Machine Learning

Dr. Mehran Safayani

safayani@iut.ac.ir

safayani.iut.ac.ir

https://www.aparat.com/mehran.safayani

https://github.com/safayani/machine_learning_course

Department of Electrical and computer engineering, Isfahan university of technology, Isfahan, Iran

Cost Function

$h_\theta(x_i) = \theta_0 + \theta_1 x_i$

$J(\theta_0 , \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2$

Minimize $J(\boldsymbol{\theta_0} , \boldsymbol{\theta_1})$

$\boldsymbol{\theta_0} , \boldsymbol{\theta_1}$

If $J(\theta_1) = (\theta_1 - 2)^2$

$\frac{dJ(\theta_1)}{d\theta_1} = 0$ → $\frac{dJ(\theta_1)}{d\theta_1} = 2 (\theta_1 - 2) = 0$ → $\theta_1 = 2$

Gradient Descent

Minimize $J(\theta_0, \theta_1)$

θ_0, θ_1

Minimize $J(\theta_0, \theta_1, \dots, \theta_n)$

$\theta_0, \theta_1, \dots, \theta_n$

Repeat until convergence: {

For j=0,…,n

$$\theta_j = \theta_j - \alpha \frac{dJ(\theta_0, \theta_1, \dots, \theta_n)}{d\theta_j}$$

}

α **is learning rate**

Updating all θ_j ***Simultaneously***

Convergence condition:

$$\|\theta^{t+1} - \theta^t\|_2 \leq \varepsilon$$

Gradient Descent

Correct form

temp0 = $\boldsymbol{\theta_0} - \alpha \frac{dJ(\boldsymbol{\theta_0}, \boldsymbol{\theta_1})}{d\boldsymbol{\theta_0}}$

temp1 = $\boldsymbol{\theta_1} - \alpha \frac{dJ(\boldsymbol{\theta_0}, \boldsymbol{\theta_1})}{d\boldsymbol{\theta_1}}$

$\boldsymbol{\theta_0}$ = temp0

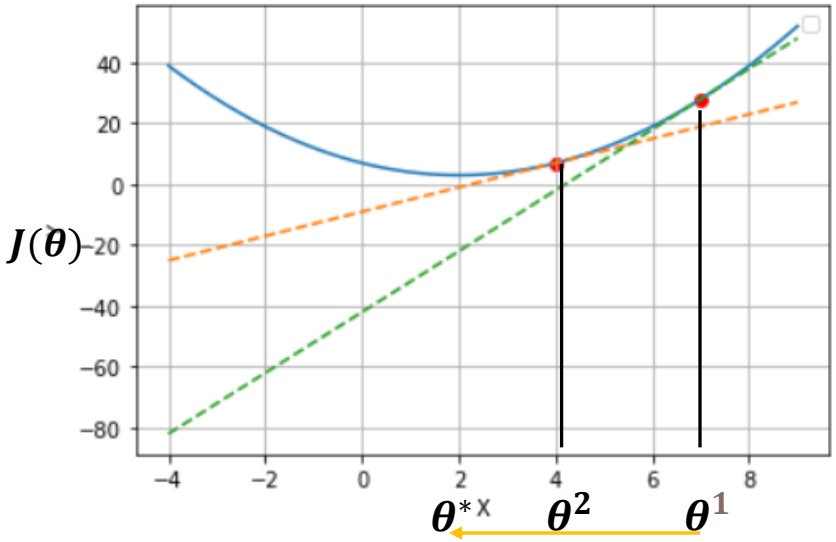
$\boldsymbol{\theta_1}$ = temp1

Incorrect form

$\boldsymbol{\theta_0} = \boldsymbol{\theta_0} - \alpha \frac{dJ(\boldsymbol{\theta_0}, \boldsymbol{\theta_1})}{d\boldsymbol{\theta_0}}$

$\boldsymbol{\theta_1} = \boldsymbol{\theta_1} - \alpha \frac{dJ(\boldsymbol{\theta_0}, \boldsymbol{\theta_1})}{d\boldsymbol{\theta_1}}$

Gradient Descent



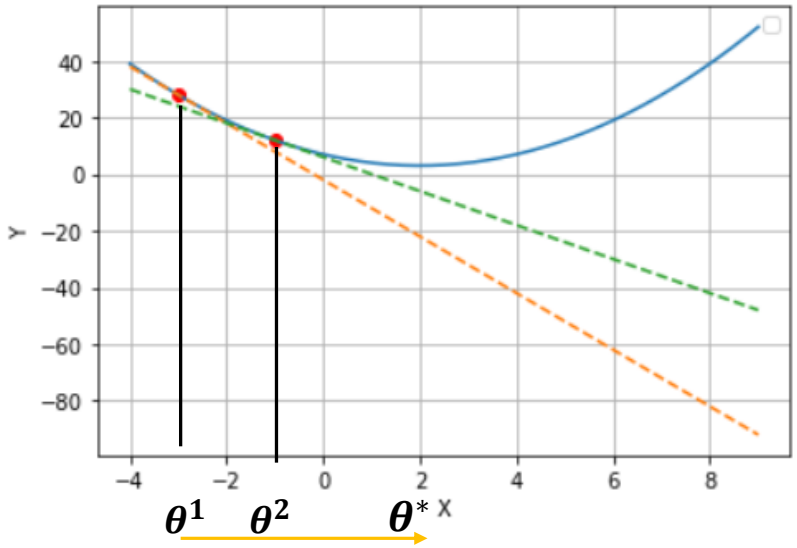
خطوط مماس نشان داده شده دارای شیب یا مشتق مثبت هستند.
درنتیجه:

$$\frac{dJ(\theta^1)}{d\theta^1} > 0\ ,\ \alpha > 0 \Rightarrow \alpha\frac{dJ(\theta^1)}{d\theta^1} > 0$$

$$\Rightarrow \theta^2 = \theta^1 - \alpha d\theta^1$$

θ کوچکتر میشود و به سمت چپ حرکت میکنیم.

Gradient Descent



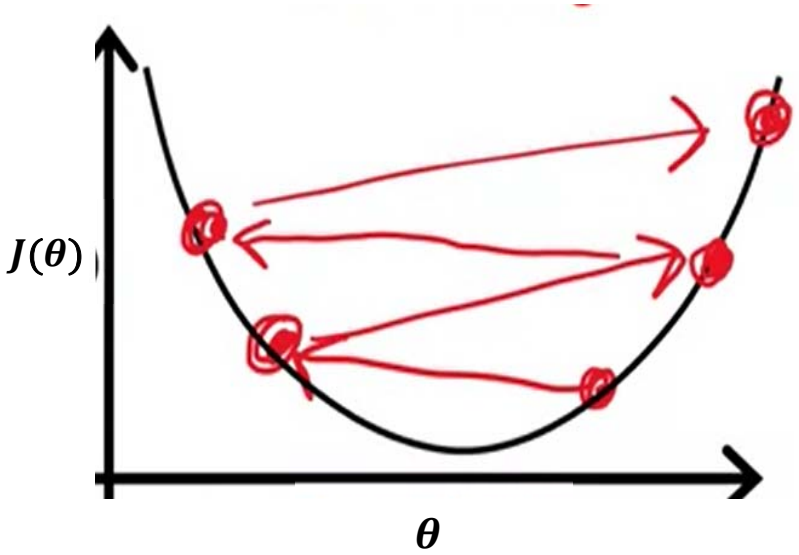
خطوط مماس نشان داده شده دارای شیب یا مشتق مثبت هستند.
درنتیجه:

$$\frac{dJ(\theta^1)}{d\theta^1} < 0\ ,\ \alpha > 0 \Rightarrow \alpha \frac{dJ(\theta^1)}{d\theta^1} < 0$$

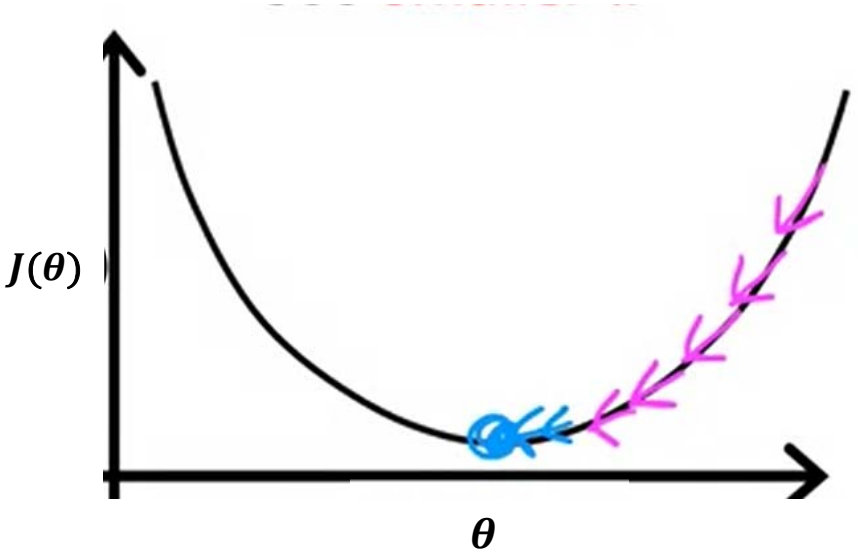
$$\Rightarrow \theta^2 = \theta^1 - \alpha d\theta^1$$

θ بزرگتر میشود و به سمت راست حرکت میکنیم.

Choosing Learning Rate

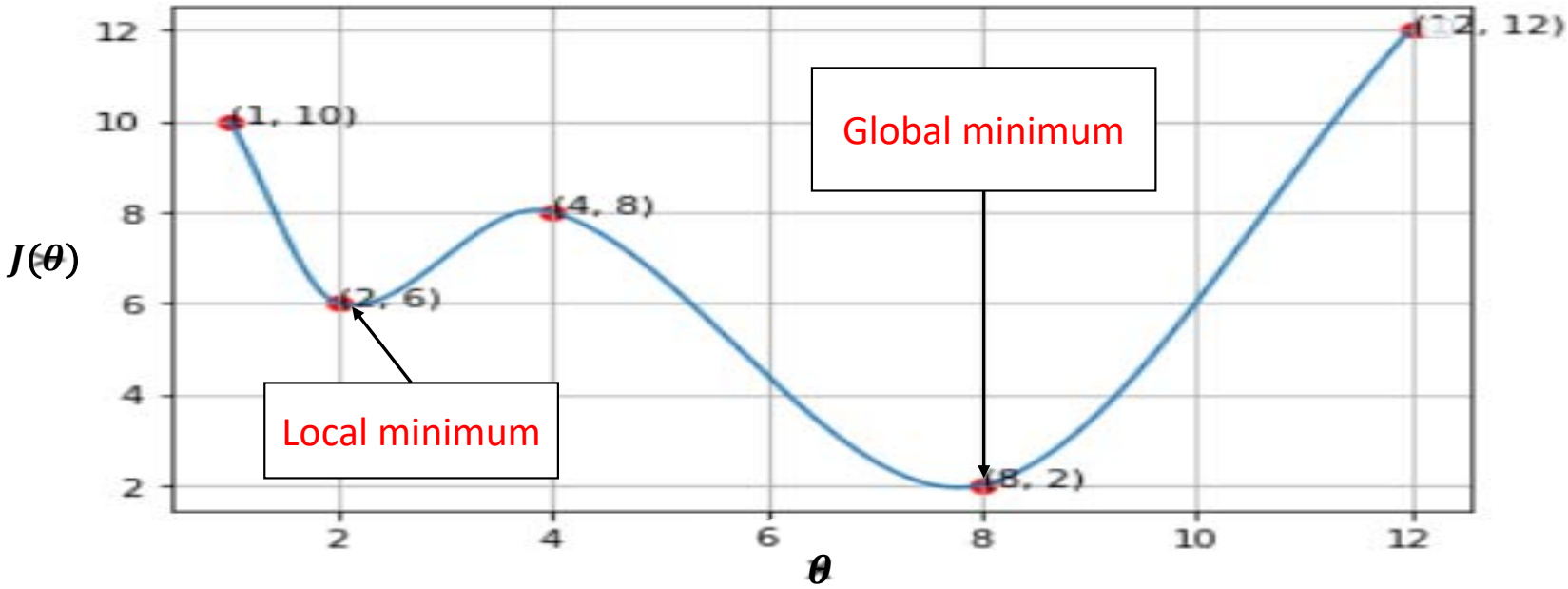


α is too large



α is small

Gradient Descent Weakness



Linear regression model

$$h_\theta(x_i) = \theta_0 + \theta_1 x_i$$

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2$$

$$\frac{dJ(\boldsymbol{\theta_0}, \boldsymbol{\theta_1})}{d\boldsymbol{\theta_0}} = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)$$

$$\frac{dJ(\boldsymbol{\theta_0}, \boldsymbol{\theta_1})}{d\boldsymbol{\theta_1}} = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)\, x_i$$

Linear regression model

Repeat until convergence: {

$$\boldsymbol{\theta_0} = \boldsymbol{\theta_0} - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)$$

$$\boldsymbol{\theta_1} = \boldsymbol{\theta_1} - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)\, x_i$$

}

بروز رسانی همزمان

$$\theta^t = \begin{bmatrix} \boldsymbol{\theta_0} \\ \boldsymbol{\theta_1} \end{bmatrix} \quad , \quad \theta^{t+1} = \begin{bmatrix} \boldsymbol{\theta_0} \\ \boldsymbol{\theta_1} \end{bmatrix} \quad , \quad d\theta = \begin{bmatrix} d\boldsymbol{\theta_0} \\ d\boldsymbol{\theta_1} \end{bmatrix}$$

Convergence condition:

- $\|\theta^{t+1} - \theta^t\|_2 = \sqrt[2]{(\theta_0^{t+1} - \theta_0^t)^2 + (\theta_1^{t+1} - \theta_1^t)^2} < \varepsilon$
- $\|d\theta\|_2 < \varepsilon$

Batch Gradient Descent

$$\frac{\boldsymbol{dJ(\theta)}}{\boldsymbol{d\theta}} = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)$$

Batch Gradient Descent

$\theta_0 \leftarrow random$, $\theta_1 \leftarrow random$

Repeat until convergence: {

J $\leftarrow$ 0 , $d\theta_1 \leftarrow$ 0 , $d\theta_0 \leftarrow$ 0

For i = 1 to m:

$h_\theta(x_i) = \theta_0 + \theta_1 x_i$

j += $(h_\theta(x_i) - y_i)^2$

$d\theta_1$ += 2 $(h_\theta(x_i) - y_i)\, x_i$

$d\theta_0$ += 2 $(h_\theta(x_i) - y_i)$

J /= 2m

$d\theta_1$ /= 2m

$d\theta_0$ /= 2m

$\theta_1 = \theta_1 - \alpha d\theta_1$

$\theta_0 = \theta_0 - \alpha d\theta_0$

}