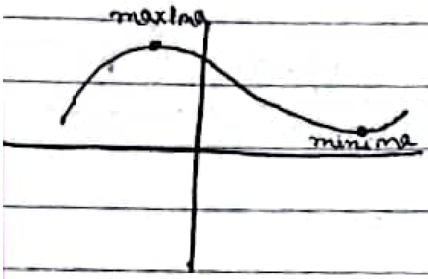
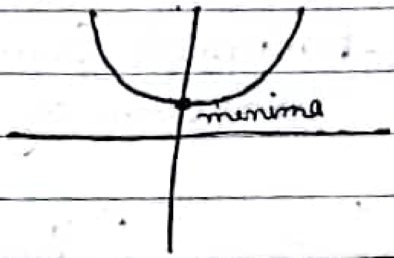
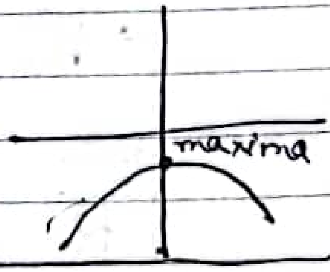
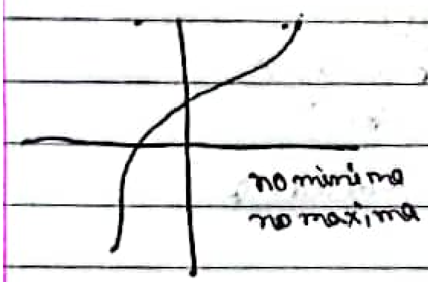
* Gradient Descendent *

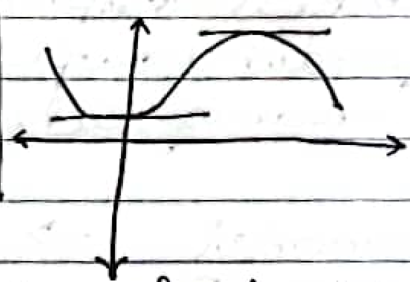








Now lets create slope line in maxima & minima



~~slope~~ slope is always '0' (zero)

$f(x) = x^2 - 3x + 2$, find minima of $f(x)$

$\because$ slope of $f(x) = 0$ @ minima

$\frac{df}{dx} = 0$

$2x - 3 = 0$

$\because \frac{d x^a}{dx} = a x^{a-1}$

$\because \frac{d\, a x^b}{dx} = a b x^{b-a}$

$x = 3/2 = 1.5$

@ $x = 1.5$ slope of $f(x)$ is '0'

So, this $x = 1.5$ can be minima or maxima

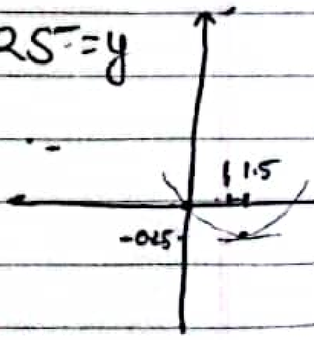
$f(x) = x^2 - 3x + 2$

$f(1.5) = (1.5)^2 - 3 \times 1.5 + 2 = -0.25 = y$

$f(1) = 1^2 - 3 \times 1 + 2 = 0$

So, $f(1) > f(1.5)$

$\therefore$ So $f(1.5)$ is minima



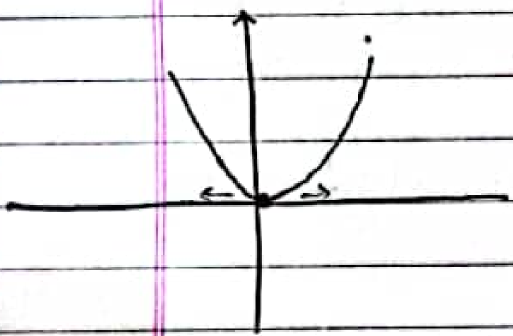
DATE : __/__/__
PAGE : ______

In linear eq^n we try to minimize

$$m^*, c^* = \underset{m,c}{\arg\min} \left\{ \sum (y_i - (mx_i + c))^2 \right\}$$

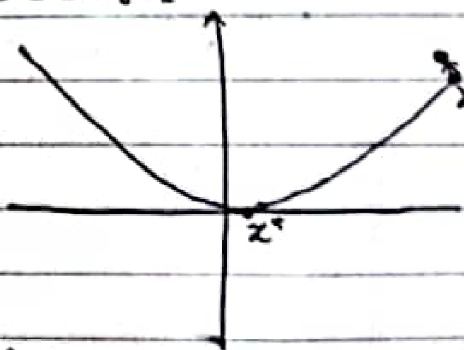
i.e., $f(m,c) = \sum (y_i - (mx_i + c))^2$

to perform this $f(m,c)$ with very fast iterative computation we use gradient ~~decend~~ descent method



- minima Right side : slope is (+) ve
- minima left side : slope is (−)ve
- As you move towards minima → slope reduce
- moving away from minima → slope increases

So in G.D. :-



→ pick @ random an initial point = x_{i0}

→ pick another x_i+1, such that x_i+1 is closer to x^*

$\therefore$ $x_i+1 = (+)x_i - (+ve)$ for +ve x_i

$x_i+1 = \underbrace{(-)x_i}_{x_i} - \underbrace{(-ve)}_{[\text{slope}]_{x_i}}$, for −ve x_i

→ $x_i+1 = x_i - r\left[\dfrac{df}{dx}\right]_{x_i}$ $\because r$ = learning rate

→ Terminate → slope = 0 or

$x_K - x_{K+1}$ is very very slow.

DATE : __/__/__
PAGE : ______

In linear regression sum of square error:-

→ $m^*, c^* = \arg\min \{ f(m, c) \}$

$\because f(m,c) = \sum (y_i - (mx + c)^2)^2$

→ pick initial m_i & $c_i = 0$

$m_{i+1} = m_i - r \left[\frac{\partial f}{\partial m} \right]_{m_i}$ $\because \partial$ (dom) = d (differentiation)

$c_{i+1} = c_i - r \left[\frac{\partial f}{\partial c} \right]_{c_i}$

should assume value of $r = 0.001$

$\left[\frac{\partial f}{\partial m} \right]_{m_i} = \nabla_m f = \frac{2}{N} \sum_{i=1}^{n} \left(y_i - (mx_i + c) \right) * (-x_i)$

$\nabla_c f = \frac{1}{N} \sum_{i=1}^{N} (y_i - (mx_i + c) * (-1)$

so,

$\nabla_m f = \frac{2}{N} \sum_{i=1}^{N} \{ (y_i - \hat{y}_i) * (-x_i) \}$ $\because mx_i + c = \hat{y}$