2. GENERALIZED LINEAR MODELS:

(a) Let W be the weights diagonal Matrix.

Using Bianomial theorum,

$$b(y_i; m_i, w_i) = \binom{m_i}{y_i} w_i^{y_i} (1-w_i)^{(m_i-y_i)}$$

Negative log of above function is

$$\ell(\beta) = -\log\left\{\prod_{i=1}^{N} P(y_i/\beta)\right\}$$

$$= -\log\left\{\prod_{i=1}^{N}\binom{m_i}{y_i} w_i^{y_i}(1-w_i)^{(m_i-y_i)}\right\}.$$

$$= -\sum_{i=1}^{N}\left\{\log\binom{m_i}{y_i} + y_i\log w_i + (m_i-y_i)\log(1-$$

Using differential calculus:

$$\nabla\ell(\beta) = \nabla_\beta\left(-\sum_{i=1}^{N}\left\{\log\binom{m_i}{y_i} + y_i\log w_i(\beta) + (m_i-y_i)\log(1-w_i(\beta)\right.\right.$$

$$= -\sum_{i=1}^{N}\left\{0 + y_i\nabla_\beta(\log w_i(\beta)) + (m_i-y_i)\nabla_\beta(\log(1-w_i(\beta)))\right\}$$

$$= -\sum_{i=1}^{N}\left\{\frac{y_i}{w_i(\beta)}\nabla_\beta(w_i(\beta)) - \frac{m_i-y_i}{1-w_i(\beta)}\nabla_\beta(w_i(\beta))\right\}.$$

$$w_i = \frac{1}{1+\exp(\mu)} = \frac{1}{V} \qquad \forall\ \mu = -x_i^T\beta$$

$$V = 1+\exp\{\mu\}$$

$$\nabla_\beta (w_i(\beta)) = \frac{-\frac{dV}{d\beta}}{V^2} = \frac{-\frac{dU}{d\beta} \exp(u)}{(1+\exp(u))^2} = \frac{x_i^T \exp\{-x_i^T\beta\}}{(1+\exp\{-x_i^T\beta\})^2}$$

$$\frac{dV}{d\beta} = \frac{dU}{d\beta} \exp\{u\}, \quad \frac{dU}{d\beta} = -x_i^T$$

$$\nabla_\beta (w_i(\beta)) = \left(\frac{1}{1+\exp\{-x_i^T\beta\}}\right)\left(\frac{\exp\{-x_i^T\beta\}}{1+\exp\{-x_i^T\beta\}}\right)(x_i^T)$$

$$= w_i \left(\frac{\exp\{-x_i^T\beta\}}{1+\exp\{-x_i^T\beta\}}\right)(x_i^T)$$

$$= w_i \left(1 - \frac{1}{1+\exp\{-x_i^T\beta\}}\right)(x_i^T)$$

$$= w_i(1-w_i)x_i^T$$

$$= (w_i(1-w_i)x_i^T)$$

$$\Rightarrow \nabla l(\beta) = -\sum_{i=1}^N \left\{ \frac{y_i}{w_i} \nabla_\beta (u_i) - \frac{m_i - y_i}{1-w_i} \cdot (w_i(1-w_i)x_i) \right\}$$

$$= -\sum_{i=1}^N \{ y_i(1-w_i)x_i - (m_i - y_i)w_i x_i \}$$

$$= -\sum_{i=1}^N \{ y_i x_i - y_i w_i x_i - m_i w_i x_i + y_i w_i x_i \}.$$

$$= -\sum_{i=1}^N \{ (y_i - m_i w_i) x_i \} = \sum_{i=1}^N \{ (m_i w_i - y_i) x_i \}.$$

$$\Rightarrow \boxed{\nabla l(\beta) = X^T(mw - y)}$$

Completing the square,

$$a + b^T x + \frac{1}{2} x^T C x = \frac{1}{2}(x+m)^T C (x+m) + v$$

$$m = C^{-1} b$$

$$v = a - \frac{1}{2} b^T C^{-1} b$$

$$a = l(\beta_0), \quad b = m\omega - y$$

$$x = X\beta - X\beta_0$$

$$C = W$$

$\Rightarrow$ $$m = W^{-1}(m\omega - y)$$

$$v = l(\beta_0) - \frac{1}{2}(m\omega - y)^T W^{-1}(m\omega - y)$$

$$l(\beta) = \frac{1}{2}(x+m)^T C (x+m) + V$$

$\Rightarrow$ $$l(\beta) = \frac{1}{2}\left((X\beta - X\beta_0) + W^{-1}(m\omega - y)\right)^T W \left((X\beta - X\beta_0) + W^{-1}(m\omega - y)\right) + l(\beta_0) - \frac{1}{2}(m\omega - y)^T W^{-1}(m\omega - y)$$

comparing this to the required form :-

$$\boxed{\begin{aligned} Z &= X\beta_0 + W^{-1}(y - m\omega) \\ W &= \sum m_i \omega_i (1-\omega_i) \end{aligned}}$$

$\cancel{\forall}$ $$q(\beta:\beta_0) = \frac{1}{2}(z - X\beta)^T W^{-1}(m\omega - y)$$

1. (C) Second order approximation is given by:

$$f(x_k + p) \approx f_k + p^T \nabla f_k + \frac{1}{2} p^T \nabla^2 f_k p$$

we need to find second order approximation $l(\beta)$ around β_0 and prove that $l(\beta) = Q(\beta; \beta_0)$

$\Rightarrow x_k = \beta_0$ & $p = \beta - \beta_0$

$$l(\beta) = l(\beta_0) + (\beta - \beta_0)^T \nabla l(\beta_0) + \frac{1}{2} (\beta - \beta_0)^T \nabla^2 l(\beta_0)(\beta - \beta_0)$$

$$\nabla l(\beta) = \sum_{i=1}^{N} \{ (m_i w_i - y_i) x_i \} = X^T (mw - y)$$

$$\nabla^2 l(\beta) = \nabla (\nabla l(\beta)) = \nabla \left(\sum_{i=1}^{N} \{ (m_i w_i - y_i) x_i^T \} \right)$$

$$= \sum_{i=1}^{N} \{ m_i \nabla_\beta (w_i) x_i^T - 0 \}$$

$$\because \nabla_\beta (w_i) = w_i (1 - w_i) x_i$$

$$\Rightarrow \boxed{\nabla^2 l(\beta) = \sum_{i=1}^{N} m_i w_i (1 - w_i) x_i x_i^T = \boxed{X^T W X}}$$

where $W = m_i w_i (1 - w_i)$ is the diagonal weights matrix.

$$\Rightarrow l(\beta) = l(\beta_0) + (mw - y)^T X (\beta - \beta_0) + \frac{1}{2} (\beta - \beta_0)^T X^T W X (\beta - \beta_0)$$