

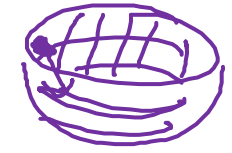
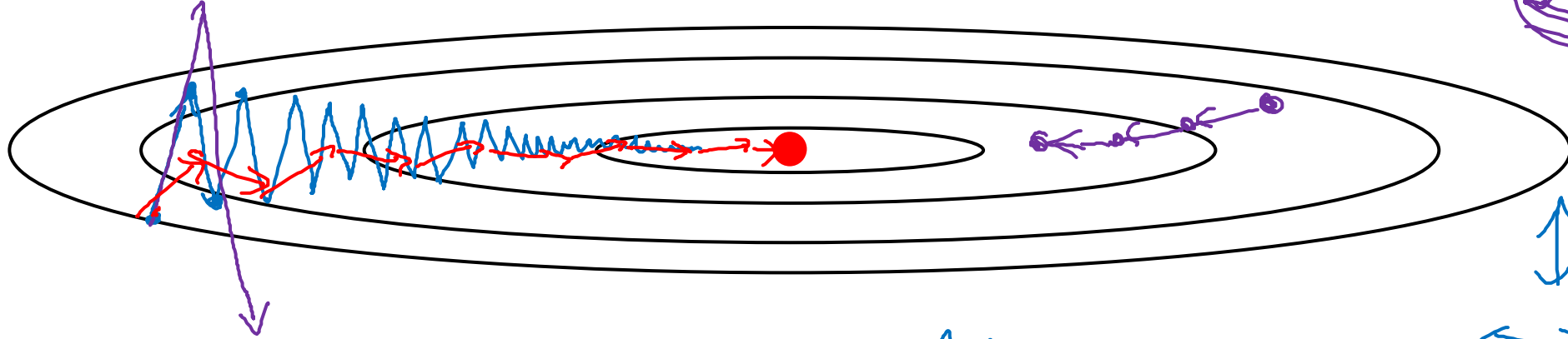


deeplearning.ai

Optimization Algorithms

Gradient descent with momentum

Gradient descent example



↑ slower learning
↔ faster learning.

Momentum:

On iteration t :

Compute dW, db on current mini-batch.

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

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

friction → velocity

acceleration

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



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

Implementation details

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

On iteration t :

Compute dW, db on the current mini-batch

$$\left. \begin{aligned} \rightarrow v_{dW} &= \beta v_{dW} + (1 - \beta) dW \\ \rightarrow v_{db} &= \beta v_{db} + (1 - \beta) db \end{aligned} \right\} \quad \left| \quad \underbrace{v_{dW} = \beta v_{dW} + dW}_{\leftarrow}$$

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

$$\frac{v_{dW}}{1 - \beta^t}$$

Hyperparameters: α, β

$$\underline{\beta = 0.9}$$

average over last ≈ 10 gradients