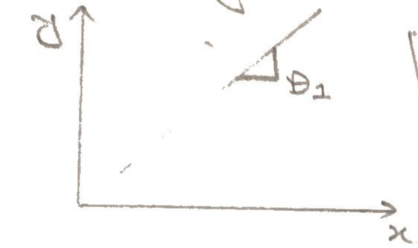
"Multiple Linear Regression"

using OLS (Ordinary Least Square) method

In simple linear regression we try to find a linear equation in a 2D plane and also try to find it by minimizing the error.



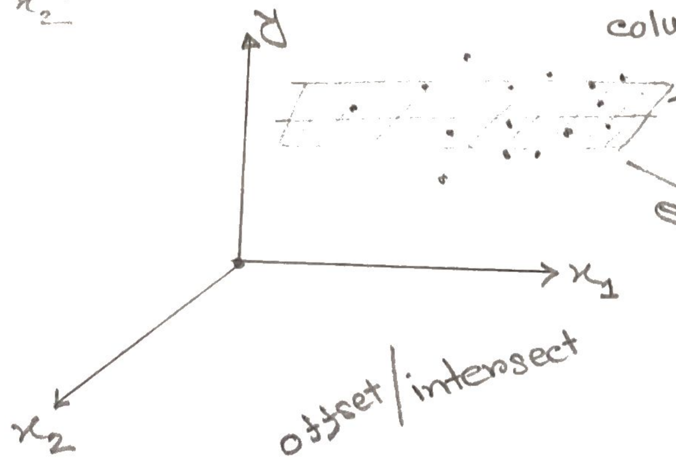
$$\boxed{y = \theta_0 + \theta_1 x}$$

Let's assume we have now 2 input feature and 1 out-come (numerical).

Note

Diamension of plane = $n x + y$

n is the number of input column.



→ a $\underline{\text{plane}}$ not a line
2D

equation

$$\boxed{y = \theta_0 + \theta_1 x_1 + \theta_2 x_2}$$

2D plane equation

offset / intersect

If there are n columns

$$\boxed{y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots\dots + \theta_n x_n}$$

→ Hyper plane. in n dimension cordinate system

$\theta \rightarrow$ weight of each column.

$\theta_1 \uparrow \rightarrow$ means x_1 has more contribution in prediction

$\theta_1 \downarrow \rightarrow$ " x_1 " less " " "

Mathmatical Formation:

Data

cgpa	iq	Placement
8	80	8
7	70	7
5	120	15
x_1	x_2	y

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

assume we know $\theta_0 \quad \theta_1 \quad \theta_2$

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

$$\hat{y}_1 = \theta_0 + 8\theta_1 + 80\theta_2$$
$$\hat{y}_2 = \theta_0 + 7\theta_1 + 70\theta_2$$
$$\hat{y}_3 = \theta_0 + 5\theta_1 + 120\theta_2$$

index notation:

$$\hat{y}_1 = \theta_0 + \theta_1 x_{11} + \theta_2 x_{12}$$
$$\hat{y}_2 = \theta_0 + \theta_1 x_{21} + \theta_2 x_{22}$$
$$\hat{y}_3 = \theta_0 + \theta_1 x_{31} + \theta_2 x_{33}$$

assume instead of 3 col we have m cols

$$\hat{y}_1 = \theta_0 + \theta_1 x_{11} + \theta_2 x_{12} + \cdots + \theta_m x_{1m}$$
$$\hat{y}_2 = \theta_0 + \theta_1 x_{21} + \theta_2 x_{22} + \cdots + \theta_m x_{2m}$$
$$\hat{y}_3 = \theta_0 + \theta_1 x_{31} + \theta_2 x_{32} + \cdots + \theta_m x_{3m}$$

If we have n row / entries

$$\hat{y}_n = \theta_0 + \theta_1 x_{n1} + \theta_2 x_{n2} + \theta_3 x_{n3} + \cdots + \theta_m x_{nm}$$

to make things simple

$$\hat{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \theta_0 + \theta_1 x_{11} + \cdots + \theta_m x_{1m} \\ \theta_0 + \theta_1 x_{21} + \cdots + \theta_m x_{2m} \\ \vdots \\ \theta_0 + \theta_1 x_{n1} + \cdots + \theta_m x_{nm} \end{bmatrix}$$

Shape $n \times 1$

shape: $n \times 1$

Augmented Form (Decompose using DOT)

$$\hat{Y} = \underbrace{\begin{bmatrix} 1 & x_{11} & \cdots & x_{1m} \\ 1 & x_{21} & \cdots & x_{2m} \\ \vdots & & & \\ 1 & x_{n1} & \cdots & x_{nm} \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_m \end{bmatrix}}_{\theta}$$

$$\boxed{\hat{Y} = X\theta} \longrightarrow (1)$$

$$\boxed{\hat{Y} = X\theta}$$

$\hat{Y}$ → hypothesis/predicted value matrix $(n \times 1)$

θ → coefficient matrix $[(m+1) \times 1]$

X → Input/Feature matrix $[n \times (m+1)]$

m is the dimension

Error Function:

Simple Linear

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

convert it into matrix form

Y → actual value matrix

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \qquad \hat{Y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_m \end{bmatrix}$$

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

$$e^T . e = \underbrace{\begin{bmatrix} y_1 - \hat{y}_1 & y_2 - \hat{y}_2 & \cdots & y_m - \hat{y}_m \end{bmatrix}}_{\text{shape } (1 \times m)} \times \underbrace{\begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_m - \hat{y}_m \end{bmatrix}}_{\text{shape } (m \times 1)}$$

$e^T . e$ → Shape 1×1 → a single value

$$e^T e = (y_1-\hat{y}_1)^2 + (y_2-\hat{y}_2)^2 + \cdots + (y_m-\hat{y}_m)^2$$

$$e^T e = \sum_{i=1}^{m} (y_i-\hat{y}_i)^2$$

$$\boxed{(AB)^T = B^T A^T}$$

$\therefore$ Loss Function For MLR: $\boxed{E = e^T e} \rightarrow (1)$

we have minimize this

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

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

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

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

$\Rightarrow$

Final Loss function

let $y = A$, $\hat{y} = B$

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

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

$A^T B = c \rightarrow$ Let

$c = c^T$

means we need to prove

$c = A^T B$ is a symetric matrix

$$y^T\hat{y} = y^T X\theta$$

$$= (1\times n).[n\times(m+1)].[(m+1)\times 1]$$

$$= (1\times n).(n\times 1) = [1\times 1] \rightarrow \text{single value symetric}$$

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

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

$$\boxed{E = y^T y - 2y^T X\theta + \theta^T X^T X\theta}$$

(4)

E is a function of θ

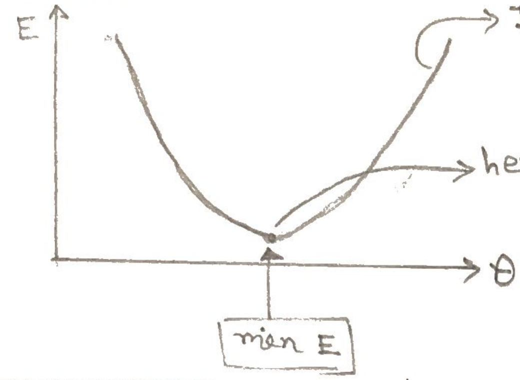
$E(\theta)$

Aristoderm

$$E(\theta) = Y^TY - 2Y^TX\theta + \theta^TX^TX\theta$$

Now find a value of θ for which $E(\theta)$ will be minimum.

It's a convex function



here slope is 0

matrix diff.

$$\boxed{\frac{d}{dA}\left(AxA^T\right) = 2A^Tx}$$

when A is symetric $\boxed{A^T=A}$

$$\boxed{\frac{dE}{d\theta} = 0} \rightarrow \frac{dE}{d\theta} = 0 - 2Y^TX + 2\theta^TX^TX = 0$$

$$\Rightarrow \theta^TX^TX = Y^TX$$

$$\Rightarrow \theta^TX^TX\left(X^TX\right)^{-1} = Y^TX\left(X^TX\right)^{-1}$$

[mutipling both X^TX]

$\boxed{AA^{-1} = 1}$

$$\Rightarrow \theta^TI = Y^TX\left(X^TX\right)^{-1}$$

$$\Rightarrow \theta^T = Y^TX\left(X^TX\right)^{-1} \qquad [AI = A]$$

$$\Rightarrow \left(\theta^T\right)^T = \left(Y^TX\left(X^TX\right)^{-1}\right)^T$$

$$\Rightarrow \theta = \left[\left(X^TX\right)^{-1}\right]^T\left[Y^TX\right]^T \qquad [(AB)^T = B^TA^T]$$

This is a symetric matrix

$$\Rightarrow \theta = \left[\left(X^TX\right)^{-1}\right]^T\left(X^TY\right)$$

$$\Rightarrow \boxed{\theta = \left(X^TX\right)^{-1}.X^T.Y} \rightarrow (5)$$

we already have X and Y so we can find θ and use $\hat{Y} = X\theta \longrightarrow$ we can predict

⇒ Prove $(X^T X)^{-1}$ is symetric: $\underbrace{\left[(X^T X)^{-1}\right]^T}_{LHS} = \underbrace{(X^T X)^{-1}}_{RHS}$

$X \xrightarrow{\text{shape}} [n \times (m+1)]$

$X^T \xrightarrow{\text{shape}} [(m+1) \times n]$

$X^{-1} \longrightarrow [n \times (m+1)]$

$$LHS = \left[\left[(m+1) \times n\right]\left[n \times (m+1)\right]\right]^T$$

$$= \left[(m+1) \times (m+1)\right]^T$$

$$= \left[(m+1) \times (m+1)\right]$$

$$RHS = \left[\left[(m+1) \times n\right]\left[n \times (m+1)\right]\right]^{-1}$$

$$= \left[(m+1) \times (m+1)\right]^T = \left[(m+1) \times (m+1)\right]$$

$\boxed{LHS = RHS} \longrightarrow (X^T X)^{-1}$ is a symetric matrix

Problem with OLS

For multivariable linear regression we need to find the inverse of X. which is much much computationally expensive. $O(n^3)$.

$$(X^T X)^{-1} \to \left[\left((m+1) \times n\right)\left[n \times (m+1)\right]\right]^{-1}$$

$$\to \left[(m+1) \times (m+1)\right]^{-1} \longrightarrow O\left((m+1)^3\right)$$

As dimension becomes larger
It will be quiet tough to
find inverse