(4) The error function $L(w)$ is

$$L(w) = \sum_{i=1}^{N} \left[y^{(i)} \log\left(h_w(x^{(i)})\right) + (1-y^{(i)}) \log\left(1-h_w(x^{(i)})\right) \right]$$

$y^{(i)}$ - actual output

$x^{(i)}$ - input

$$h_w(x^{(i)}) = \frac{1}{1+e^{w^T x^{(i)}}}$$

$$\nabla L(w) = \sum_{i=1}^{N} \left[y^{(i)} \frac{\partial}{\partial w} \log\left(h_w(x^{(i)})\right) + (1-y^{(i)}) \frac{\partial}{\partial w} \left(\log(1-h_w(x^{(i)})\right) \right]$$

$$= \sum_{i=1}^{N} \left[y_i \frac{e^{-w^T x^{(i)}}}{1+e^{-w^T x^{(i)}}} x^{(i)} - (1-y^{(i)}) \cdot \frac{(1+e^{-w^T x^{(i)}})}{(1+e^{-w^T x^{(i)}})^2} \frac{e^{-w^T x^{(i)}}}{e^{-w^T x^{(i)}}} x^{(i)} \right]$$

$$= \sum_{i=1}^{N} \left[y_i (1-h_w(x^{(i)})) x^{(i)} - (1-y^{(i)}) \frac{x^{(i)}}{1+e^{-w^T x^{(i)}}} \right]$$

$$\cancel{= \sum_{i=1}^{N} (y_i - h_w)}$$

$$= \sum_{i=1}^{N} \left(y_i - h_w(x^{(i)})\right) x^{(i)}$$

$$\boxed{\nabla L(w) = X^T (y-h)}$$; gradient

Hessian, $\nabla(\nabla L(w)) = \frac{\partial}{\partial w}\left(\sum_{i=1}^{N}\left(y_i - \frac{1}{1+e^{-w^T x^{(i)}}}\right) x^{(i)}\right)$

$$= -x^T \sum_{i=1}^{N} \frac{1+e^{-w^T x^{(i)}} \left[y_i e^{-w^T x^{(i)}} \cdot (-x^{(i)})\right] - \left(y_i + y_i e^{-w^T x^{(i)}} - 1\right) \cdot}{1+e^{-w^T x^{(i)}}}$$

on solving ;

$$\nabla\nabla L(w) = \sum_{i=1}^{N} \frac{e^{-w^T x^{(i)}} \cdot x^{(i)} \left(x^{(i)}\right)^T}{\left(1+e^{-w^T x^{(i)}}\right)^2} \quad ; \text{ hessian}$$

$$= \sum \left(\frac{e^{-w^T x^{(i)}}}{1+e^{-w^T x^{(i)}}}\right)\left(\frac{1}{1+e^{-w^T x^{(i)}}}\right) x^{(i)} \left(x^{(i)}\right)^T$$

$$= \sum_{i=1}^{N} (1-h_w(x))\, h_w(x)\, x^{(i)} x^{(i)T}$$

$$\therefore \boxed{H(w) = X^T A X \; ;} \quad A = (1-h_w(x))(h_w(x))$$

Hessian

$n - H^{-1} (Lw$

Tylor series expansion of first two terms

$$f(w) = f + (\nabla L(w))^T (w - W_{old}) + \frac{1}{2}(w - W_{old})^T H . (w - W_{old})$$

Simply in matrix notation,

$$f(w) = w^T P w + w^T Q + C$$

where $P = \frac{1}{2}H$, $b = [\nabla L(w)] - H . W_{old}$

$$C = f + (\nabla L(w))^T W_{old} + \frac{1}{2}(W_{old})^T H (W_{old})$$

$$f'(w) = 2Pw + 2Q$$

$$w^* = -\frac{1}{2} P^{-1} Q$$

$$w^* = -\frac{1}{2}(2H^{-1}) . (\nabla L(w) - H W_{old})$$

$$\therefore \boxed{w_{new} = W_{old} - H^{-1} \nabla L(w)}$$ updated weight

$$w_{new} = W_{old} - (X^T A X)^{-1} (X^T (y - h))$$

$$= (X^T A X)^{-1} [(X^T A X) W_{old} - X^T (y - h)]$$

$$\boxed{w_{new} = (X^T A X)^{-1} X^T A [X W_{old} - A^{-1} (y - h)]}$$; $A = (1 - h_w(x^{(i)})) . h_w(x^{(i)})$

Observe that this looks like W_A after weighted least square