



Machine Learning

Linear regression
with one variable

Gradient descent
intuition

Gradient descent algorithm

repeat until convergence {

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

}

Learning Rate

rate

derivative

(simultaneously update
 $j = 0$ and $j = 1$)

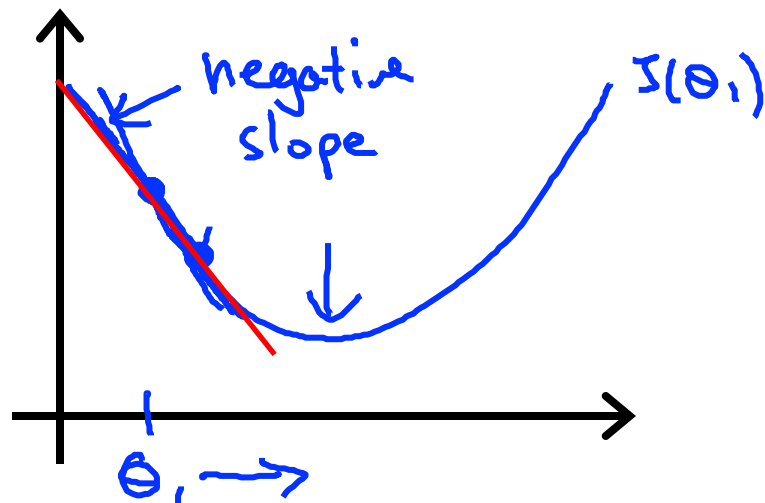
$$\min_{\theta_1} J(\theta_1)$$

$$\theta_1 \in \mathbb{R}.$$



$$\theta_1 := \theta_1 - \alpha \left(\frac{\partial}{\partial \theta_1} J(\theta_1) \right) \geq 0$$

Diagram illustrating the update rule for θ_1 when the slope is positive. The derivative $\frac{\partial}{\partial \theta_1} J(\theta_1)$ is positive, and the update rule is $\theta_1 := \theta_1 - \alpha \left(\frac{\partial}{\partial \theta_1} J(\theta_1) \right)$, where α is a positive number. The result is ≥ 0 .



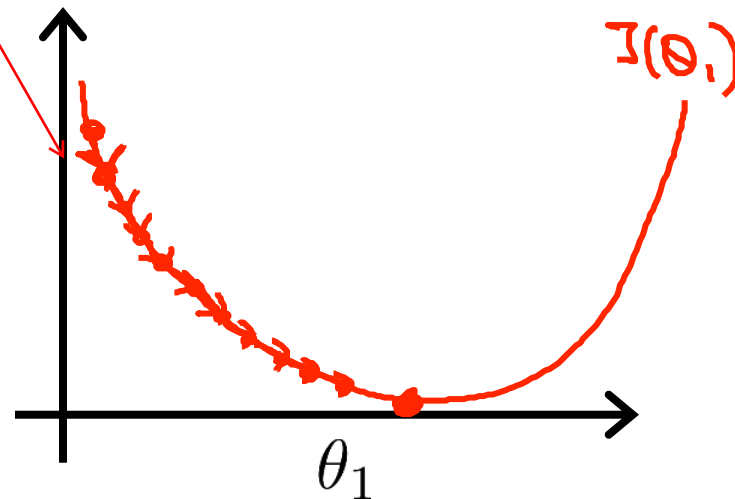
$$\theta_1 := \theta_1 - \alpha \left(\frac{\partial}{\partial \theta_1} J(\theta_1) \right) \leq 0$$

Diagram illustrating the update rule for θ_1 when the slope is negative. The derivative $\frac{\partial}{\partial \theta_1} J(\theta_1)$ is negative, and the update rule is $\theta_1 := \theta_1 - \alpha \left(\frac{\partial}{\partial \theta_1} J(\theta_1) \right)$, where α is a positive number. The result is ≤ 0 .

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

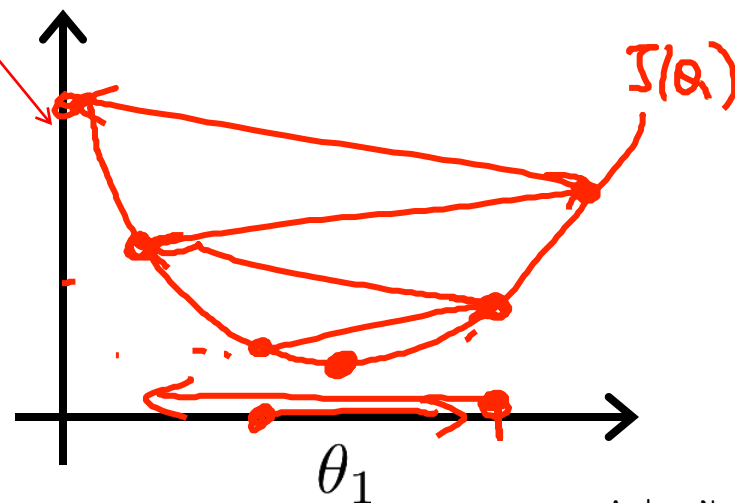
- If α is too small, gradient descent can be slow.

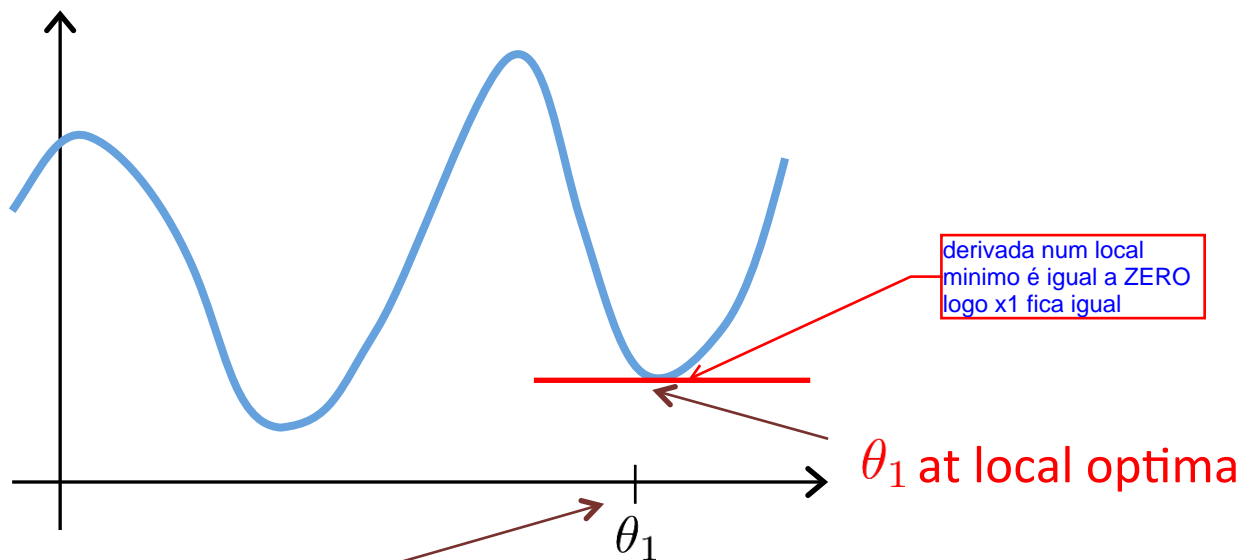
pequenas passos



- If α is too large, gradient descent can overshoot the minimum. It may fail to converge, or even diverge.

grandes passos





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

$$x_1 = x_i - x^*0$$

Gradient descent can converge to a local minimum, even with the learning rate α fixed.

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

a medida que nos aproximamos do mínimo a derivada é cada vez menor

As we approach a local minimum, gradient descent will automatically take smaller steps. So, no need to decrease α over time.

