



Machine Learning

Linear Regression with multiple variables

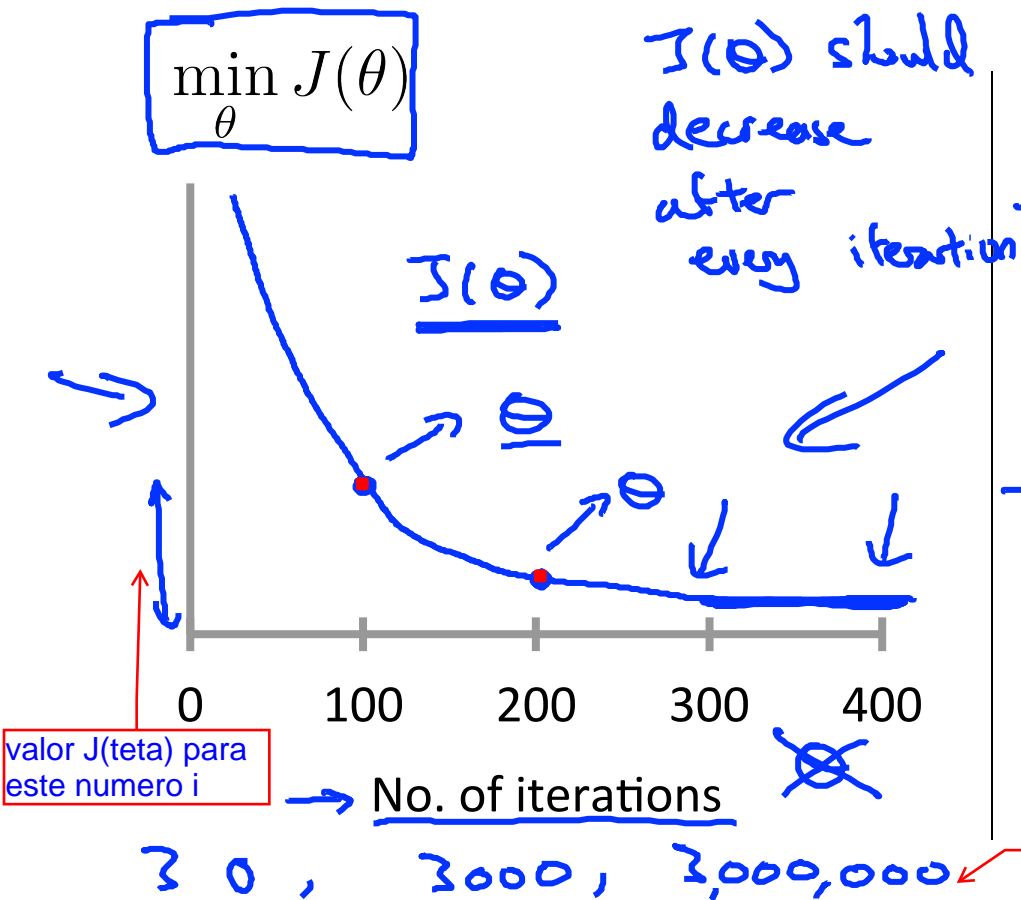
Gradient descent in
practice II: Learning rate

Gradient descent

$$\rightarrow \theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

- “Debugging”: How to make sure gradient descent is working correctly.
- How to choose learning rate α .

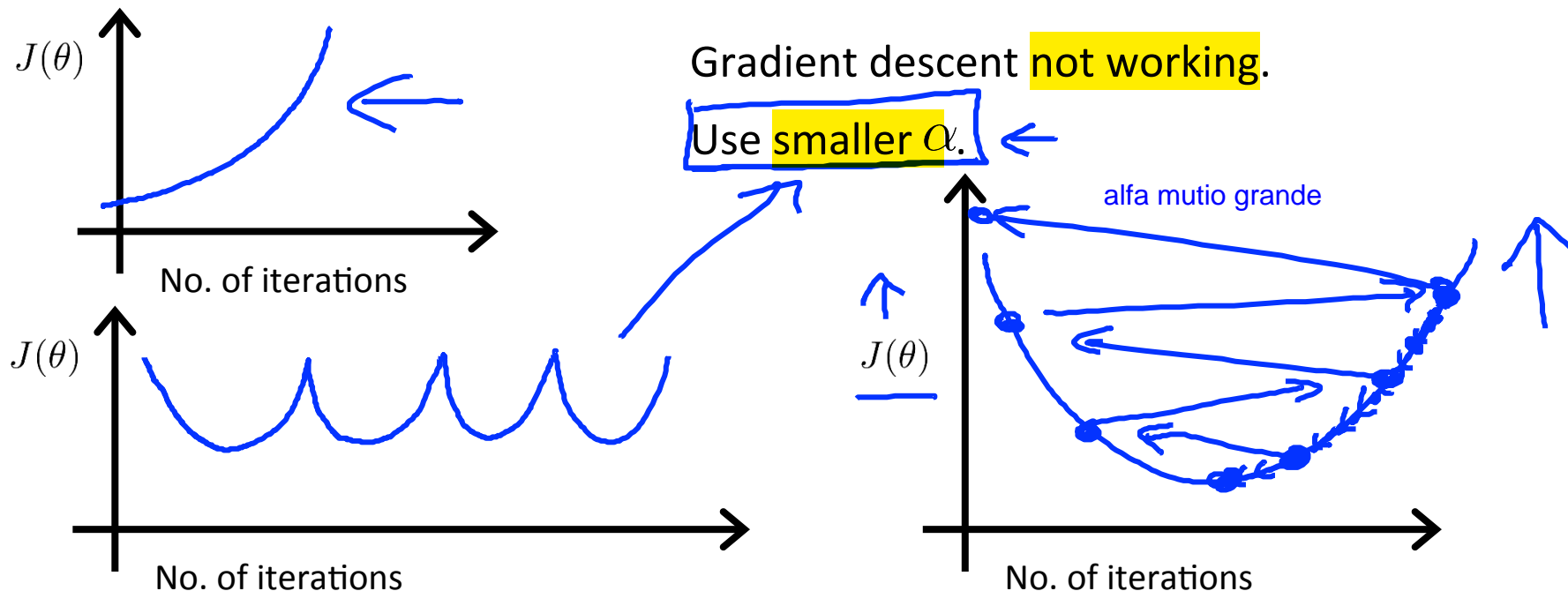
Making sure gradient descent is working correctly.



→ Example automatic convergence test:

→ Declare **convergence** if $J(\theta)$ decreases by less than 10^{-3} in one iteration.

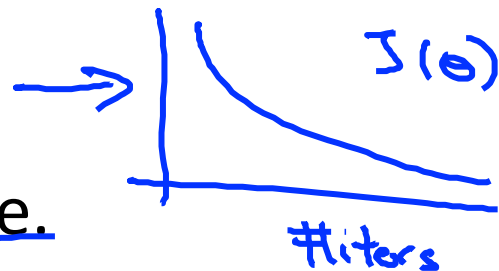
Making sure gradient descent is working correctly.



- For sufficiently small α , $J(\theta)$ should decrease on every iteration.
- But if α is too small, gradient descent can be slow to converge.

Summary:

- If α is too small: slow convergence.
- If α is too large: $J(\theta)$ may not decrease on every iteration; may not converge. (Slow converge also possible.)



To choose α , try

$\dots, 0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 1, \dots$

Handwritten blue annotations below the sequence of values: upward arrows point to 0.001, 0.03, 0.1, and 1. Curved arrows connect 0.001 to 0.003, 0.003 to 0.01, 0.01 to 0.03, and 0.03 to 0.1. Each curved arrow is labeled with $\approx 3\times$, indicating a threefold increase between consecutive values.