6. Gradient Descent with Momentum
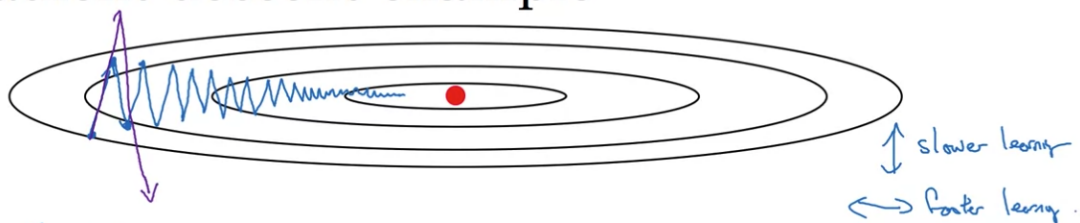
≣ Tags	Momentum

Gradient descent example

Gradient descent example

Gradient descent example

Momentum:

On iteration t:

Compute dW, db on current mini-batch.
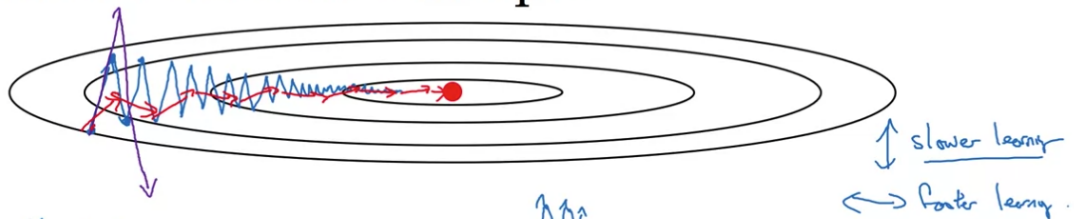
$$V_{dW} = \beta V_{dW} + (1-\beta)dW$$

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

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

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

↑ slower learning

↔ faster learning.

Andrew Ng

Gradient descent example

↑ slower learning
↔ faster learning

Momentum:
On iteration t:
Compute dW, db on current mini-batch.
$$v_{dW} = \beta v_{dW} + (1-\beta)dW$$
$$v_{db} = \beta v_{db} + (1-\beta)db$$

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

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

Andrew Ng

Implementation details

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

On iteration t:

Compute dW, db on the current mini-batch

$$v_{dW} = \beta v_{dW} + \cancel{(1-\beta)}dW$$
$$v_{db} = \beta v_{db} + \cancel{(1-\beta)}db$$
$$W = W - \alpha v_{dW}, \quad b = b - \alpha v_{db}$$

Hyperparameters: α, β $\beta = 0.9$

average over last ≈ 10 gradients

Andrew Ng