# MATHEMATICS BEHIND BINOMIAL LOGISTIC REGRESSION WITH GRADIENT DESCENT

## 1. PROBLEM SETUP

We have:

- Feature vector: `x`
- Target variable: `y ∈ {0, 1}`
- Parameters: weights `w` and bias `b`

Goal: Model `P(y=1 | x; w, b)` using the logistic function.

## 2. HYPOTHESIS FUNCTION

The linear combination:

$$z = \mathbf{w}^T\mathbf{x} + b$$

The sigmoid function maps `z` to a probability:

$$\&haty = \sigma(z) = 1 / (1 + e^{-z})$$

`&haty` is the predicted probability that `y=1`. The sigmoid squashes the output to the range `(0, 1)`.

## 3. DECISION RULE

Predicted class:

$$\&haty_{class} = \begin{cases} 1 & \text{if} \quad \&haty \geq 0.5 \\ 0 & \text{otherwise} \end{cases}$$

## 4. COST FUNCTION (CROSS-ENTROPY LOSS)

The loss for a single data point:

$$L(\text{\&haty}, y) = -[\ y \cdot \log(\text{\&haty}) + (1 - y) \cdot \log(1 - \text{\&haty})\ ]$$

For `m` training samples:

$$J(\mathbf{w}, b) = -1/m \sum_{i=1}^{m} [\ y^{(i)} \log(\text{\&haty}^{(i)}) + (1 - y^{(i)}) \log(1 - \text{\&haty}^{(i)})\ ]$$

## 5. GRADIENT COMPUTATION

Using calculus:

$$\partial J / \partial w_j = 1/m \sum_{i=1}^{m} (\ \text{\&haty}^{(i)} - y^{(i)}\ )\ x_j^{(i)}$$
$$\partial J / \partial b = 1/m \sum_{i=1}^{m} (\ \text{\&haty}^{(i)} - y^{(i)}\ )$$

## 6. GRADIENT DESCENT UPDATES

For each iteration:

$$w_j := w_j - \alpha \cdot \partial J / \partial w_j$$
$$b := b - \alpha \cdot \partial J / \partial b$$

where `α` is the learning rate.

## 7. SUMMARY OF THE LEARNING PROCESS

1. Initialize `w` and `b` to zero (or small random values).
2. Compute `z`, apply the sigmoid to get `&haty`.
3. Compute the cost `J` (cross-entropy loss).
4. Compute gradients $\partial J / \partial w_j$ and $\partial J / \partial b$.
5. Update `w` and `b` using gradient descent.
6. Repeat until convergence.