

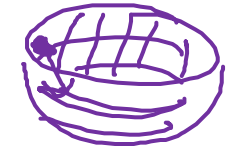
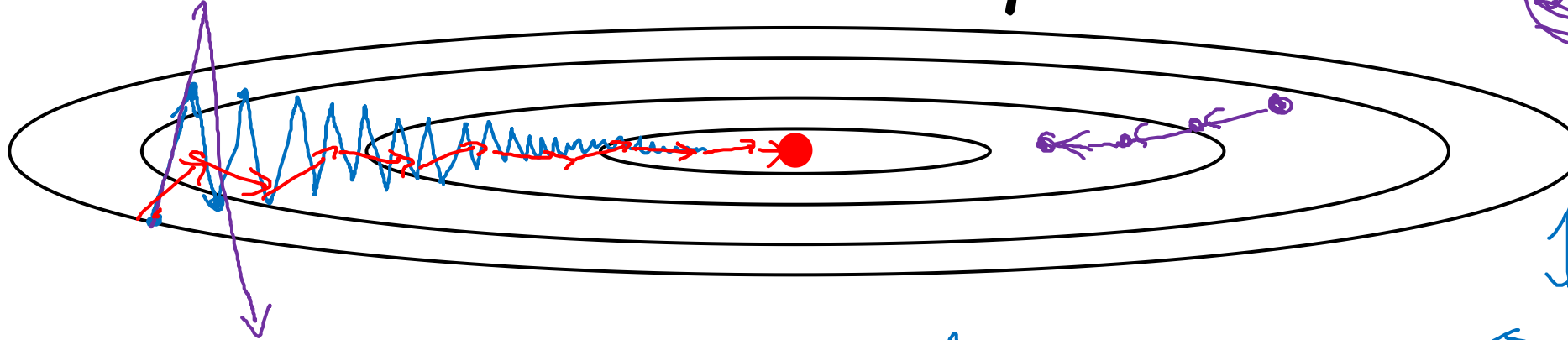


deeplearning.ai

Optimization Algorithms

Gradient descent with momentum

Gradient descent example



↑ slower learning
↔ faster learning

Momentum

On iteration t :

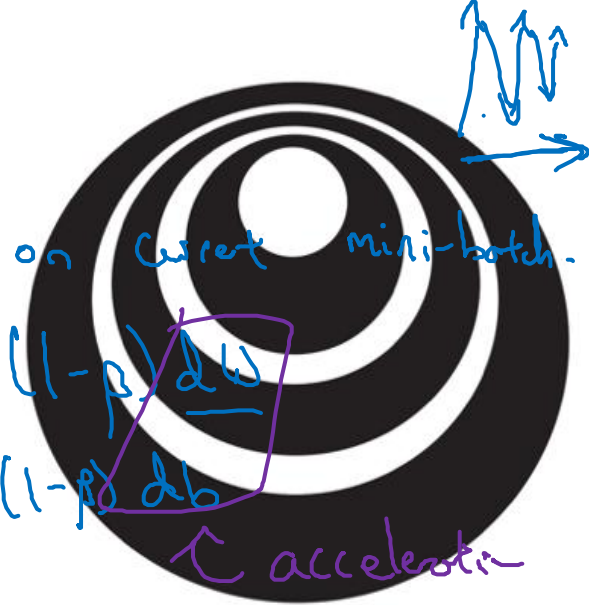
Compute dW, db on current mini-batch.

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

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

friction ↑ velocity

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



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

deelearning.ai

Implementation details

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

On iteration t :

Compute dW, db on the current mini-batch

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

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

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

Hyperparameters: α, β
↑ ↑

$\beta = 0.9$
average over last ≈ 10 gradients