Multiple Linear Regression (MLR) - Short Note

Definition:
Multiple Linear Regression (MLR) is an extension of simple linear regression where a dependent variable (target) is predicted based on multiple independent variables (features). It is represented as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \epsilon$$

where:

- Y = Dependent variable (target)
- $X_1, X_2, ..., X_n$ = Independent variables (features)
- β_0 = Intercept
- $\beta_1, \beta_2, ..., \beta_n$ = Coefficients of independent variables
- ϵ = Error term

Example Problem

A company wants to predict the sales (Y) based on the amount spent on TV ads (X_1) and radio ads (X_2). The dataset includes:

TV Ads ($1000s)	Radio Ads ($1000s)	Sales ($1000s)
10	5	25
15	8	30
20	12	35

Using regression analysis, the estimated equation is:

$$Y = 5 + 1.5X_1 + 2X_2$$

Solved Example

Predict sales if TV Ads = $12,000 and Radio Ads = $7,000.

$$Y = 5 + (1.5 \times 12) + (2 \times 7)$$

$$Y = 5 + 18 + 14 = 37$$

Thus, predicted sales = **$37,000**.

Exercise

Given the regression equation:

$$Y = 10 + 2X_1 + 3X_2$$

Find Y when $X_1 = 8$ and $X_2 = 6$.

Moderately Tough Solved Example - Multiple Linear Regression

A real estate company wants to predict the **price of a house (in $1000s)** based on its **size (square feet)** and **number of bedrooms**. The multiple linear regression equation is derived as:

$$Y = 50 + 0.08X_1 + 15X_2$$

where:

- Y = Predicted price of the house (in $1000s)
- X_1 = Size of the house (sq. ft)
- X_2 = Number of bedrooms

Problem Statement

Predict the price of a house that is **2000 sq. ft** in size and has **3 bedrooms**.

Solution

Using the given equation:

$$Y = 50 + (0.08 \times 2000) + (15 \times 3)$$

$$Y = 50 + 160 + 45$$

$$Y = 255$$

Thus, the predicted price of the house = **$255,000**.

Challenging Multiple Linear Regression Problem

Problem Statement

A university wants to predict a student's **final exam score (Y)** based on their **study hours per week (X_1)**, **attendance percentage (X_2)**, and **number of practice tests taken (X_3)**. After running a multiple linear regression analysis, the estimated equation is:

$$Y = 20 + 2.5X_1 + 0.4X_2 + 3X_3$$

Using this equation, predict the **final exam score** for a student who:

- Studies **15 hours per week**
- Has an **80% attendance rate**
- Has taken **5 practice tests**

Solution

Substituting the given values into the equation:

$$Y = 20 + (2.5 \times 15) + (0.4 \times 80) + (3 \times 5)$$

$$Y = 20 + 37.5 + 32 + 15$$

$$Y = 104.5$$

Since exam scores are typically out of 100, we assume a **maximum possible score of 100**, so the final predicted score is **100 (capped at max score limit).**

Interpretation

- If no study hours, attendance, or practice tests were considered, the baseline score would be **20** (intercept).
- **Each additional study hour** increases the predicted score by **2.5 points**.
- **Each additional 1% in attendance** contributes **0.4 points**.
- **Each extra practice test** improves the score by **3 points**.