Logistic Regression

- It is used for binary classification problems-
- sigmoid function is used in it

$$\text{sigmoid}(z) = \frac{1}{1+e^{-z}}$$

Step 1:

It start by Combining input functions in single score using linear equation:

$$z = w_1 \cdot x_1 + w_2 \cdot x_2 + \cdots + w_n \cdot x_n + b$$

Step 2: Sigmoid Function

$$P(y=1|x) = \sigma(z) = \frac{1}{1+e^{-z}}$$

e = base of natural log. (≈ 2.718)

$\sigma(z)$ = ensure output is always in range $(0,1)$

Step 3:

$$y = \begin{cases} 1, & \text{if } P \geq t \\ 0, & \text{if } P < t \end{cases}$$

Step 4: Train model, minimize error, error measured in Log-Loss

$$J(w,b) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log(P^{(i)}) + (1-y^{(i)})\log(1-P^{(i)})\right]$$

m = number of training examples
$y^{(i)}$ = Actual label for i-th example (0 or 1)

$P^{(i)} = \sigma(z^{(i)})$ = Predicted probability for i-th eg

Step 5: (Gradient Descent)

$$w_j = w_j - \alpha\frac{\partial J}{\partial w_j}, \quad b = b - \alpha\frac{\partial J}{\partial b}$$

α = learning rate
$\frac{\partial J}{\partial b}$ and $\frac{\partial J}{\partial w_j}$ are gradient of cost function.

Step 6 Final Model:

$$P(y=1|x) = \frac{1}{1+e^{-(w_1 \cdot x_1 + w_2 \cdot x_2 + \cdots + w_n \cdot x_n + b)}}$$