deeplearning.ai

Optimization
Algorithms

Gradient descent
with momentum

Gradient descent example

slower learning

faster learning
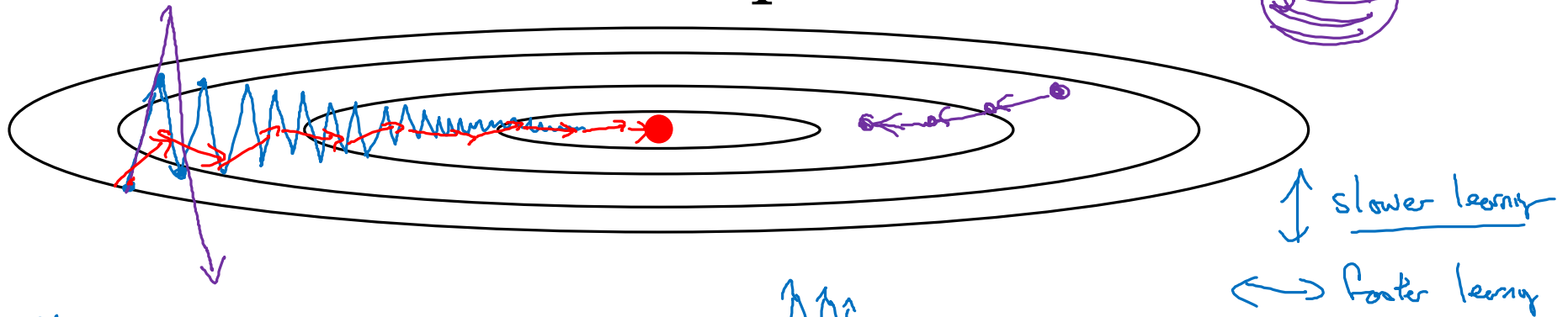
Momentum:

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

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

← acceleration

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

deeplearning.ai

Andrew Ng

Implementation details

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

On iteration t:

    Compute dW, db on the current mini-batch

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

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

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

$v_{dw} = \beta v_{dw} + dW \leftarrow$

$\dfrac{v_{dw}}{1-\beta t}$

Hyperparameters: α, β $\beta = 0.9$

average over last ≈ 10 gradients

Andrew Ng