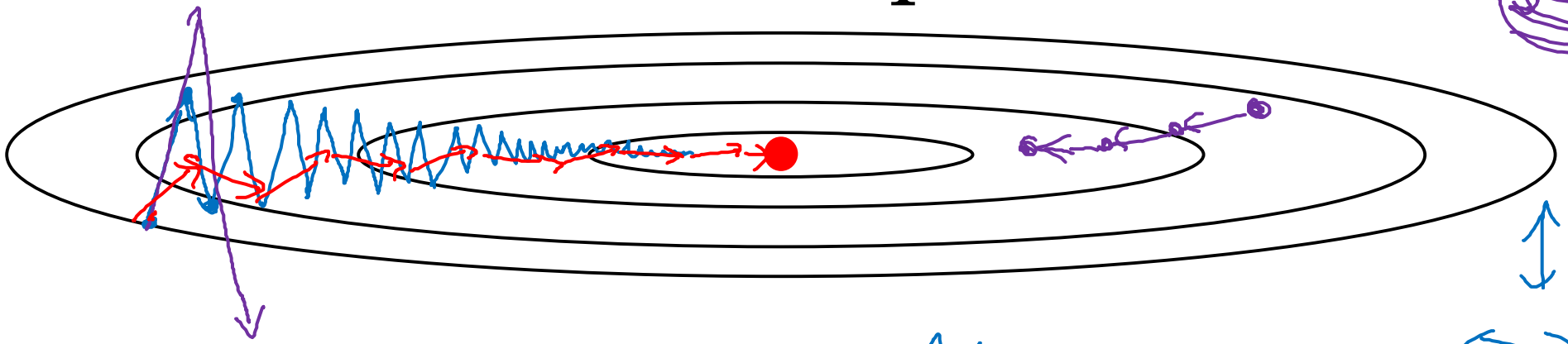
deeplearning.ai

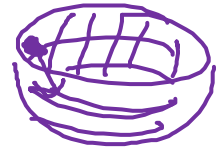
# Optimization Algorithms

---

# 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

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

To make the learning in the horizontal direction faster, to make the oscillation in the vertical direction smaller.

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

# Implementation details

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

On iteration $t$:

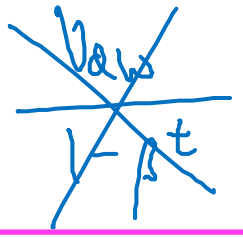
Compute $dW, db$ on the current mini-batch

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

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

$V_{dW} = \beta V_{dW} + dW$

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




Hyperparameters: $\alpha, \beta$

$\beta = 0.9$

average over last $\approx 10$ gradients