



deeplearning.ai

# Optimization Algorithms

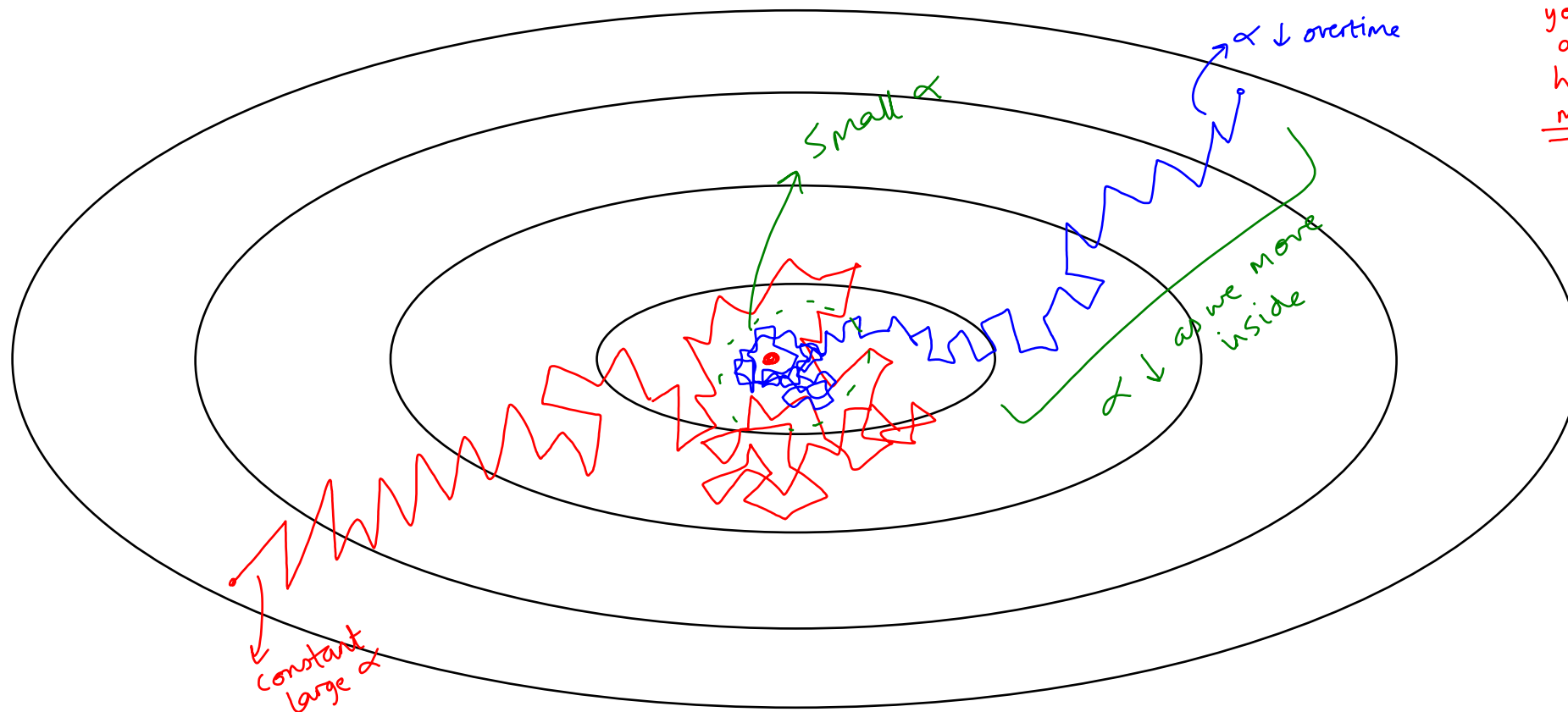
---

## Learning rate decay

# Learning rate decay

— Slowly  $\downarrow$  learning Rate over time  
as you reach the optima

- with large  $\alpha$ , you will oscillate near the optima
- with smaller  $\alpha$  near optima, you will still oscillate in a Region near the optima, however, the Region will be much smaller



# Learning rate decay

1 epoch = 1 pass through training set  $\Rightarrow$  training has multiple passes through the set

$$\alpha = \frac{1}{1 + \underbrace{\text{decay rate} \times \text{epoch num}}_{\substack{\text{Additional hyper param} \\ \text{to tune}}}} \times \alpha_0$$

$$\alpha_0 = 0.2$$
$$\text{decay rate} = 1$$

eg

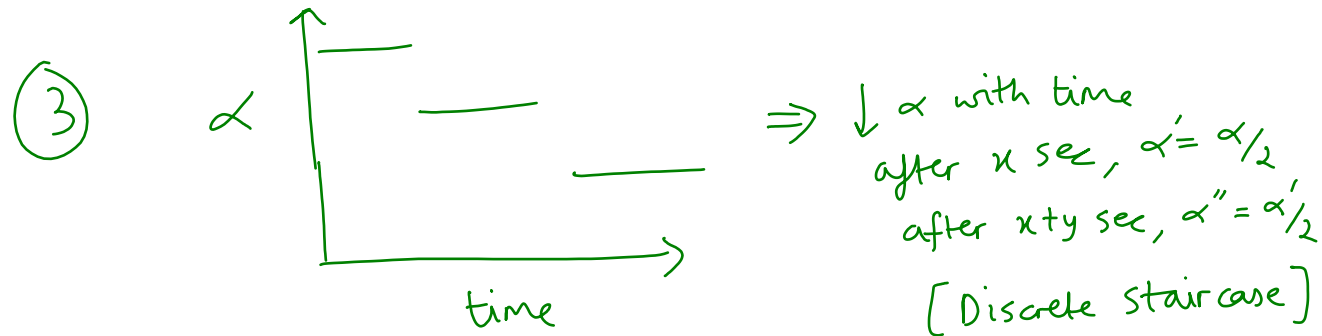
Epoch	$\alpha$
1	0.1
2	0.67
3	0.5
4	0.4
$\vdots$	$\vdots$

$\alpha \downarrow$  as epoch  $\uparrow$

# Other learning rate decay methods

①  $\alpha = (0.95)^{\text{epoch num}} \cdot \alpha_0$  [exponential decay]

②  $\alpha = \frac{k}{\sqrt{\text{epoch-num}}} \cdot \alpha_0$  or  $\frac{k \cdot \alpha_0}{\sqrt{t}}$   
(minibatch number "t")



④ Manual Decay