Multiple Linear Regression

Multiple Linear Regression is a Supervised machine Learning algorithm used to predict a continuous target variable using two or more independent variables.

→ It models a linear relationship b/w features & target

It finds a straight-line relationship between

→ dependent variable (Target) → what we want to predict

→ Multiple Independent variable (Features) → factors impacting prediction.

Ex:- Suppose we want to predict house price using:

Feature	meaning
i) Size	Bigger house → higher price
ii) Bedrooms	more rooms → higher price
iii) Distance to city	far → lower price

So the model learns:-

House Price = Function (Size, Bedrooms, Distance)

Mathematical Equation :- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$

↓ Hyperplane in n-dim co-ordinate

y → Predicted values

β_0 → intercept (base prediction)

$x_1, x_2 \ldots$ → Features

$\beta_1, \beta_2, \ldots$ → Co-efficients (importance of Each feature)

Coefficients tell how much the target changes when one feature increases by 1 unit.

→ The model finds the best coefficients, so the line fits the data points.

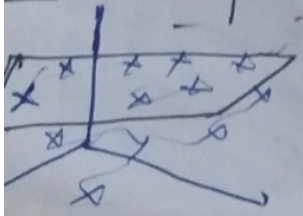
→ It reduces the error between Actual value vs Predicted

Ex:-

cgpa	iq	package

→ Goal

to find the values of $\beta_0, \beta_1, \beta_2$ (coefficients)



$$E = e^T e = (y-\hat{y})^T (y-\hat{y})$$

$$= (y^T - \hat{y}^T)(y-\hat{y})$$ ← symmetric means same

$$= y^T y - \hat{y}^T y - \hat{y} y^T + \hat{y}^T \hat{y}$$

$$\boxed{E = y^T y - 2 y^T \hat{y} + \hat{y}^T \hat{y}} \rightarrow Eq(3)$$

now replace $\hat{y} \longrightarrow x\beta$

$$E = y^T y - 2 y^T x\beta + (x\beta)^T (x\beta)$$

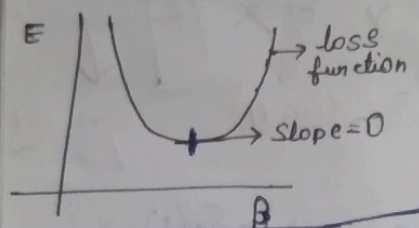
$$\boxed{E = y^T y - 2 y^T x\beta + \beta^T x^T x\beta} \rightarrow Eq(4)$$

here E is a function of β

$y \longrightarrow$ data output
$x \longrightarrow$ data input } data will not change. So here E is will change, when β changes

$E(\beta) \rightarrow$ find such value of β matrix for which E is min

$$\boxed{\frac{dE}{d\beta} = 0}$$



$$E = 0 - 2y^T x + 2\beta^T x^T x = 0$$

$$2\beta^T x^T x = 2 y^T x$$

$$\boxed{\beta^T x^T x = y^T x}$$

multiply by $(x^T x)^{-1}$ on b.s

$$\beta^T x^T x (x^T x)^{-1} = y^T x (x^T x)^{-1}$$

$$\beta^T I = y^T x (x^T x)^{-1}$$

$$\beta^T = y^T x (x^T x)^{-1}$$

apply (T) transpose on b.s

$$(\beta^T)^T = [y^T x (x^T x)^{-1}]^T$$

$$\beta = [(x^T x)^{-1}]^T (y^T x)^T$$

$$\boxed{\beta = [x^T x)^{-1} \; x^T y} \rightarrow Eq(5)$$

OLS → ordinary least square methodology

$$\frac{d(\beta^T x^T x \beta)}{d\beta} = 2\beta^T x^T x$$

x^T A x

$$\because x^T A x = 2 x^T A$$

only if A Symmetric.

$$A = (x^T x)^T = x^T x$$

$A^T = A$ → Symmetric

$$(x^T x)^{-1\,T} \Rightarrow$$

$$[A^{-1}]^T = A^{-1}$$

$(x^T x)^{-1}$ Symmetric

Mathematical Formulation:-

Ex:-

x_1 cgpa	x_2 iq	y Placement
8 x_{11}	80 x_{12}	8
7 x_{21}	70 x_{22}	7
5 x_{31}	120 x_{32}	15

$y_1 = 8$; $y_2 = 7$; $y_3 = 15$

$\hat{y}_1 = 8$, $\hat{y}_2 = 8$, $\hat{y}_3 = 8$

$$\begin{cases} \hat{y}_1 = \beta_0 + \beta_1 8 + \beta_2 80 \\ \hat{y}_2 = \beta_0 + \beta_1 7 + \beta_2 70 \\ \hat{y}_3 = \beta_0 + \beta_1 5 + \beta_2 120 \end{cases}$$

m cols $\longrightarrow$ n students/rows

now for n

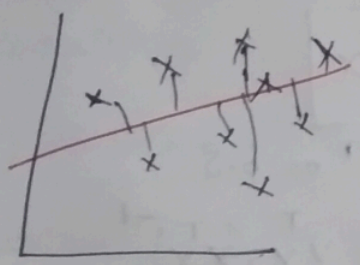
$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \vdots \\ \hat{y}_n \end{bmatrix}_{[n\times 1]} = \begin{bmatrix} \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \beta_3 x_{13} + \dots + \beta_m x_{1m} \\ \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_m x_{2m} \\ \beta_0 + \beta_1 x_{31} + \beta_3 x_{32} + \dots + \beta_m x_{3m} \\ \vdots \\ \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_m x_{nm} \end{bmatrix}_{[n\times 1]} = \hat{y}$$

$$\hat{y} = \begin{bmatrix} 1 x_{11} + x_{12} + \dots x_{1m} \\ 1 x_{21} + x_{22} + \dots x_{2m} \\ 1 x_{n1} + x_{n2} + \dots x_{nm} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}_{(m+1)}$$

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

$(n\times 1) = n\times(m+1)\ \ (m+1)$

Loss function for SLR



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

$\longrightarrow$ for mLR need to write matrix form.

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{[n\times 1]} \quad \hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}_{[n\times 1]}$$

$$e = y - \hat{y} \Rightarrow \begin{bmatrix} y_1 - \hat{y} \\ y_2 - \hat{y} \\ \vdots \\ y_n - \hat{y} \end{bmatrix}_{n\times 1}$$

$$e^T \cdot e = \underset{1\times n}{\begin{bmatrix} y_1 - \hat{y} & y_2 - \hat{y} & \dots & y_n - \hat{y} \end{bmatrix}} \underset{n\times 1}{\begin{bmatrix} y_1 - \hat{y} \\ y_2 - \hat{y} \\ \vdots \\ y_n - \hat{y} \end{bmatrix}} \Rightarrow \text{o/p } [1$$

$$e^T \cdot e = \left[(y_1 - \hat{y})^2 + (y_2 - \hat{y})^2 + \dots (y_n - \hat{y}_n)^2\right]$$

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

$$E = e^T \cdot e \Rightarrow \text{error function for mLR Eq(2)}$$

Problems

linear regression → OLS method
→ Gradient Descent

for 10 i/p:- OLS → $[(X^T X)^{-1} X^T Y]$

Here inverse function will consume more time

Time complexity = $O(n^3)$

$(X^T X)^{-1}$ Chol

$(m+1) \times n \quad n \times (m+1)$

$(m+1)(m+1)$

$\Rightarrow (m+1) \times (m+1)$

11×11 } for 10 i/p

$(100)^3 \Rightarrow (1000000)$ //

for high dimensional data like 100 i/p cols

$(100 \times 100)^3 \Rightarrow (10000)^3$ } → will give result
→ but too slow

i) Sensitive to outliers :-

→ OLS Squares errors → outliers get huge weight
→ Even one Extreme point can be change the regression line completely.

ii) Multicollinearity :-

→ when independent variables are highly correlated
→ Coefficients become unstable and interpretation becomes misleading
→ Hard to identify which feature actually impacts the target.

iii) Heteroscedasticity (Non-Constant variance of errors)

→ residual variance changes with predictors
→ Standard errors become biased → hypothesis tests become unreliable.

iv) Linearity Assumption:- works only if the relationship b/w is non linear → OLS gives poor predictions