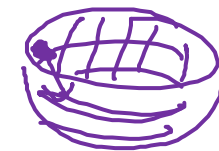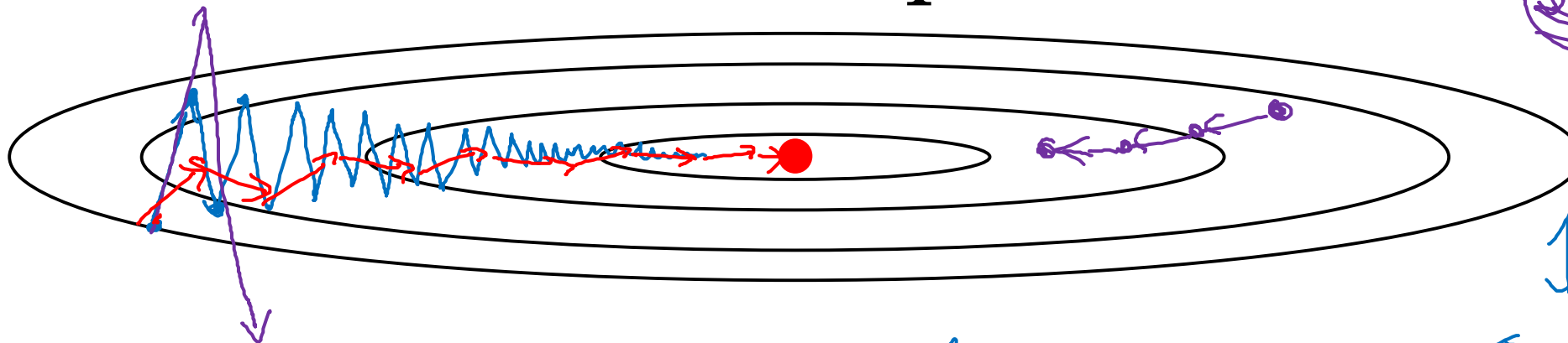Optimization
Algorithms

Gradient descent
with momentum

deeplearning.ai

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 $\longrightarrow$ $\uparrow$ velocity $\uparrow$ acceleration

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

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

Implementation details

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

On iteration t:

 Compute dW, db on the current mini-batch

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

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

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

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

Hyperparameters: $\alpha, \beta \qquad\qquad \beta = 0.9$

average over last ≈ 10 gradients