① Mathematical Formulation for Gradient Descent on n dimensional data set

3 cols

		cgpa x_1	iq x_2	lpa y
x_{11}	x_{12}	8.1	9.3	3.2
x_{21}	x_{22}	7.5	9.5	3.5

$\rightarrow y_1, y_2$

Intercept

$$\underset{(lpa)}{y} = \beta_0 + \beta_1 \underset{(cgpa)}{x_1} + \beta_2 \underset{(iq)}{x_2}$$

* we have to find 3 coefficients $(\beta_0, \beta_1, \beta_2)$

① Random values

$\beta_0 = 0$, $\beta_1, \beta_2 = 1$

② Epochs = 100, learning rate (α) = 0.1

$\beta_0 = \beta_0 - \alpha$ slope $\leftarrow$ ($b_{new} = b_{old} - \alpha$ (slope))

$\beta_1 = \beta_1 - \alpha$ (slope)

$\beta_2 = \beta_2 - \alpha$ (slope)

Loss function $(\beta_0, \beta_1, \beta_2) \longrightarrow$ 4d graph

$$\frac{\partial L}{\partial \beta_0}, \frac{\partial L}{\partial \beta_1}, \frac{\partial L}{\partial \beta_2} \longrightarrow \text{slope}$$

for n-dimensional data we have to calculate $(n+1)$ slope and then update.

* How to calculate derivative of Loss function w.r.t intercept

* $\frac{\partial L}{\partial \beta_0} =$

$\frac{\partial (L)}{\partial \beta_0}$ — $\boxed{\text{Loss function} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$ — MSE

$n = 2$ (No of indepent columns)

so

$$\text{Loss func} = \frac{1}{2} \left[(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 \right]$$

* Now we have to calculate $(\hat{y})$ and we have (y)

cgpa	iq	lpa
8.1 (x_{11})	9.3 (x_{12})	3.2
7.5 (x_{21})	9.5 (x_{22})	3.5

$\rightarrow y_1, y_2$

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

so,

$$\hat{y}_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12}$$

$$\hat{y}_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22}$$

then

$$= \frac{1}{2} \left[(y_1 - \beta_0 + \beta_1 x_{11} + \beta_2 x_{12})^2 + (y_2 - \beta_0 \right.$$

then

$$\text{loss function} = \frac{1}{2} \left[(y_1 - \beta_0 - \beta_1 x_{11} - \beta_2 x_{12})^2 + (y_2 - \beta_0 - \beta_1 x_{21} - \beta_2 x_{22})^2 \right]$$

Page No.
Date

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

$(0) \quad (-1) \quad 0 \quad (0)$

(Derivative)

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

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

(n)

④ For (n) dimension

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

↓ In simple

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

↓

(Slope of the β_0)

② $\boxed{\frac{\partial L}{\partial \beta_1}}$ Loss function $= \frac{1}{2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

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

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

$$\boxed{\frac{\partial}{\partial \beta_1} -\beta_1 x_{11} = -x_{11}}$$

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

Now,

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

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

for n dimension

$$= -\frac{2}{n}\left[(y_1 - \hat{y}_1)(x_{11}) + (y_2 - \hat{y}_2)(x_{21}) \ldots (y_n - \hat{y}_n)(x_{n1})\right]$$

in simple

$$\frac{\partial L}{\partial \beta_1} = -\frac{2}{n}\sum_{i=1}(y_i - \hat{y}_i)(x_{n1})$$

$$\boxed{\frac{\partial L}{\partial \beta_1} = \frac{-2}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)(x_{i1})}$$

slope of β_1

x_1 col 1 Represents (x_1)

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

slope of β_2

x_1	x_2	y
x_{11}	x_{12}	
x_{21}	x_{22}	
x_{31}	x_{32}	

x_{i2} x_{12}

Represent col 2

x_1	x_2	$\cancel{x_{in}}$
x_{11}	x_{12}	
x_{21}	x_{22}	
x_{31}	x_{32}	
x_{41}	x_{42}	
⋮	⋮	
x_{i1}	x_{i2}	

for (m columns) βm

$$\boxed{\frac{\partial L}{\partial \beta m} = \frac{-2}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)(x_{im})}$$

(General formula)

(y-hat)

$\hat{y}_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \beta_3 x_{13}$

x_1	x_2	x_3	y

$\hat{y}_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \beta_3 x_{23}$

So,

$$\hat{y} = \beta_0 + [x_{11} \; x_{12} \; x_{13}] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$ ← Dot product

$$\boxed{\hat{y} = np.dot(coef\_, x\_train) + \beta_0}$$

$\hat{y} = $ ~~np.mean~~

$\hat{y} = $ np.dot (x_train, coef_) intercept (β_0)

$= (353, 10)(10, 1) + \beta_0$

$(353, 1) + \beta_0$

$\hat{y} = (353, 1)$

In python

$\hat{y} = $ np.dot (x_train, self.coef_) self.intercept