/2$.