Multiple Linear Regression - Math Summary

1. Model (Hypothesis)

For d input features:

$$\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_d x_{id}$$

Vector form:

$$\hat{y} = X\beta$$

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. Error (Residual)

$$e_i = y_i - \hat{y}_i$$

Vector form:

$$\mathbf{e} = y - X\beta$$

4. Cost Function (Mean Squared Error)

$$J(\beta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\|y - X\beta\|^2$$

5. Normal Equation

To minimize $J(\beta)$:

$$X^\top X\beta = X^\top y$$

6. Closed-Form Solution

$$\beta = (X^\top X)^{-1}X^\top y$$

If $X^\top X$ is not invertible:

$$\beta = (X^\top X)^{+}X^\top y$$

where $(\cdot)^+$ denotes the pseudo-inverse.

7. Prediction

For a new sample x_{new}:

$$\hat{y}_{\text{new}} = x_{\text{new}}^{\top}\beta$$

Expanded:

$$\hat{y}_{\text{new}} = b_0 + \sum_{j=1}^{d} b_j x_{\text{new},j}$$

8. Feature Scaling (Standardization)

For each numeric feature:

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

9. Mean Squared Error

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

10. Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

11. Interpretation of Coefficients

$$b_j = \text{change in } \hat{y} \text{ for a unit change in } x_j \text{ (holding other features constant)}$$