Multiple Linear Regression – Geometric Intuition

1. What is Multiple Linear Regression?

- **Multiple Linear Regression (MLR)** is an extension of **Simple Linear Regression**.
- It predicts a target variable y using **two or more independent variables** x1,x2,...,xp.
- The general equation is:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p$$

Where:

- b0 = Intercept (value of y when all xi=0)
- b1,b2,...,bp = Coefficients (impact of each feature on yy)
- x1,x2,...,xp = Independent features

2. Geometric Intuition

- In **Simple Linear Regression**, the best-fit line is a straight line in **2D space** (x vs y).
- In **Multiple Linear Regression**, the relationship exists in a **multi-dimensional space**:
  - With **2 features (x_1, x_2)**, the regression model forms a **plane** instead of a line.
  - With **3 features**, the model forms a **3D hyperplane**.
  - With **p features**, it forms a **p-dimensional hyperplane**.
- The goal is to find this **hyperplane** that **minimizes the distance (errors)** between actual data points and the predicted surface.

3. How the Model Fits the Data

- The model finds coefficients b1,b2,...,bp by minimizing the **Sum of Squared Errors (SSE)**:

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

- In geometric terms:
  - Each data point is a **vector in p-dimensional space**.
  - Linear regression projects these points onto a **regression hyperplane**.
  - The solution found is the one where the **distance between points and the hyperplane is minimized**.

4. Interpretation of Coefficients

- bj (coefficient for feature xj) tells:
  - ❖ The expected **change in y** for a **one-unit increase** in xj**, keeping all other variables constant**.
- This allows us to understand the **influence of each independent variable** on the target.

---

5. Visualization Example

- Imagine predicting **house price (y)** using:
  - ❖ **Size (x_1)** and **Number of rooms (x_2)**.
- The data points are plotted in **3D space** (x_1, x_2, y).
- The regression finds a **flat surface (plane)** that **best fits** these points.
- Predictions for new houses are made by projecting them onto this plane.

➢ Key Takeaway

- **Simple Linear Regression** = Best-fit **line** in 2D.
- **Multiple Linear Regression** = Best-fit **hyperplane** in higher dimensions.
- The model still aims to **minimize errors**, just in a **multi-feature space**.
