3-1-6 Gradient Descent Method 2

$$\theta = \theta - a \frac{\partial J(\theta)}{\partial \theta} \rightarrow \text{update } \theta.$$

Model : $\hat{y} = \theta_1 x_1 + \theta_0$ parameters (θ_1, θ_0)

Gradient

$$\nabla_{\vec{\theta}} \mathcal{L}^{(i)}(\vec{\theta}) = \left(\frac{\partial \mathcal{L}^{(i)}(\vec{\theta})}{\partial \theta_1}, \frac{\partial \mathcal{L}^{(i)}(\vec{\theta})}{\partial \theta_0} \right)$$

$$\nabla_{\vec{\theta}} J(\vec{\theta}) = \left(\frac{\partial J(\vec{\theta})}{\partial \theta_1}, \frac{\partial J(\vec{\theta})}{\partial \theta_0} \right)$$

Gradient descent

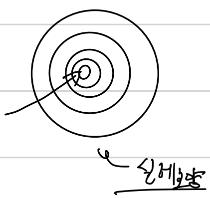
$$(\theta_1, \theta_0) = (\theta_1, \theta_0) - a \nabla_{\vec{\theta}} \mathcal{L}^{(i)}(\vec{\theta})$$

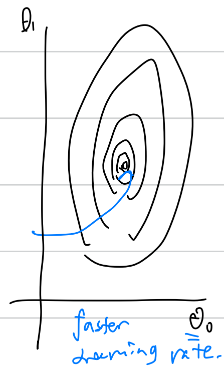
$$(\theta_1, \theta_0) = (\theta_1, \theta_0) - a \nabla_{\vec{\theta}} J(\vec{\theta})$$

$$\downarrow$$

$$\theta_1 = \theta_1 - a \frac{\partial J(\vec{\theta})}{\partial \theta_1}$$

$$\theta_0 = \theta_0 - a \frac{\partial J(\vec{\theta})}{\partial \theta_0}$$





Model : $\hat{y} = \theta_n x_n + \theta_{n-1} x_{n-1} + \dots \theta_1 x_1 + \theta_0$

Parameters : $(\vec{w}, b) = (\underbrace{\theta_n, \theta_{n-1}, \theta_{n-2} \cdots \theta_1, \theta_0}_{\text{'n+1' learning parameters}})$

plug in

$\downarrow$

Loss, Cost.

$\vdots$

Mathematically same function shape even though we have many parameters.

$\theta_1, \theta_2 \ldots \theta_n$ are 'respectively' updated.

<Key Expressions>

$$\vec{\theta} := \vec{\theta} - a \nabla_{\vec{\theta}} \mathcal{L}^{(i)}(\vec{\theta})$$

$$\vec{\theta} := \vec{\theta} - a \nabla_{\vec{\theta}} J(\vec{\theta})$$