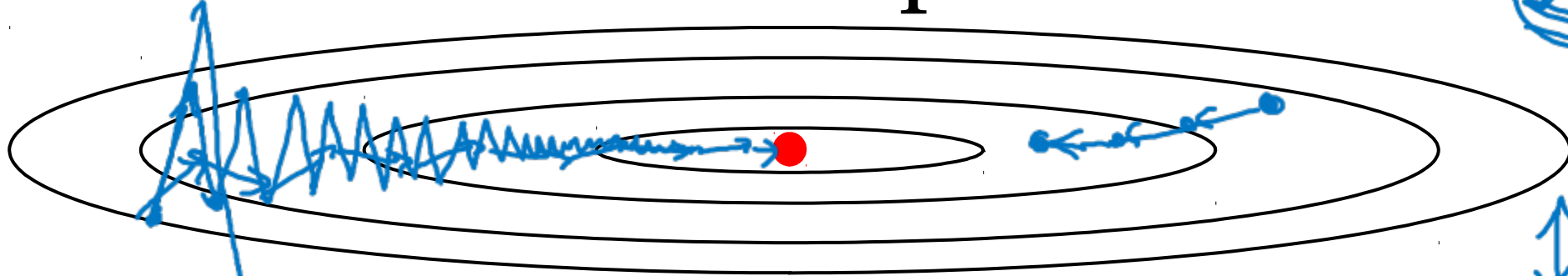**deeplearning.ai**

Optimization Algorithms

Gradient descent with momentum

Gradient descent example

slower learning

faster learning.

Momentum:

On iteration t:

Compute dW, db on current mini-batch.

$V_{dW} = \beta V_{dW} + (1-\beta) dW$

$V_{db} = \beta V_{db} + (1-\beta) db$

friction ⟶ velocity ⟵ acceleration

"$V_\theta = \beta V_\theta + (1-\beta) \theta_t$"

$W := W - \alpha V_{dW}$, $b := b - \alpha V_{db}$

deeplearning.ai

Andrew Ng

Implementation details

$V_{dw} = 0, \quad V_{db} = 0$

On iteration t:

Compute on the current mini-batch

$$\rightarrow v_{dW} = \beta v_{dW} + (1 - \beta)dW \Big\}$$

$$\rightarrow v_{db} = \beta v_{db} + (1 - \beta)db \Big\}$$

$$W = W - \alpha v_{dW}, \quad b = b - \alpha v_{db}$$

$V_{dw} = \beta V_{dw} + \quad dW \leftarrow$

Hyperparameters:

$\beta = 0.9$

average over last $\times$ 10 gradients

Andrew Ng