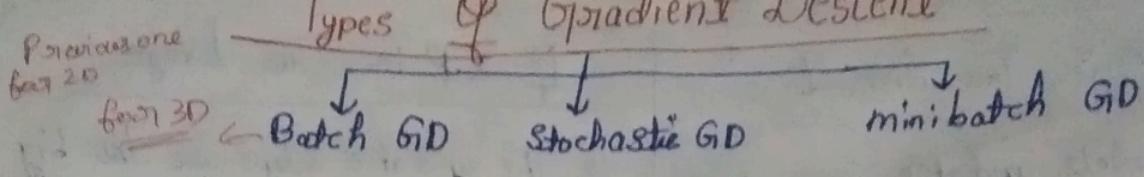


## Types of Gradient Descent



### i) Batch Gradient Descent:

→ uses the entire datasets to complete the gradient before updating the parameters

#### advantage

→ very accurate & stable updates

→ Guaranteed to move smoothly toward minimum

#### disadvantage

→ very slow for large data

→ Requires a lot of memory

### Mathematical Formulation:-

→ Consider 3 cols dataset  $\Rightarrow$

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

(lpa)                    (cgpa)                    (iv)

$x_1$	$x_2$	$y$
Cgpa	iv	lpa
8.1	93	3.2
7.5	95	3.5

i) Random variables values  $\Rightarrow \beta_0 = 0, \beta_1 = \beta_2 = 1$

ii) epoch = 100,  $\eta = 0.1$

$$\left\{ \begin{array}{l} \beta_0 = \beta_0 - \eta \text{ slope} \\ \beta_1 = \beta_1 - \eta \text{ slope} \\ \beta_2 = \beta_2 - \eta \text{ slope} \end{array} \right\}$$

→ so here Loss function is depending on  $\{\beta_0, \beta_1, \beta_2\}$  these three

$$\left( \frac{\partial L}{\partial \beta_0} \right), \left( \frac{\partial L}{\partial \beta_1} \right), \left( \frac{\partial L}{\partial \beta_2} \right)$$

$$\begin{aligned} L &= \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= \frac{1}{2} \left[ (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 \right] \end{aligned}$$

$$\begin{aligned} \frac{\partial L}{\partial \beta_0} &= \frac{1}{2} \left[ (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 \right] \\ &= \frac{1}{2} \left[ 2(y_1 - \hat{y}_1)(-1) + 2(y_2 - \hat{y}_2)(-1) \right] \\ &= \frac{-2}{2} \left[ (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 \right] \end{aligned}$$

→ for  $n$

$$\frac{\partial L}{\partial \beta_0} = \frac{-2}{n} \left[ (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \dots + (y_n - \hat{y}_n)^2 \right]$$

$$\frac{\partial L}{\partial \beta_0} = \frac{-2}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

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

$$L = \frac{1}{2} [(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2]$$

$$L = \frac{1}{2} \left[ \left( y_1 - \underbrace{\beta_0 + \beta_1 x_{11} + \beta_2 x_{12}}_{\hat{y}_1} \right)^2 + \left( y_2 - \underbrace{\beta_0 + \beta_1 x_{21} + \beta_2 x_{22}}_{\hat{y}_2} \right)^2 \right]$$

$$\frac{\partial L}{\partial \beta_1} = \frac{1}{2} \left[ 2(y_1 - \hat{y}_1)(-x_{11}) + 2(y_2 - \hat{y}_2)(-x_{21}) \right]$$

$$\frac{\partial L}{\partial \beta_2} = \frac{-2}{n} \left[ (y_1 - \hat{y}_1)(x_{12}) + (y_2 - \hat{y}_2)(x_{22}) + \dots + (y_n - \hat{y}_n)(x_{n2}) \right]$$

$\rightarrow$  form  $\boxed{\frac{\partial L}{\partial \beta_1} = \frac{-2}{n} \sum_{i=1}^n (y_i - \hat{y}_i) (\underbrace{x_{i1}}_{\text{entire column}})}$

$\left. \begin{array}{l} \text{So here } \\ x_{i1} \text{ represents} \\ \text{entire column} \end{array} \right\}$

$x_{11}$
$x_{21}$
$x_{31}$
$\vdots$
$x_{n1}$

$\rightarrow x_{i1} \Rightarrow$  1st column of data

$\rightarrow \beta_1 \Rightarrow$  weight or values of 1 col

$$\boxed{\frac{\partial L}{\partial \beta_2} = \frac{-2}{n} \sum_{i=1}^n (y_i - \hat{y}_i) (\underbrace{x_{i2}}_{\text{entire column}})}$$

$x_{i2} \rightarrow$  2nd column  
of dataset

$\rightarrow$  so for m cols  $\{\beta_0 - \beta_m\}$

$$\boxed{\frac{\partial L}{\partial \beta_m} = \frac{-2}{n} \sum_{i=1}^n (y_i - \hat{y}_i) (\underbrace{x_{im}}_{\text{entire column}})}$$

$x_{im} \rightarrow$  m<sup>th</sup> column  
of dataset