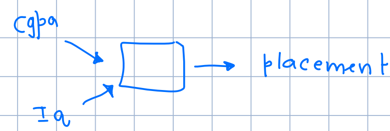
MLR : When we have multiple input Columns.

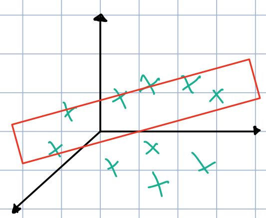
1000 students

cgpa	iq	placement



In Simple linear Reg, we had our data in 2D plane, but now its in bigger dimensions.

eg above cgpa, iq, package is in 3D



Equation of Plane = $\underbrace{\beta_0}_{\text{intercept}} + \beta_1 \underbrace{x_1}_{\text{cgpa}} + \beta_2 \underbrace{x_2}_{\text{iq}}$

If more colms : $\underbrace{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \dots \beta_n x_n}_{\text{hyper plane}}$

So, our goal is to find out $\beta_0, \beta_1, \beta_2 \dots \beta_n$

The values of B's tell How much that column effects our result

Always Check the Shape of data.

Mathematical formulation

β_0

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

$$\hat{y}_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} \dots \beta_m X_{1m}$$
$$\hat{y}_2 = \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} \dots \beta_m X_{2m}$$
$$\hat{y}_3 = \beta_0 + \beta_1 X_{31} + \beta_2 X_{32} \dots \beta_m X_{3m}$$
$$\vdots$$
$$\hat{y}_n = \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \dots \beta_m X_{nm}$$

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

To Calculate $\hat{y}$'s :

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

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

$$\hat{y} = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & \dots & X_{1m} \\ 1 & X_{21} & X_{22} & X_{23} & \dots & X_{2m} \\ 1 & & & & & \\ \vdots & & & & & \\ 1 & X_{n1} & X_{n2} & X_{n3} & \dots & X_{nm} \end{bmatrix}_{n \times (m+1)} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix}_{(m+1) \times 1}$$

$$\boxed{\underbrace{\hat{Y} = X\beta}_{\text{matrices}}}$$

$$\text{Error func} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \leftarrow \text{minimize}$$

Convert it into matrix form.

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1} \qquad \hat{Y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}_{n \times 1}$$

$$e = Y - \hat{Y} = \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix}$$

$$e^T e = \begin{bmatrix} y_1 - \hat{y}_1 & y_2 - \hat{y}_2 & \dots & y_n - \hat{y}_n \end{bmatrix}_{1 \times n} \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix}_{n \times 1} = [\ ]_{1 \times 1}$$

$$e^T e = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \dots (y_n - \hat{y}_n)^2$$

$$= \sum_{n=1}^{n} (y_i - \hat{y}_i)^2 \qquad \text{(loss func)}$$

$$\boxed{E = e^T e} \rightarrow \text{loss function of multiple linear reg.}$$

↳ we have to minimize this

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

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

$$E = y^T \cdot y - 2\, y^T \cdot \hat{y} + \hat{y}^T \cdot \hat{y}$$

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

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

$$\Rightarrow \boxed{E = y^T \cdot y - 2y^T \cdot X\beta + \beta^T X^T \cdot X\beta}$$

This $E(\beta)$ is func of Beta β.

To Prove: $y^T \hat{y} = \hat{y}^T y$

or

$$A^T B = B^T A$$

$$(A^T B)^T = B^T A$$

Now I have to prove $(A^T B) = (A^T B)^T$

I need to show $A^T B$ is a symmetric Matrix

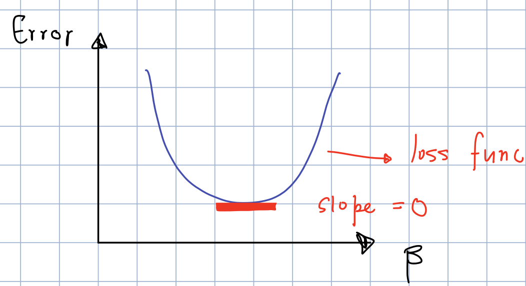
$$y^T \hat{y} = \underbrace{\boxed{y^T}}_{1 \times n} \underbrace{X}_{n \times (m+1)} \underbrace{B}_{(m+1) \times 1}$$

Sizes above suggest that

$$1 \times n \, (n \times 1) = 1 \times 1$$

We will get 1×1 matrix and its always Symmetric.

We have to find such Value for β matrix for which E is minimum



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

$$E = y^T \cdot y - 2y^T \cdot X\beta + \beta^T X^T \cdot X\beta$$

$$\frac{dE}{d\beta} = 0 + -2 y^T X + \underbrace{\frac{d}{d\beta} \beta^T \cdot X^T \cdot X\beta}_{\text{matrix differentiation}}$$

$$= 0 + -2 y^T X + 2\beta^T X^T \cdot X$$

here $X^T \cdot X$ should be symmetric:

$$(X^T \cdot X)^T = X^T \cdot X \quad \text{yes its Symmetric :)}$$

Matrix diff:

$$= \frac{d}{dx} x^T A x$$

$$= 2x^T A$$

This is only True if A is matrix

$$E = -2y^T X + 2\beta^T X^T X$$

$$= \cancel{2}\beta^T X^T X = \cancel{2} y^T X$$

$\Rightarrow \quad \beta^T X^T X = y^T X$

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

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

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

Transpose both Sides

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

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

$(X^T X)^{-1}$ is symmetric : To prove

let $A = X^T X$

$A A^{-1} = I$

$(A A^{-1})^T = I^T$

$(A^{-1})^T (A)^T = I$

$(A^{-1})^T A = I$

$(X^T X)^T = X^T X$

$A^T = A$

$(A^{-1})^T A A^{-1} = I A^{-1}$

$\boxed{(A^{-1})^T = A^{-1}}$

so A^{-1} is symetric

$\Rightarrow (X^T X)^{-1}$ is symmetric

$$\boxed{\beta = (X^T X)^{-1} X^T y} \longrightarrow OLS$$

this will minimize error; "β"'s value