# Logistic Regression - Math Summary

## 1. Model (Hypothesis)

For binary classification with $d$ input features, logistic regression models the **log-odds** as a linear function:

$$z_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_d x_{id}$$

The predicted probability is obtained via the sigmoid function:

$$\hat{p}_i = P(y_i = 1 \mid x_i) = \sigma(z_i) = \frac{1}{1 + e^{-z_i}}$$

## 2. Design Matrix

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1d} \\ 1 & x_{21} & x_{22} & \dots & x_{2d} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nd} \end{bmatrix} \qquad \beta = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_d \end{bmatrix}$$

## 3. Linear Predictor (Logit)

$$\mathbf{z} = X\beta$$

Each component:

$$z_i = \log\left(\frac{\hat{p}_i}{1 - \hat{p}_i}\right)$$

## 4. Sigmoid Function

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

This maps any real-valued input to a probability in $(0, 1)$.

## 5. Predicted Probabilities

$$\hat{\mathbf{p}} = \sigma(X\beta)$$

where $\sigma(\cdot)$ is applied element-wise.

## 6. Loss Function (Binary Cross-Entropy)

The cost function to minimize is the negative log-likelihood:

$$J(\beta) = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log(\hat{p}_i) + (1-y_i)\log(1-\hat{p}_i)\right]$$

## 7. Gradient of the Loss

Let:

$$\hat{\mathbf{p}} = \sigma(X\beta)$$

Then the gradient with respect to $\beta$ is:

$$\nabla_\beta J(\beta) = \frac{1}{n} X^\top(\hat{\mathbf{p}} - \mathbf{y})$$

## 8. Optimization (Gradient Descent)

Logistic regression has no closed-form solution. Parameters are learned iteratively:

$$\beta^{(t+1)} = \beta^{(t)} - \alpha\frac{1}{n} X^\top(\hat{\mathbf{p}} - \mathbf{y})$$

where $\alpha$ is the learning rate.

## 9. Prediction (Classification Rule)

Given a threshold $\tau$ (usually $\tau = 0.5$):

$$\hat{y}_i = \begin{cases} 1 & \text{if } \hat{p}_i \geq \tau \\ 0 & \text{otherwise} \end{cases}$$

## 10. Feature Scaling (Standardization)

For each numeric feature:

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}$$

This improves numerical stability and convergence of gradient descent.

## 11. Interpretation of Coefficients

Each coefficient $b_j$ represents the change in **log-odds** for a unit increase in $x_j$ (holding other features constant):

$$\log\left(\frac{p}{1-p}\right) = b_0 + \sum_{j=1}^{d} b_j x_j$$

Exponentiating:

$$e^{b_j} = \text{odds ratio associated with } x_j$$

## 12. Model Evaluation Metrics

Common evaluation metrics include:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP} \qquad \text{Recall} = \frac{TP}{TP + FN}$$

$$\text{ROC AUC} = P\left(s(x^+) > s(x^-)\right)$$