



deeplearning.ai

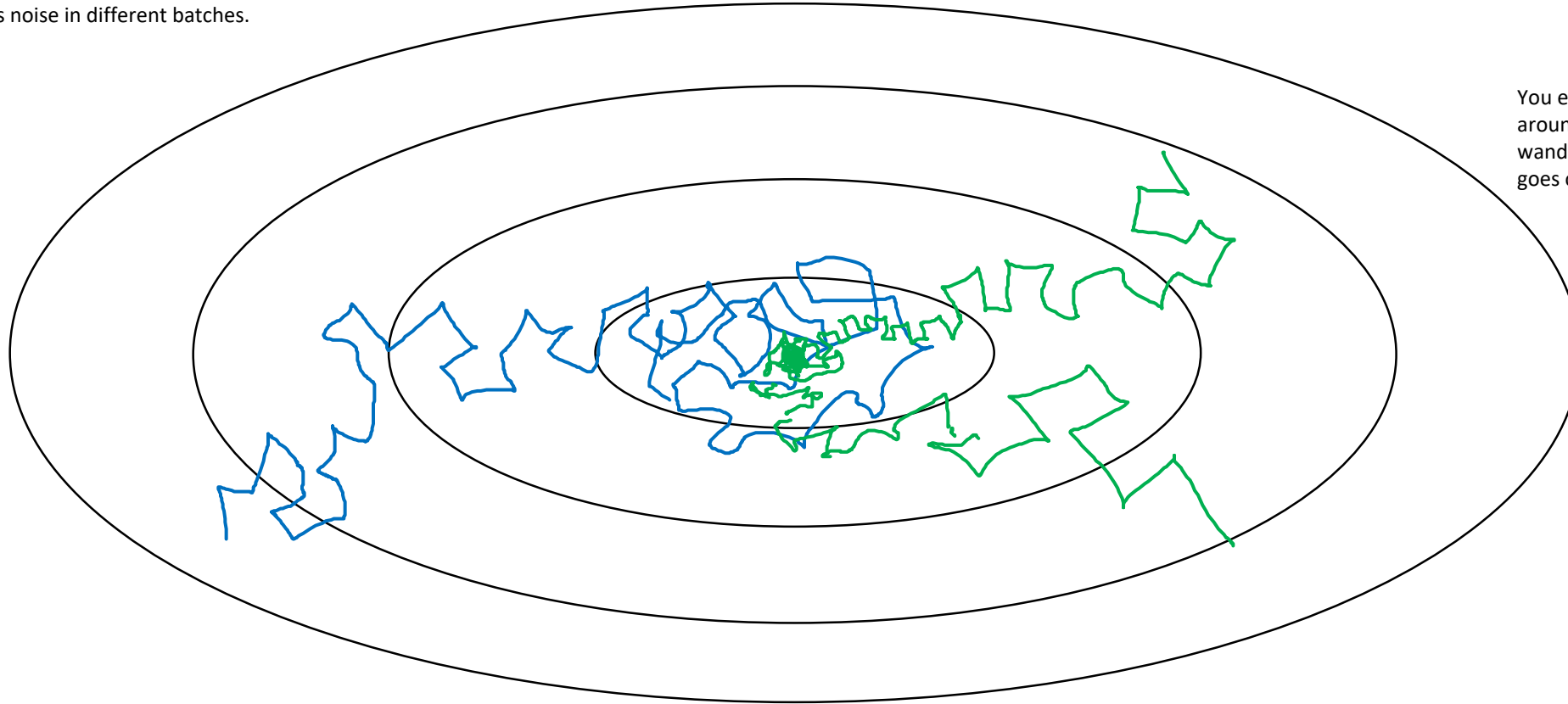
Optimization Algorithms

Learning rate decay

Learning rate decay

Blue line - fixed learning-rate α
Never converge, wandering around the optimal
because there's noise in different batches.

Slowly reduce α



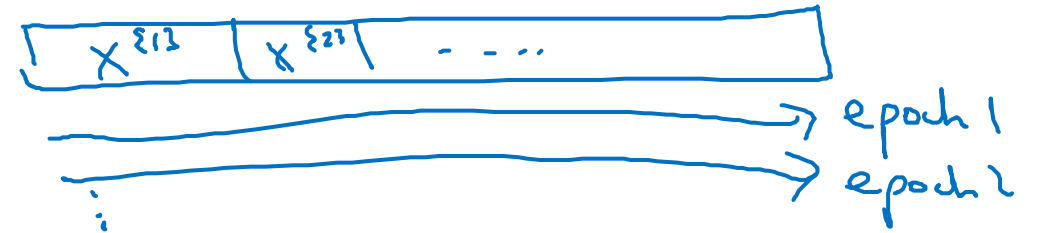
You end up oscillating in a tighter region around this minimum, rather than wandering far away, even as training goes on and on.

Learning rate decay

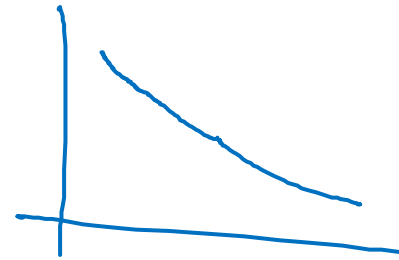
1 epoch = 1 pass through data.

$$\alpha = \frac{1}{1 + \text{decay-rate} * \text{epoch-num}} \alpha_0$$

Epoch	α
1	0.1
2	0.67
3	0.5
4	0.4
\vdots	\vdots




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



Other learning rate decay methods

formula {

$$\alpha = 0.95^{\text{epoch-num}} \cdot \alpha_0 \quad - \text{exponentially decay.}$$
$$\alpha = \frac{k}{\sqrt{\text{epoch-num}}} \cdot \alpha_0 \quad \text{or} \quad \frac{k}{\sqrt{t}} \cdot \alpha_0 \quad t \text{ mini-batch number}$$


discrete staircase

Manual decay.

works if you have only small number of models