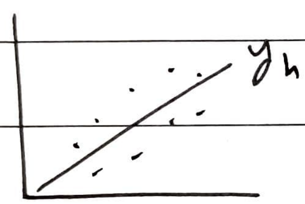
Linear Regression

$y_h = mx + c$ (y - hypothesis)

$m \rightarrow$ Slope, coefficient

$c \rightarrow$ Intercept, bias



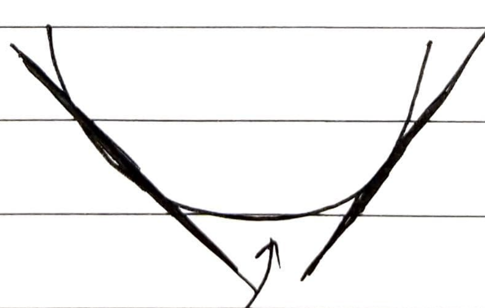
$\text{error} = (y_a - y_h)$ ($\because y_a$ = y-actual)

$$e^2 = (y_a - y_h)^2$$

$$\frac{\partial e^2}{\partial m} = \frac{\partial (y_a - mx - c)^2}{\partial m}$$

$$= 2(y_a - mx - c)(-x)$$

$$\frac{\partial e^2}{\partial c} = 2(y_a - mx - c)(-1)$$



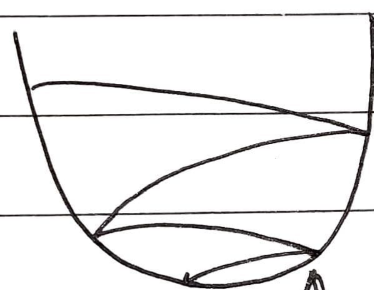
Need to reach here for min error.

$$\Delta m = -\frac{\delta e^2}{\delta m} \times lr$$ (learning rate)

$$\Delta c = -\frac{\delta e^2}{\delta c} \times lr$$

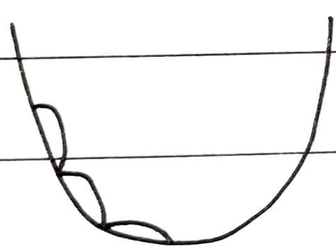
New slope, $m = m + \Delta m$

New ~~slop~~ c (bias), $c = c + \Delta c$



Δ value should decrease with time

Obtuse angle.



Acute angle

Closed form :

$$y_n = mx + c$$

$$y_n = X\theta$$

$$= m_1 x_1 + m_2 x_2 + \dots + c$$

$$= \begin{bmatrix} 1 & x_1 & x_2 & x_3 & \dots \end{bmatrix} \begin{bmatrix} c \\ m_1 \\ m_2 \\ \vdots \end{bmatrix}$$

$$\underset{100 \times 5}{\begin{bmatrix} & \end{bmatrix}} \underset{5 \times 1}{\begin{bmatrix} & \end{bmatrix}} = \underset{100 \times 1}{\begin{bmatrix} & \end{bmatrix}}$$

100 examples,
5 features

Error:

$$e = (X\theta - y_n)$$

$$\frac{\partial e^2}{\partial \theta} = 2(X\theta - y)X \quad (\text{Say } y_n = y)$$

$$\Delta\theta = -\frac{\partial e^2}{\partial \theta} \times lr$$

$$e = (X\theta - y)$$

$$e^2 = (X\theta - Y)^2$$

$$= (X\theta - Y)^T (X\theta - Y)$$

$$= (\theta^T X^T - Y^T)(X\theta - Y)$$

$$e^2 = \Sigma(\theta^T \underbrace{X^T X}\theta - \theta^T X^T Y - Y^T X\theta + Y^T Y)$$

when, $\frac{de^2}{d\theta} = 0$, then best value of θ is found.

<u>Rules:</u>

$\frac{\partial(\theta A)}{\partial\theta} = A^T$	$\frac{\partial(\theta^T A)}{\partial\theta} = A$

$$\frac{\partial(\theta^T A\theta)}{\partial\theta} = 2A\theta$$

$$\frac{de^2}{d\theta} = 0 \Rightarrow 2X^T X\theta - 2X^T Y = 0$$

$$\boxed{\theta = (X^T X)^{-1} X^T Y}$$

$\theta \rightarrow 5, 1$

$\theta^T \rightarrow 1, 5$

Close form:

Final meaningful data but valid for small data.

- statistic (one pt. update)
- complete (all pre-programs)
- minibatch (big data) (can't do close form)

* Can make complete to minibatch by passing chunks batches to gradient descend.

(i) **Epoch** means one pass over the full training set.

(ii) **Batch** means that you use all your data to compute the gradient during one iteration.

(iii) **Minibatch** means you only take a subset of all your data during one iteration.

* One iteration means single ~~step~~ update step.