<Intuition2> - learning rate

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

i) If α (learning rate) is too small

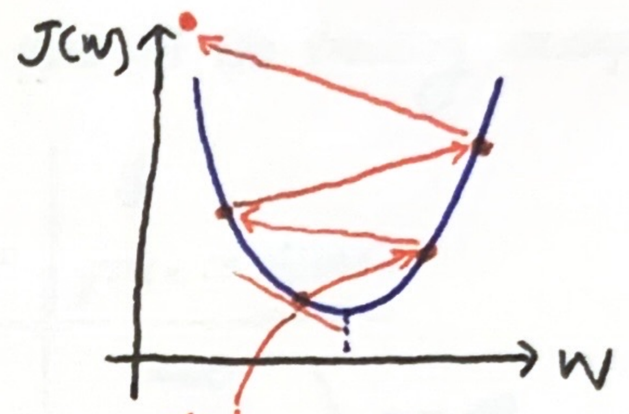
⇒ Gradient Descent works slowly

= Take too long time before getting close to the minimum

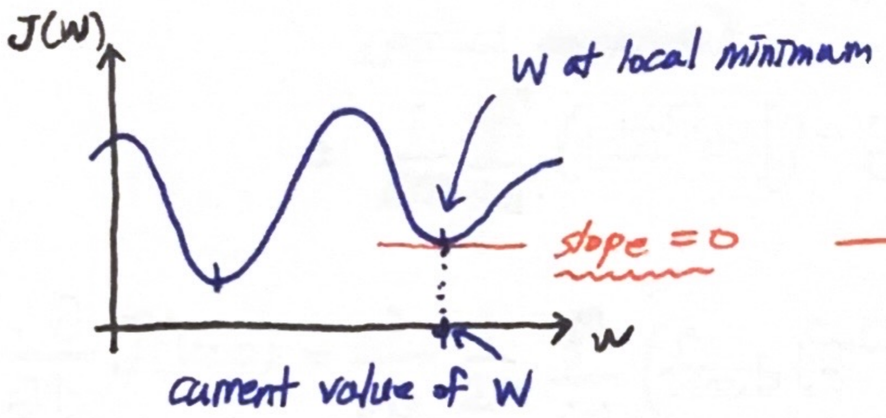
ii) If α (learning rate) is too large

⇒ Gradient Descent may overshoot or never reach minimum

= fail to converge, diverge



<Intuition 3> - more than 1 local minimums



$$w := w - \alpha \frac{\partial}{\partial w} J(w) = w - \alpha \cdot 0$$

$\frac{\partial}{\partial w} J(w) = 0$

∴ update new w = previous w

⇒ If w is already at local minimum, gradient descent leaves w unchanged because derivative term is zero so there is no update of w

it also means...

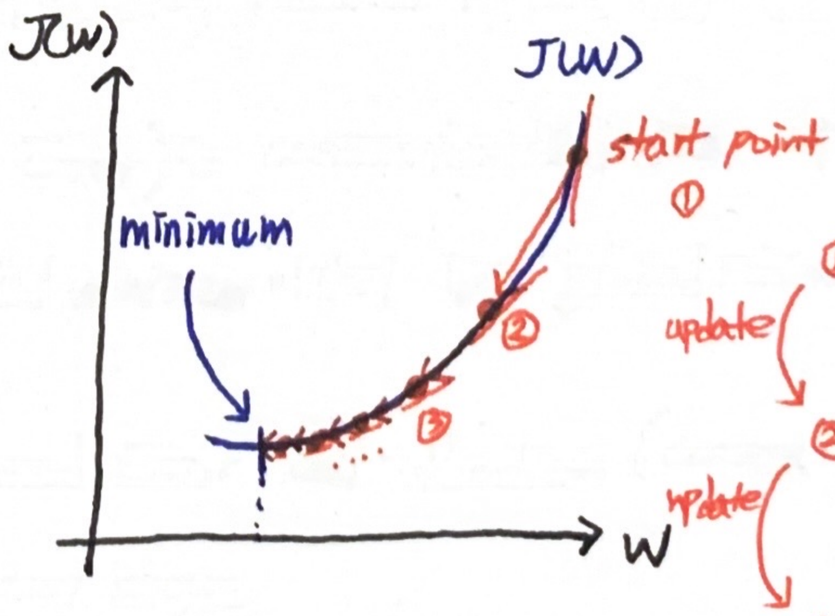
"Can reach local minimum with fixed α" (learning rate)

- As we approach a local minimum

→ derivative becomes smaller

⇒ update steps become smaller

⇒ "can reach minimum without" decreasing learning rate α



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

① $\frac{\partial}{\partial w} J(w)$ = slope is large

② $\frac{\partial}{\partial w} J(w)$ = slope gets smaller

③ $\frac{\partial}{\partial w} (J(w)$ = slope gets smaller

⋮

"Derivate becomes smaller automatically"

⇒ "Gradient descent becomes smaller"