Linear Regression :-

1. check if data is linear or sort of linear.

2. in 99.99% Case real world data is sort of linear (not perfectly straight)

3. Linear Regression makes a best fit straight line on this sort of linear data.

4. This line is chosen to minimize the error. (difference between predicted and actual values)

* Assumption of Linear Regression.

1. Linear Relationship :-
input and output should have a strai ght-line relationship.

2. No multicollinearity :-
input features should not be highly related to each other.

3. Normality of Residuals :-
The prediction errors should follow a normal (bell-shaped) distribution.

Homoscedasticity :-
Error should have constant spread across all values of input.

5. No autocorrelation :-
Errors shouldn't follow a pattern they should be random.

* Simple Linear Regression Mathematical Formulation :-

(m, b)

Closed form solution

(direct formula)

OLD
(ordinary least square)

(Low dimension)

Non-closed form solution

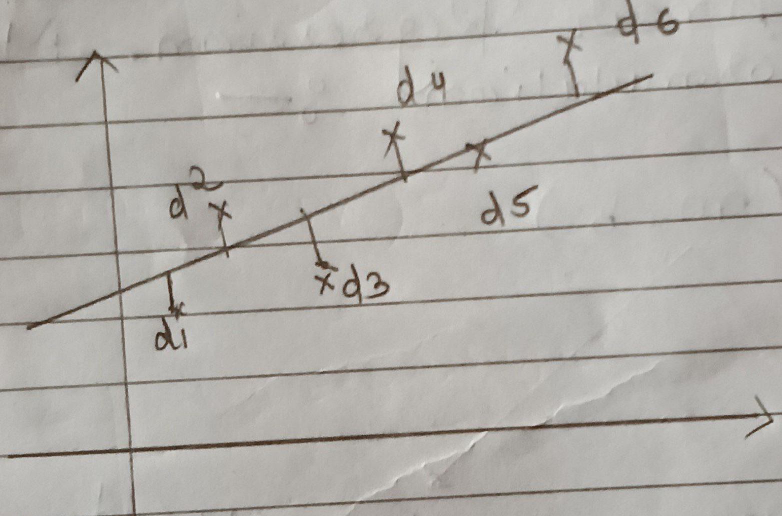
[Gradient discent]

(High dimension)

$$\bar{Y} = mx + b$$

$$b = \bar{Y} - mx$$

$$m = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$



$$E = d_1 + d_2 + d_3 + \ldots\ldots d_n$$

Positive and negative error can cancel out each other so here we use square of error.

why modulus is not use instead of square.?

→ Because modulus is not differentiable at origin. and square is differentiable at all points.

$E = d_1^2 + d_2^2 + d_3^2 + \dots + d_n^2$

$$\boxed{E = \sum_{i=1}^{n} d_i^2}$$ ......1 Error function

$d_i = y_i - \hat{y_i}$ ..(2).. (actual - predicted)

put (2) in (1)

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

$$\hat{y_i} = mx_i + b$$

$$E = \sum_{i=1}^{n} (y_i - mx_i - b)^2$$

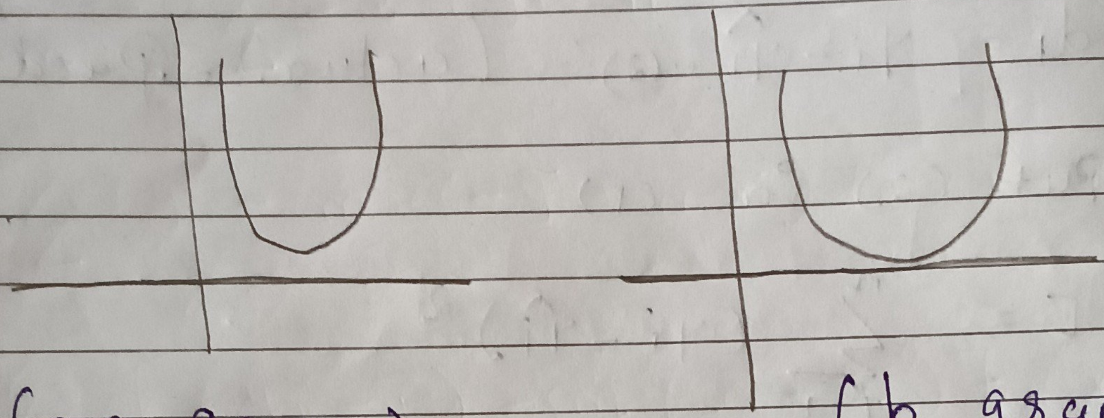
$$E(m,b) = \sum_{i=1}^{n} (y_i - mx_i - b)^2$$

Let $b = 0$

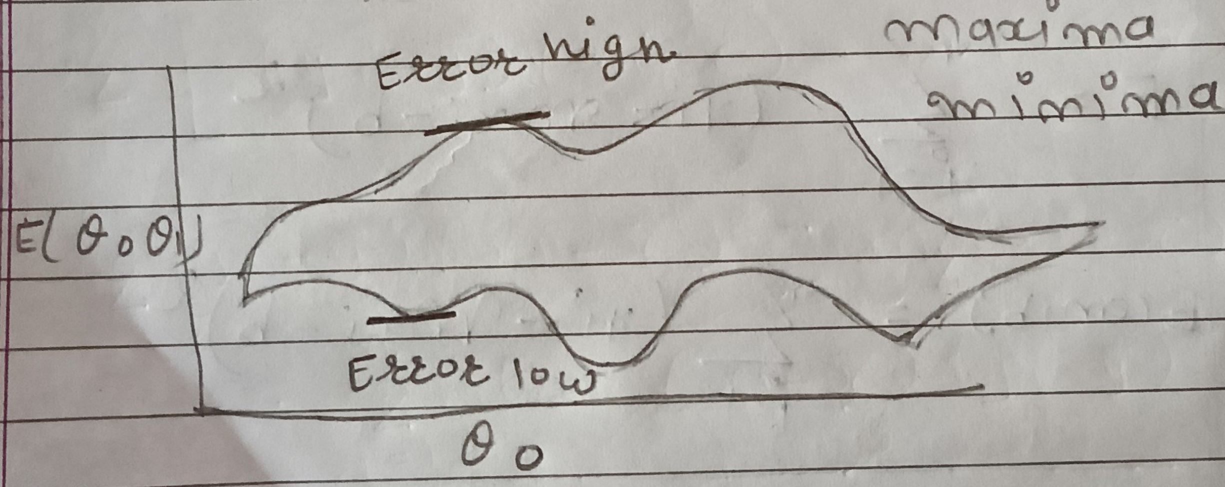
$$E(m) = \sum_{i=1}^{n} (y_i - mx_i)$$

Let $m = 1$

$$E(m, b) = \sum_{i=1}^{n} (y_i - x_i - b)^2$$



(m graph) (b graph)



$$\frac{\partial E}{\partial b} = \sum_{i=1}^{n} \frac{\partial}{\partial b} (y_i - mx_i - b)^2 = 0$$

$$\frac{\partial E}{\partial b} = \sum_{i=1}^{n} 2 (y_i - mx_i - b)(-1) = 0$$

$$\frac{\partial E}{\partial b} = \sum_{i=1}^{n} -2(y_i - mx_i - b) = 0$$

divide both side by -2

$$\sum_{i=1}^{n} (y_i - mx_i - b) = 0$$

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} mx_i - \sum_{i=1}^{n} b = 0$$

divede both side by n

$$\frac{\sum_{i=1}^{n} y_i}{n} - \frac{\sum_{i=1}^{n} mx_i}{n} - \frac{\sum_{i=1}^{n} b}{n} = 0$$

$$\bar{y} - m\bar{x} - \frac{\cancel{n} b}{\cancel{n}} = 0$$

$$\left[\frac{y_i}{n} = \bar{y} \,,\ \frac{x_i}{n} = \bar{x} \right] \therefore \text{mean}$$

$$\left[\frac{b + b + b + \cdots + b}{n} \right] n \text{ times } \frac{nb}{n}$$

$$\bar{y} - m\bar{x} - b = 0 \quad \ldots\ldots (3)$$

$$b = \bar{y} - m\bar{x}$$

$$E = \sum (y_i - mx_i - \bar{y} + m\bar{x})^2 = 0$$

$$\frac{\partial E}{\partial m} = \sum \frac{\partial}{\partial m} (y_i - mx_i - \bar{y} + m\bar{x})^2 = 0$$

$$\frac{\partial E}{\partial m} = \sum -2 (y_i - mx_i - \bar{y} + m\bar{x})(-x_i + \bar{x}) = 0$$

$$\sum (y_i - mx_i - \bar{y} + m\bar{x})(x_i - \bar{x}) = 0$$

$$\sum \left[(y_i - \bar{y}) - m(x_i - \bar{x}) \right] (x_i - \bar{x}) = 0$$

$$\sum \left[(y_i - \bar{y})(x_i - \bar{x}) - m(x_i - \bar{x})^2 \right] = 0$$

$$\sum (y_i - \bar{y})(x_i - \bar{x}) - m\sum (x_i - \bar{x})^2 = 0$$

$$m = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad \ldots\ldots (4)$$

Multiple Linear Regression :-

it predicts an outcome using multiple input variables.

it finds a mathematical equation that best fits the relationship between those input and the output.

Mathematical formulation :-

Predected value :-

$$\hat{y} = \begin{bmatrix} \hat{y_1} \\ \hat{y_2} \\ \hat{y_3} \end{bmatrix} = \begin{bmatrix} \beta_0 & B_1 X_{11} & B_1 X_{12} \dots & B_1 X_{13} \\ B_0 & B_1 X_{21} & B_1 X_{22} & B_1 X_{23} \\ \vdots & & & \\ B_0 & B_1 X_{1001} & B_1 X_{1002} & B_1 X_{1003} \end{bmatrix}$$

Suppose rows $\rightarrow n$, columns $\rightarrow m$

$$\hat{y} = \begin{bmatrix} \hat{y_1} \\ \hat{y_2} \\ \vdots \\ \hat{y_n} \end{bmatrix} = \begin{bmatrix} B_0 & B_1 X_{11} & B_2 X_{12} & \dots & B_m X_{1m} \\ \beta_0 & B_1 X_{21} & B_2 X_{22} & \dots & B_m X_{2m} \\ \vdots & & & & \\ B_0 & B_1 X_{n1} & B_2 X_{n2} & \dots & B_m X_{nm} \end{bmatrix}$$

$A * B = AB$

$$\begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1m} \\ 1 & X_{21} & X_{22} & \cdots & X_{2m} \\ \vdots & & & & \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nm} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix}$$

$$\hat{y} = X\beta \quad \cdots\cdots\cdots \quad (1)$$

Error function from simple linear regression

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

we can write it as

$e^T e$ where $e = (y_i - \hat{y_i})$

$$\begin{bmatrix} (y_1 - \hat{y_1}) & (y_2 - \hat{y_2}) & (y_3 - \hat{y_3}) \\ & (y_n - \hat{y_n}) & \end{bmatrix} \begin{bmatrix} (y_1 - \hat{y_1}) \\ (y_2 - \hat{y_2}) \\ \vdots \\ (y_n - \hat{y_n}) \end{bmatrix}$$

$$(y_1 - \hat{y_1})^2 + (y_2 - \hat{y_2})^2 + \cdots (y_n - \hat{y_n})^2$$

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

So $E = e^T e$

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

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

from eq (1)

$E = \left[y^T - (X\beta)^T \right] \left[y - (X\beta) \right]$

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

Prove $y^T X\beta = (X\beta)^T y$

Let $y^T = A$, $X\beta = B$

$A^T B = B^T A$

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

here we have to prove $A^T B = (A^T B)^T$

Let $A^T B = C$ then $(A^T B)^T = C^T$

$C = C^T$

$$(y^T X \beta)^T = y^T X \beta$$

$$\hat{y} = \begin{bmatrix} - \\ - \\ - \\ \vdots \\ - \end{bmatrix}$$

$$(n \times 1) \rightarrow (1 \times n)$$

$$(y^T X B)^T = y^T X B$$

$$(1 \times n)\,(n \times (m+1))\,[(m+1) \times 1]$$

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

$$1 \times 1 = [1]$$

$$= [1]^T = [1]$$

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

$$E = y^T y - 2y^T X \beta + (X B)^T X B \quad \ldots\ldots (4$$

$$\frac{dE}{d\beta} = \frac{d}{d\beta}\left[y^T y - 2y^T X\beta + \cancel{(X\beta)^T X\beta}\right.$$

$$(\underbrace{\beta^T}_{A}\ \underbrace{X^T X}_{X}\ \underbrace{\beta}_{A})$$

$$\frac{dE}{d\beta} = 0 - 2y^T X + 2X^T X B^T = 0$$

$$-2y^TX + 2X^TB^T = 0$$

$$\cancel{2}X^TXB^T = \cancel{2}y^TX$$

$$B^T = \frac{y^TX}{X^TX}$$

$$B^T = y^Tx\,(X^Tx)^{-1}$$

$$(B^T)^T = \left[y^Tx\,(x^Tx)^{-1}\right]^T$$

$$B = \underbrace{\left[(X^TX)^{-1}\right]^T}_{\text{square matrix}} \left[y^Tx\right]^T$$

$$\boxed{B = (X^TX)^{-1}\,X^Ty}$$