

Gradient descent - I

↳ It is an optimization algorithm used to minimize a fn, most commonly a loss fn in ML

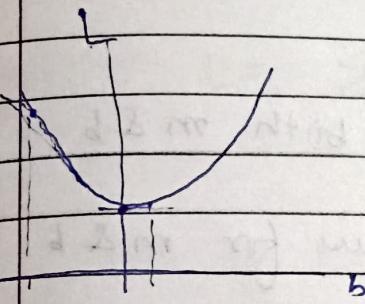
$$\hat{y}_i = mx_i + b$$

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

assume $b = 78.35$ (we know)

$$L = \sum (y_i - 78.35 + x_i - b)^2$$

L depends on square of b



→ I want b for which L is min

Step-1 - Select a random b_{init}

2 - find slope at first pt

→ If slope is -ve then move right

→ If slope is +ve then move left.

$$\therefore b_{\text{new}} = b_{\text{old}} - \text{slope}$$

$$b_{\text{new}} = b_{\text{old}} - \eta \text{slope} \quad \eta \rightarrow \text{learning rate}$$

↳ we wait so met

there is no drastic changes

Generally.. $\eta = 0.01$

in slope

⇒ when to stop -

$$i) b_{\text{old}} - b_{\text{new}} \geq 0.001$$

ii) Iterations

• mathematical formula -

Step-1 :- Start with random value of b

Step-2 :- for i in epochs:

$$b_{new} = b_{old} - \eta \text{ Slope}$$

$\eta = 0.01$

↳ finding Slope -

$$\frac{d}{db} \left(\sum_i (y_i - \hat{y}_i)^2 \right) = \sum_i \frac{d}{db} (y_i + mx_i - b)^2$$

$$\text{Slope} = \sum_i 2(y_i - mx_i - b) = \infty \quad (\text{initially } b=0)$$

$n \text{ Slope} \rightarrow \text{step size}$

→ Now we are considering both m & b .

1) initialize random value for m & b

$$m = 1 \text{ and } b = 0$$

2) epochs = 100, $\Delta\sigma = 0.01$

for i in epochs :

$$b = b - n \text{ Slope}$$

$$m = m - n \text{ Slope}$$

$$L = \sum_i (y_i - \hat{y}_i)^2 =$$

$$L(m, b) = \sum_i (y_i - mx_i - b)^2$$

$$b\text{-slope} = \frac{\partial L}{\partial b} \sum_i (y_i - mx_i - b)$$

$$= -2 \sum_i (y_i - mx_i - b)$$

$$m\text{-slope} = \frac{\partial L}{\partial m} = -2 \sum (y_i - mx_i - b) x_i$$

- If learning rate is too low then step size will be small and it will take too long to converge so model become slow
 - If learning rate is too high then it will do zigzag motion and may not reach to min pt
 - Effect of Loss fn -

$$L = \sum (y_i - \hat{y}_i)^2 \rightarrow \text{convex } f^n$$

↳ have only 1 min.
and that is global minimum

Lecture -2

Gradient Descent

L Batch 9D

Stochastic 9D

u minibarn 9D

- Mathematical formulae :-
is assuming 3 columns & 2 rows

x_1	x_2	y
8.1	93	3.2
7.5	95	3.5

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

Step.1 - choose random values

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

2) epoch = 100, lr = 0.1

$$\beta_0 = \beta_0 - n \text{ Slope} \quad \frac{\partial L}{\partial \beta_0}$$

$$\beta_1 = \beta_1 - n \text{ Slope} \quad \frac{\partial L}{\partial \beta_1}$$

$$\beta_2 = \beta_2 - n \text{ Slope} \quad \frac{\partial L}{\partial \beta_2}$$

$$L(\beta_0, \beta_1, \beta_2)$$

→ If we have n -dimensions then $(n+1)$ dimensions

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

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

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

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

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

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

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

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

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

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

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

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

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

$x_{i1} \Rightarrow$ represent values of 1st column

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