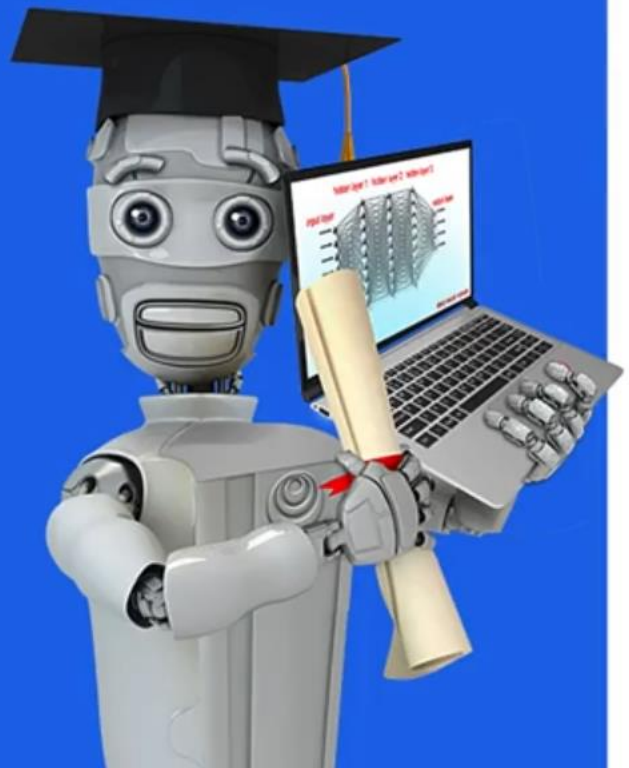
Stanford
ONLINE

DeepLearning.AI



Training Linear Regression

Learning Rate

$$w = w - \alpha \frac{d}{dw} J(w)$$

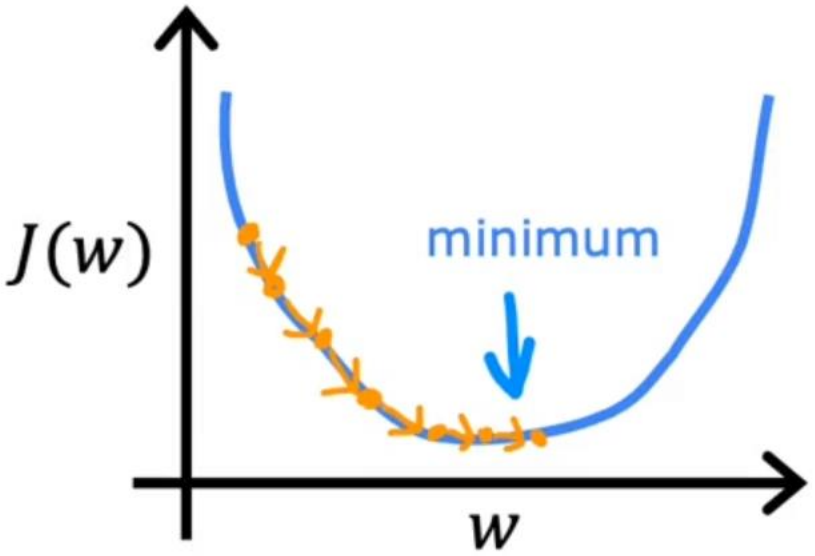
If α is too small…

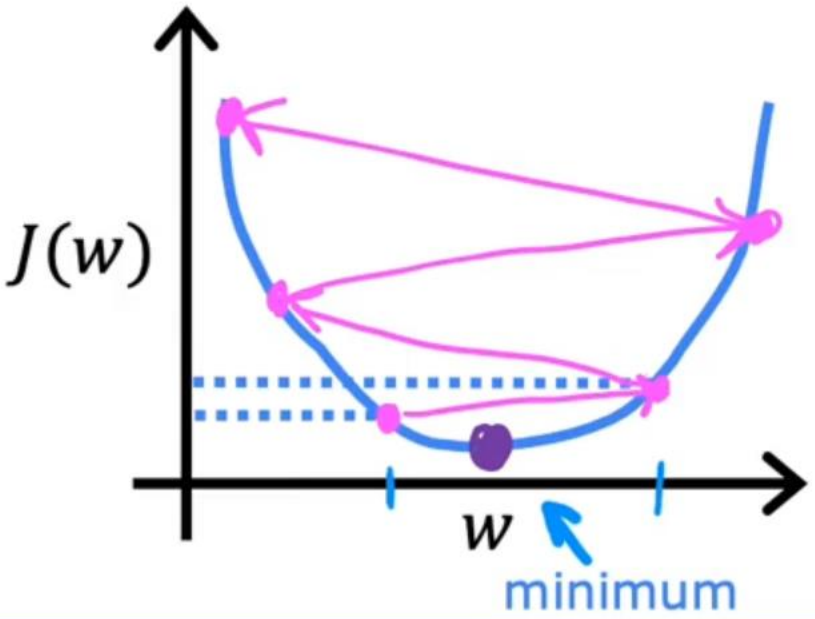
Gradient descent may be slow.

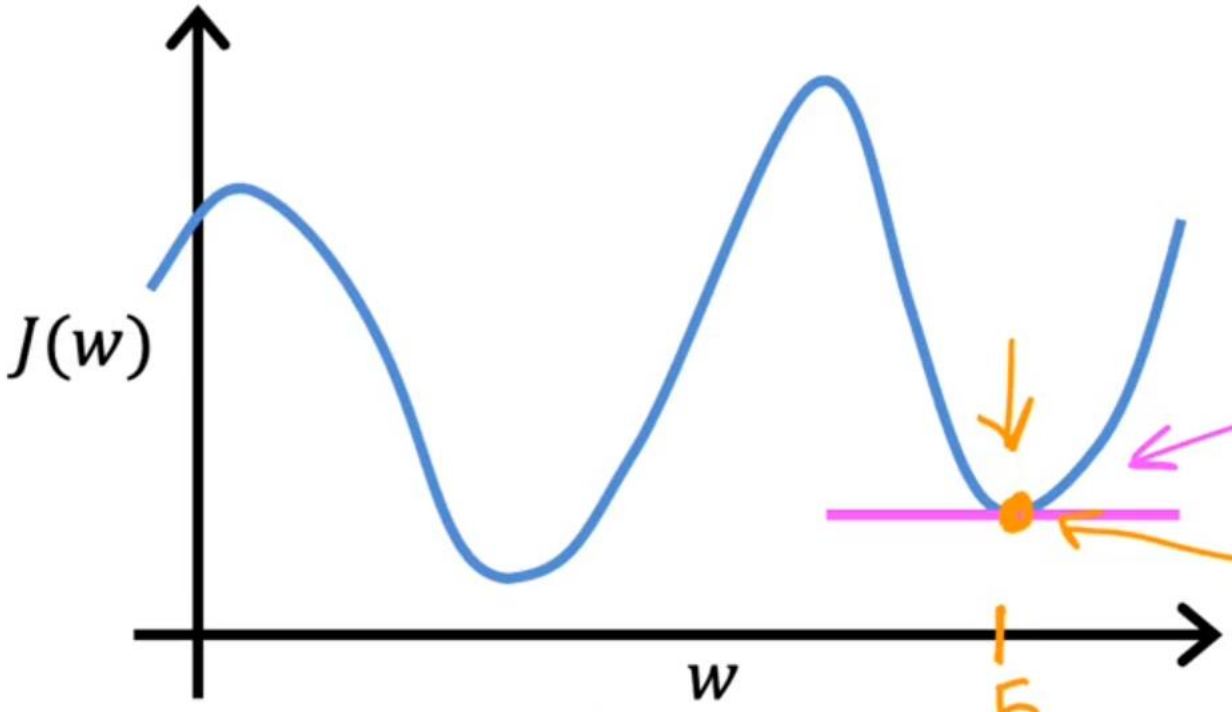
If α is too large…

Gradient descent may:

- Overshoot, never reach minimum
- Fail to converge, diverge







$$w = w - \alpha \frac{d}{dw} J(w)$$

0.1

=0

$5 - 0.1 \cdot 0$

$w = w - \alpha \cdot 0$

$5 \quad 5$

$w = w$

Can reach local minimum with fixed learning rate

α

$$w = w - \alpha \frac{d}{dw} J(w)$$

smaller

not as large

large

Near a local minimum,
- Derivative becomes smaller
- Update steps become smaller

Can reach minimum without decreasing learning rate α