Q4 a)

→ The likelyhood function of logistic regression can be written as

$$P(t/\omega) = \prod_{n=1}^{n} y_n^{t_n} (1-y_n)^{1-t_n}$$

where data set $\{\phi_n, t_n\}$, $t_n \in \{0,1\}$ &

$\phi_n = \cancel{\phi(x_n)}\ \phi(x_n)$ with $n = 1, 2 \ldots n$.

By taking negative likelyhood, we will get the error function.

$$E(\omega) = -\sum_{n=1}^{n}\left[t_n \log(y_n) + (1-t_n)\log(1-y_n)\right]$$

where $y_n = \sigma(\omega^T \cancel{\phi_n})$

$$E(\omega) = -\sum_{n=1}^{n}\left[t_n \log\left(\frac{1}{1+e^{-\omega^T x}}\right) + (1-t_n)\log\left(1-\frac{1}{1+e^{-\omega^T x}}\right)\right]$$

$$= -\sum_{n=1}^{n}\left[t_n\left[\log\left(\frac{1}{1+e^{-\omega^T x}}\right) - \log\left(1-\frac{1}{1+e^{-\omega^T x}}\right)\right] + \log\left(1-\frac{1}{1+e^{-\omega^T x}}\right)\right]$$

$$= -\sum_{n=1}^{n}\left[t_n \log e^{\omega^T x} + \log\left(\frac{e^{-\omega^T x}}{1+e^{-\omega^T x}}\right)\right]$$

$$= -\sum_{n=1}^{n}\left[t_n \omega^T x + \log\left(\frac{1}{1+e^{\omega^T x}}\right)\right]$$

$$= -\sum_{n=1}^{n}\left[t_n \omega^T x - \log\left(1+e^{\omega^T x}\right)\right]$$

$$E(\omega) = \sum_{n=1}^{n}\left[\log\left(1+e^{\omega^T x}\right) - t_n \omega^T x\right]$$

where $x = \cancel{\phi} \phi_n$

Gradient :

It is 1^{st} order derivative of error function

$$\therefore \nabla E = \frac{\partial}{\partial \omega}\left[\sum_{n=1}^{n} \log(1+e^{\omega^T x}) - t_n \omega^T x\right]$$

$$= \sum_{n=1}^{n} \frac{x \cdot e^{\omega^T x}}{1+e^{\omega^T x}} - t_n x$$

$$= \sum_{n=1}^{n} \left[\frac{1}{1+e^{-\omega^T x}} - t_n\right] x$$

$$= \sum_{n=1}^{n} (y_n - t_n)\, \phi_n .$$

Hessian :

Hessian is double derivative of error function or ~~of~~ derivative of gradient

$$\nabla E(\omega) = \sum_{n=1}^{n} (y_n - t_n)\, \phi_n$$

$$= \Phi^T (y - t) .$$

where Φ is $N \times m$ design matrix whose n^{th} row is given by ϕ_n^T .

$$\therefore \text{Hessian}(H) = \nabla\nabla E(\omega) \frac{\partial^2}{\partial \omega} \sum \left[\frac{1}{1+e^{-\omega^T x}} - t_n\right] x$$

$$= \sum_{n=1}^{n} \left(\frac{1}{1+e^{\omega^T x}}\right)^2 e^{-\omega^T x} . x . x$$

$$= \sum_{n=1}^{n} \left(\frac{e^{-\omega^T x}}{1+e^{\omega^T x}}\right)\left(\frac{1}{1+e^{\omega^T x}}\right) x^T x \qquad (\because x^2 = x^T x)$$

$$= \sum_{n=1}^{n} (1 - y_n)\, y_n\, x^T x .$$

where $x = \phi_n$

$\therefore H = \Phi^T R \Phi$ where $R_{nn} = y_n(1-y_n)$

Update equation:

Newton-Raphson gives update equation as

$$w_{new} = w_{old} - H^{-1} \nabla E$$

$$w_{new} = w_{old} - (\Phi^T R \Phi)^{-1} \Phi^T (y-t)$$

$$= (\Phi^T R \Phi)^{-1} \{ \Phi^T R \Phi \, w_{old} - \Phi^T (y-t) \}$$

$$= (\Phi^T R \Phi)^{-1} \Phi^T R z$$

where $z = \Phi w_{old} - R^{-1}(y-t)$.

Algorithm

1) Initialize w, with any value (w_0)

2) for $i = 1$ to ...:

   i) Calculate $E(w)$ with quadratic function $\bar{E}(w)$ using w_0 (i.e calculate gradient).

   ii) Calculate the hessian.

   iii) Update the value of w.

3) stop when $|w_{new} - w_{old}| \approx 0$

4) we got $\frac{\partial}{\partial w} E(w) = 0$.

Q4 b)

→ We know that,

Gradient of logistic regression.

$$\nabla E(w) = \phi^T (y - t)$$

& Hessian $H = \phi^T R \phi$ where $R_{nn} = y_n(1 - y_n)$

∴ update equation:

$$w_{new} = w_{old} - (\phi^T R \phi)^{-1} \phi^T (y - t)$$

$$= (\phi^T R \phi)^{-1} [\phi^T R \phi \, w_{old} - \phi^T (y - t)]$$

$$= (\phi^T R \phi)^{-1} \times \phi R z$$

where $z = \phi w_{old} - R^{-1}(y - t)$

~~This eq~~

we can see the R is not constant & is calculated again & again with each iteration.

For this reason, algorithm is called iterative reweighted least squares

Q4 c

→ We can say that the error function $E(w)$ is convex when its hessian is positive definite

i.e $\forall w \neq 0 \in \mathbb{R}^m$: $w^T H w > 0$.

We know that,

$$H = \Phi^T R \Phi, \quad \text{where } R_{nn} = y_n(1-y_n)$$

$$\therefore \quad w^T H w = w^T \Phi^T R \Phi w$$

As, $0 < y_n < 1$, therefore we can write

$$R = R^{1/2} . R^{1/2} .$$

where $R^{1/2} = \text{diag}\left(\sqrt{y_n(1-y_n)}\right)$

$$\therefore \quad w^T \Phi^T R \Phi w \Rightarrow w^T \Phi^T R^{1/2} R^{1/2} \Phi^T w$$

$$= \left(R^{1/2} \Phi^T w\right)^T R^{1/2} \Phi^T w$$

$$= \| R^{1/2} \Phi^T w \|^2 \quad (\because x^2 = x^T x)$$

which will be always positive.

$\therefore$ Error function $E(w)$ is convex.