Q4 a) Logistic regression objective function:

parameter vector

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

log-likelihood function — observed class — predicted probability

Gradient (first derivative)
(wrt parameter vector w)

$$\nabla L(w) = X^T(y-p)$$

↳ gradient of log-likelihood

Hessian (Second derivative)

$$H = -X^T W X$$

W → diagonal matrix with elements $p_i(1-p_i)$

Hessian matrix — design matrix

Newton-Raphson Update equation.

$$w_{new} = w_{old} - (H^{-1} \nabla L(w_{old}))$$

Updated parameter — Current parameter

→ Newton-Raphson Optimization algorithm
(for logistic regression)

1. Initialise parameter vector w. (can be zeroes as well)

2. Repeat the following steps until convergance.

[Convergence: parameter values becomes very small]

i) Compute the predicted probabilities p using the logistic sigmoid function and the current parameter values.

ii) Compute $\nabla L(w)$ & H using equations.

iii) Compute update direction as $\Delta w = H^{-1} \nabla L(w)$

iv) Update w as $w = w - \Delta w$

3. Return the estimated parameter vector w when convergence is reached (or after a certain no. of iterations)

Q4 b) In logistic regression,

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

& Sum of squares error

$$E(w) = \sum_{i=1}^{N} w_i (y_i - \hat{y}_i)^2$$

They are related through the choice of weights.

In logistic regression, the weights w_i are determined by the diagonal elements of the matrix W

$$W = \text{diag}(p_1(1-p_1), p_2(1-p_2), p_3(1-p_3) \ldots p_N(1-p_N))$$

These weights are computed based on the predicted probabilities p_i, and they reflect the uncertainty in model's predictions.

Higher uncertainity (closer to 0.5) $\Leftrightarrow$ lower weights
lower uncertainity (closer to 0 or 1) $\Leftrightarrow$ higher weights.

Since, it iteratively updates the parameter vector w by treating the logistic regression problem as a weighted least squares problem at each iteration, where the weights are determined by the predicted probabilities p_i, this gives the term "iterative reweighted least squares" (IRLS)

Q4 c) Convex function $\Rightarrow$ unique minimum.
(From previous parts)

Error function (negative log likelihood) for logistic regression is

$$J(w) = -\sum_{i=1}^{N} \left[y_i \log(\sigma(w^T x_i)) + (1-y_i) \log(1-\sigma(w^T x_i)) \right]$$

where $\sigma(z) = \frac{1}{1+e^{-z}}$ is sigmoid function.

The elements of Hessian matrix are the second partial derivatives of $L(w)$ wrt w_i & w_j

$$H = \sum_{i=1}^{N} \sigma(w^T x_i)(1-\sigma(w^T x_i)) x_i x^T$$

To check convexity,
A function is convex if & only if its Hessian matrix is negative definite i.e. all its eigen values are negative & it will have a unique minimum

Positive semi definiteness of a matrix M is defined as $V^T M v \geq 0$ $\forall$ non zero vectors V

For the logistic error function, the Hessian positive semi-definiteness is evident from

$$V^T H v = \sum_{i=1}^{N} \sigma(w^T x_i)\left(1-\sigma(w^T x_i)\right)(V^T x_i)^2$$

$\therefore$ This is always non-negative

Since the error function is convex, the local minima is actually global minima.

For logistic regression, Newton Raphson method will converge to the unique global minima of error function.

Update rule :-

$$w_{new} = w_{old} - H^{-1} \nabla J(w)$$

where $\nabla J(w)$ = gradient of error function.

To check convexity,
A function is convex if & only if its Hessian matrix is negative definite i.e. all its eigen values are negative & it will have a unique minimum.