Gradient Descent

$$\nabla J(\theta) = \frac{\partial J}{\partial \theta} = -\frac{1}{m}\sum_{i=1}^{n}(y_i - \hat{y}_i)\cdot x_i$$

$$\text{MSE} \quad \frac{1}{2m}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

objective is to reduce the loss

$y = mx + c$

$$\theta_{new} = \theta - \eta \frac{\partial J}{\partial \theta}$$

$$\hat{y}_i = \theta_0 + \theta_1 X$$

X	Y_{true}
1	2
2	2.8
3	3.6

$\theta_0 =$ $\eta = 0.1$

$\theta_0 = 0.5$

$\theta_1 = 0.5$

$\hat{y} = ?$

$$\hat{y}_1 = 0.5 + 0.5 * 1 = 1$$

$$\hat{y}_2 = 0.5 + 0.5 * 2 = 1.5$$

$$\hat{y}_3 = 0.5 + 0.5 * 3 = 2$$

$$MSE = \frac{1}{2 \times 3}\left((2-1)^2 + (2.8-1.5)^2 + (3.6-2)^2\right)$$

$MSE_1 = 0.875$

$MSE_2 = 1.75$

For θ_0 do

$$\frac{\partial J}{\partial \theta_0} = -\frac{1}{m}\sum_{i=1}^{n}(y_i - \hat{y}_i) = -\frac{1}{3}((2-1) + (2.8-1.5) + (3.6-2)) = -1.3$$

For θ_1 do

$$\frac{\partial J}{\partial \theta_1} = -\frac{1}{m}\sum_{i=1}^{n}((y_i - \hat{y}_i) * x) = -\frac{1}{3}((2-1)*1 + (2.8-1.5)*2 + (3.6-2)*3) = -2.8$$

$$\theta_{0_{new}} = \theta_0 - \eta \frac{\partial J}{\partial \theta_0} = 0.5 - 0.1[-1.3] = 0.63$$

$$\theta_{1_{new}} = \theta_1 - \eta \frac{\partial J}{\partial \theta_1} = 0.5 - 0.1[-2.8] = 0.78$$

$$\hat{y} = 0.63 + 0.78x$$

Calculate MSE = ? 0.18

= 0.372

Stochastic Gradient Descent

X	Y
1	3
2	5
3	7
4	9
5	11

$y = wx + b$

Initial

$w = 0.5$

$b = 0.1$

$\eta = 0.01$

for first tuple

$x = 1 \quad y = 3$

$\hat{y} = (0.5) * 1 + 0.1 = 0.6$

compute the gradient

$$\frac{\partial J}{\partial w} = (y_i - \hat{y}_i) \times x = (3 - 0.6) \times 1$$
$$= 2.4$$

$$\frac{\partial J}{\partial b} = (y_i - \hat{y}_i) = (3 - 0.6)$$
$$= 2.4$$

Update weight $= w_0 - \eta \frac{\partial J}{\partial w} = 0.5 - (0.01) \times 2.4$
$= 0.476$

Update bais $= b - \eta \frac{\partial J}{\partial b} = 0.1 - (0.01) \times 2.4$
$= 0.076$

Given data

A	B	y_{true}
3	1	4
5	0	6
2	4	10
3	6	8
2	4	2

Linear model = $\hat{y} = \theta_0 + \theta_1 A + \theta_2 B$

$\theta_0 = 0, \quad \theta_1 = 1 \quad \theta_2 = 2$

$\alpha = 0.1$