



deeplearning.ai

# Optimization Algorithms

---

## Exponentially weighted averages

GD 보다 더 빠른 optimization 알고리즘들이 있는데 그걸 이해하려면 exponentially weighted average라는 놈을 알아야 한다.

# Temperature in London

$$\theta_1 = 40^\circ\text{F} \quad 4^\circ\text{C} \quad \leftarrow$$

$$\theta_2 = 49^\circ\text{F} \quad 9^\circ\text{C}$$

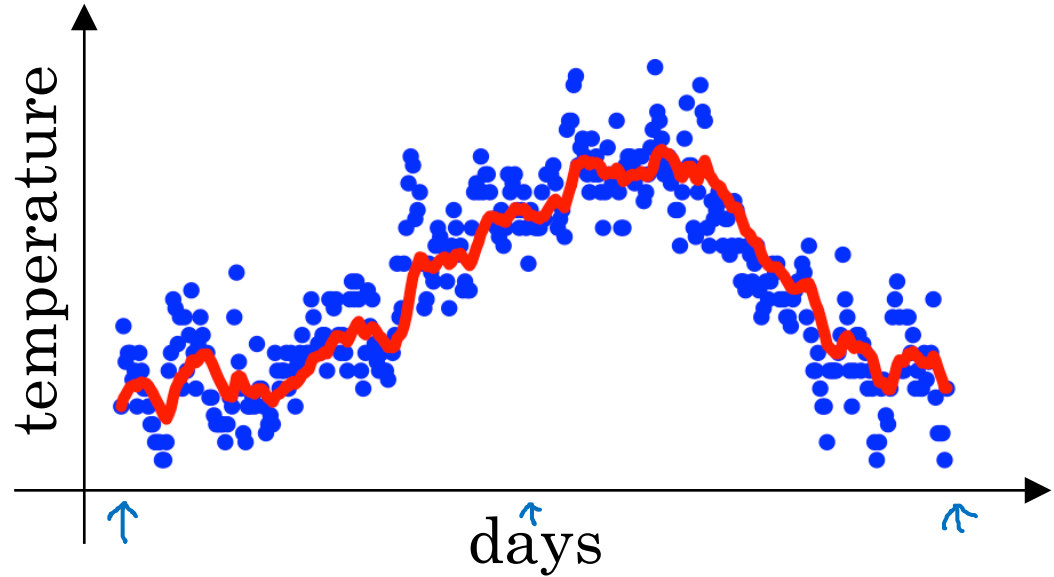
$$\theta_3 = 45^\circ\text{F} \quad \vdots$$

$\vdots$

$$\theta_{180} = 60^\circ\text{F} \quad 15^\circ\text{C}$$

$$\theta_{181} = 56^\circ\text{F} \quad \vdots$$

$\vdots$



$$V_0 = 0 \quad (V_0 \text{는 } 0 \text{으로 초기화})$$

$$V_1 = 0.9 V_0 + 0.1 \theta_1$$

$$V_2 = 0.9 V_1 + 0.1 \theta_2$$

$$V_3 = 0.9 V_2 + 0.1 \theta_3$$

$\vdots$

$$V_t = 0.9 V_{t-1} + 0.1 \theta_t$$

# Exponentially weighted averages <sup>moving</sup>

$$V_t = \beta V_{t-1} + (1-\beta) \theta_t \leftarrow$$

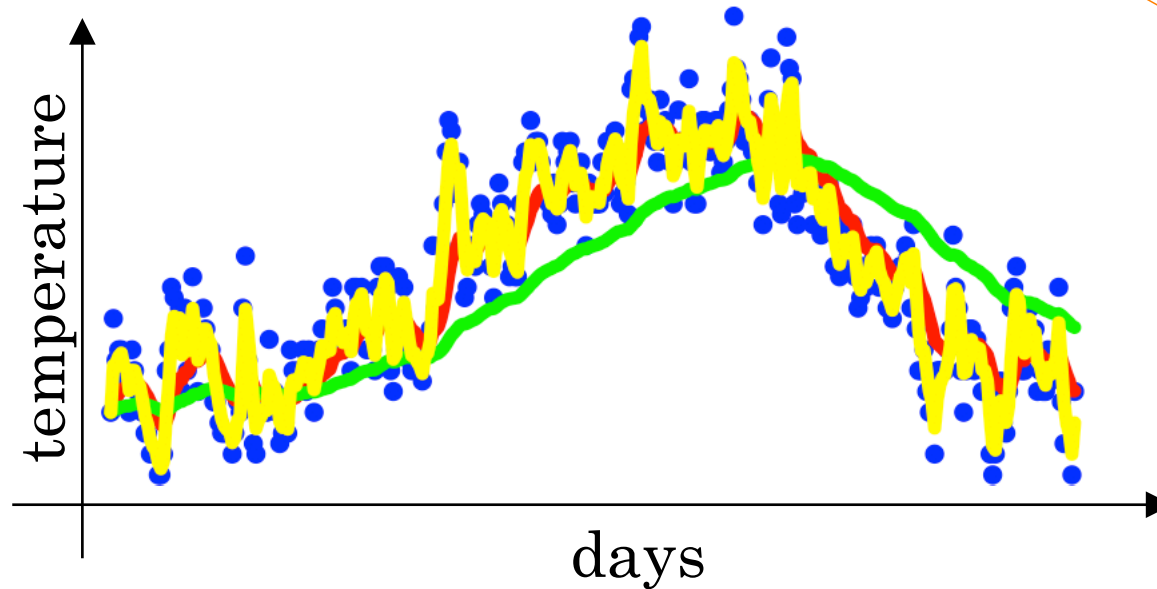
$\beta = 0.9$  :  $\approx 10$  days' temperature.  
( $= \frac{1}{1-0.9}$ )

$\beta = 0.98$  :  $\approx 50$  days

$\beta = 0.5$  :  $\approx 2$  days

$V_t$  is approximately  
as approximately  
average over  
 $\rightarrow \approx \left(\frac{1}{1-\beta}\right)$  days' temperature.

$$\frac{1}{1-0.98} = 50$$



큰 베타값을 쓰겠다는 것은 즉, previous value에 더 가중치를 주겠다는 의미. 그 결과 오늘의 급격한 온도 변화에 민감하지 않게 되고 그림이 smooth하지만 오른쪽으로 shift된 모양으로 그려진다.