Gradient descent

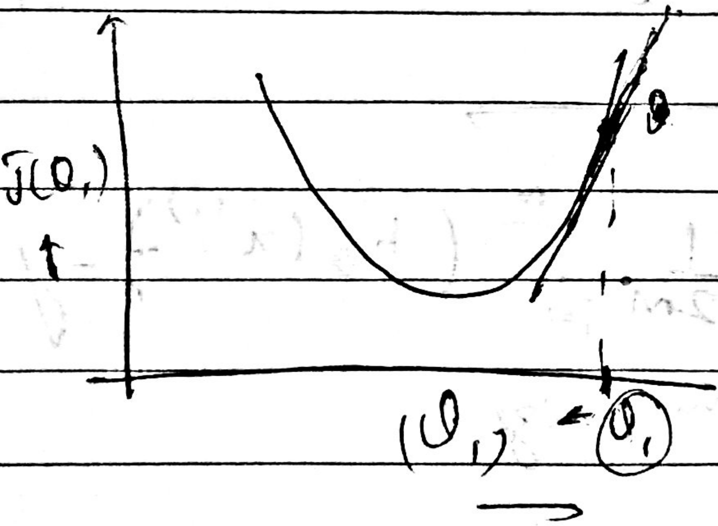
$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

learning rate (huge steps downhill)

$j = 0$ & $j = 1$

<u>Simultaneously</u> update θ_0 & θ_1

$\alpha \rightarrow$ determines how are your steps.



$J(\theta_1) = 1$

$$\theta_1 = \theta_1 - \boxed{\alpha \frac{d}{d\theta_1} J(\theta_1)}$$

≥ 0 derivative = slope

$\therefore \theta_1 = \theta_1 - \alpha(\text{+ve no})$

therefore, θ_1 will move towards left.

→ If α is too small, it will take tiny steps

→ If α is too large, it may overshoot.

Gradient descent algorithm.

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$ (until the convergence)

for ($j = 1$ & $j = 0$ simultaneously)

Linear Regression Model.

$h_\theta(x) = \theta_0 + \theta_1 x$ (hypothesis)

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

squared error cost function.

$$\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) = \frac{\partial}{\partial \theta_0} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

$$= \frac{\partial}{\partial \theta_0} \frac{1}{2m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)} \right)^2$$

Gradient descent algo

$$\theta_0\ j = 0 : \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)$$

$$\theta_1\ j = 1 : \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) \cdot x^{(i)}$$

'Batch' gradient descent.

~~θ_0~~ :

$2 = \theta_0 + \theta_1(4)$

$4\theta_1 + \theta_0 = 2$

$2\theta_1 + \theta_0 = 1$

$2\theta_1 = 1$

$\theta_1 = 0.5$

$h(6) = \theta_0 + \theta_1 x$

$= -1 + 2(6)$

$= 11$

Tower of Hanoi

2 + 6 L 10

$$\begin{bmatrix} 1 & 0 & 3 \\ 2 & 1 & 5 \\ 3 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 6 \\ 2 \end{bmatrix} , \begin{bmatrix} \cancel{7} \\ 13 \\ \end{bmatrix}$$

2 + 8 = 10

House	3 competing hypotheses
2104	$h_\theta(x) = -40 + 0.25x$
1416	$h_\theta(x) = 200 + 0.1x$
1534	$h_\theta(x) = -150 + 0.4x$
852

10 hypothesis

$$\begin{bmatrix} 1 & 2104 \\ 1 & 1416 \\ 1 & 1534 \\ 1 & 852 \end{bmatrix} \begin{bmatrix} -40 & 200 & -150 \\ 0.25 & 0.1 & 0.4 \end{bmatrix}$$

★ Matrix multiplication is not commutative.

$$A \times B \neq B \times A.$$

★ Associative law.

$$A \times (B \times C) = (A \times B) \times C.$$

Matrix is associative.

★ for identity matrix,

$$A \cdot I = I A = A$$