(4).

ML Estimate $E(w) = -\ln(L)$.

$$E_n(w) = -\left(t_n \ln(\sigma(w^T x_n)) + (1-t_n)\ln(1-\sigma(w^T x_n))\right).$$

$$\frac{d}{dz}\sigma(z) = \sigma(z)(1-\sigma(z)).$$

Where,

t_n = Observed target data point n.

$\sigma(x)$ = logistic sigmoid function

x_n = Input feature vector

(a).

(i) Gradient :-

The Gradient Error function $E(w)$ to the parameter vector w in logistic regression is given by.

$$\nabla E_n(w) = \left[\frac{\partial E_n(w)}{\partial w_1}, \frac{\partial E_n(w)}{\partial w_2}, \ldots, \frac{\partial E_n(w)}{\partial w_d}\right]$$

$$\frac{\partial}{\partial w_i}\left(t_n \ln(\sigma(w^T x_n))\right) = t_n \frac{1}{\sigma(w^T x_n)}\sigma(w^T x_n)(1-\sigma(w^T x_n))x_{n,i}.$$

$$\frac{\partial}{\partial w_i}\left((1-t_n)\ln(1-\sigma(w^T x_n))\right) = (1-t_n)\frac{1}{1-\sigma(w^T x_n)}(-\sigma(w^T x_n)(1-\sigma(w^T x_n))x_{n,i}).$$

$$\nabla E_n(w) = -\left(t_n(1-\sigma(w^T x_n)) - (1-t_n)\sigma(w^T x_n)\right)x_n.$$

$$= -(t_n - \sigma(w^T x_n))x_n$$

$$\nabla E(w) = -\sum_{n=1}^{N}(t_n - \sigma(w^T x_n))x_n = X^T(y-T).$$

T, y, X are vectors of size N of datapoints t, $\sigma(w^T x_i)$, x respectively.

Hessian Matrix :-

The Error function $E(w)$ is given by

$$H(w) = \frac{\partial}{\partial w} E(w).$$

From $\frac{\partial}{\partial z} \sigma(z) = \sigma(z)(1-\sigma(z))$ and,

$$\nabla E(w) = -\sum_{n=1}^{N} (t_n - \sigma(w^T x_n)) x_n = X^T(y-T).$$

$$H(w) = \sum_{n=1}^{N} \sigma(w^T x_n)(1-\sigma(w^T x_n)) x_n x_n^T.$$

$Y = \text{diag}_N(\sigma(w^T x_i)(1-\sigma(w^T x_i)))$ and X is Coloumn Vectors $(N \times 1)$ of input.

Update Equation for the Newton Raphson Optimization Technique :-

$$w^{(t+1)} = w^{(t)} + \nabla w$$

We choose ∇w such that

$E(w^{(t+1)})$ is minimum

$$\frac{\partial}{\partial \nabla w} E(w^{(t+1)}) = \frac{\partial}{\partial \nabla w} E(w^{(t)} + \Delta w) = 0$$

$$\nabla_E(w^{(t)}) + H(w^{(t)}) \nabla w = 0$$

$$\nabla w = -H(w^{(t)})^{-1} \nabla E(w^{(t)})$$

Here Newton - Raphson update scheme iteratively updates the parameter vector w as follows:

$$w^{(t+1)} = w^{(t)} - H(w^{(t)})^{-1} \nabla E(w^{(t)}).$$

Where

parameter vector at $t = w^{(t)}$.

Hessian Matrix at $t = H(w^{(t)})$.

Gradient $t = \nabla E(w^{(t)})$.

The Algorithm Describing the Methodology:-

(i) Initialize the parameter Vector. $(w^{(0)})$.

(ii) Now Repeat Until Convergence

(a) Compute the gradient $\nabla E(w^{(t)})$ and the Hessian Matrix $H(w^{(t)})$.

(b) update the parameter vector using the Newton-Raphson update Equation.

(c) Check for Convergence Criteria.

(iii) return W^*, After Completing the step (2)

(b) similarity with WLS $w^{(t+1)} = w^{(t)} - H(w^{(t)})^{-1} \nabla E(w^{(t)})$

$$= w^{(t)} - (X^T Y X)^{-1} X^T (y - T)$$

$$= (X^T Y X)^{-1} ((X^T Y X) w^{(t)} - X^T (y - T))$$

$$\Rightarrow (X^T Y X)^{-1} (X^T Y (X w^{(t)}) - Y^{-1} (y - T))$$

$$\Rightarrow (X^T Y X)^{-1} X^T Y Z.$$

Here $Z = X w^{(t)} - Y^{-1} (y - T)$ and $W_{ML} = (X^T X)^{-1} X^T t$

We can say that Newton-Raphson update scheme is related to the weighted least squares problem.

During each iteration of the Newton-Raphson algorithm in Logistic regression Based on the gradient and Hessian of the log-likelihood, we update w, In this update phase, The weights are chosen by the logistic sigmoid function, similar to a weighted five least squares update. Additionally, the Newton-Raphson update is inversely proportional to the gradient of the log-likelihood, and the log-likelihood Hessian's function is comparable to that of the accuracy in weighted least squares. The Hessian determines the size and update direction by taking into consideration the log-likelihood functions curvature.

Now, Therefore, logistic regression can be referred to as an Iterative reweighted least squares Approach, Where the logistic sigmoid function determines the weights for each data point, and the parameter vector is updated at each iteration to minimise the Negative log-likelihood giving ambiguous data point more weight. The estimate of the logistic regression parameters is the result of the iterative process, which continuous Until Convergence.

(c)

The Hessian Matrix $H(w)$ of the error function in logistic regression is

$$H(w) = \sum_{n=1}^{N} \sigma(w^T x_n)(1-\sigma(w^T x_n)) x_n x_n^T = X^T Y X$$

For $w \in R^N ! = 0^N$

$$w^T H w = w^T X^T Y X w.$$

Now $(\sigma(w^T x_n)(1-\sigma(w^T x_n)) > 0)$ and Y is a diagonal function $Y = (Y^{\frac{1}{2}})^T Y^{1/2}$ and $Y^{1/2} = diag((\sigma(w^T x_n)(1-\sigma(w^T x_n)))^{1/2})$.

$$w^T H w = w^T X^T Z^T Z X w = \| Z X w \| > 0.$$

$$\therefore Z = Y^{1/2}$$

So, Therefore $W^T H W > 0 \quad \forall w \in R^N$.

Which suggests that $\nabla E(w)$ is rising, Indicating a Convex Error function as since the logistic sigmoid function $\sigma(w^T x_n)$ always lies b/w 0 and 1 and $1-\sigma(w^T x_n)$ is also always b/w 0 and 1 and $x_n x_n^T$ is always finitely positive, and also Here the Hessian function is also always finite positive. So, The Gradient function is increasing.

Which implies $\ln(L)$ is Concave function which shows that $E(w) = -\ln(L)$ is a Convex function.

Therefore, The Convex down error function of logistic regression which says it has an Unique global minimum value and that occurs at W_{ML}.