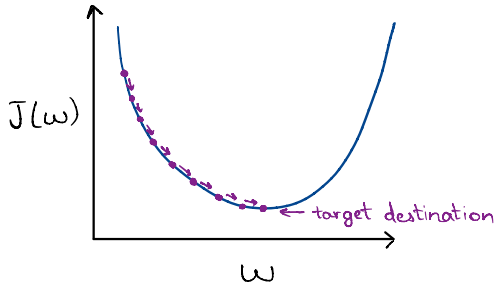


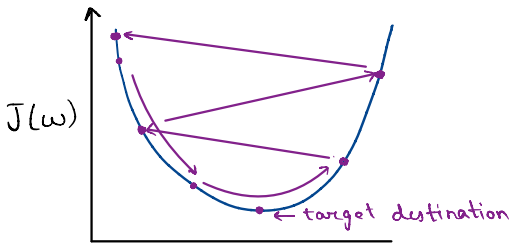
In the algorithm, α is called the learning rate.

If α is too small,



In the algorithm due to the learning rate being so small, even though we reach the target destination the algorithm takes too much time for doing that.

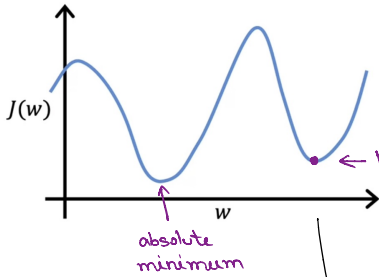
If α is too big,



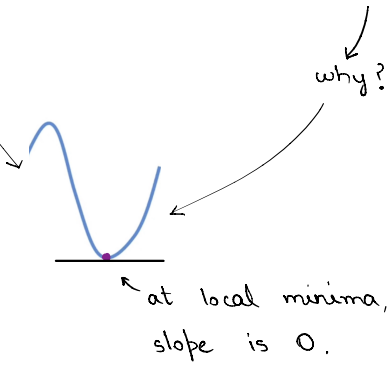
w will overshoot and pass the target distance.

In some cases, it may even diverge due to failure of convergence.

Local Minima



If algorithm has already reached local minima then it won't do anything (i.e. w will remain unchanged)



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

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

$$\Rightarrow w = w - \alpha \cdot 0$$
$$w = w$$

Now, you maybe wondering, how does gradient descent reach a local minimum if the learning rate is fixed.

→ That is because, $\frac{\partial}{\partial w} J(w)$ keeps changing based on steepness.

