

Mini-Batch Gradient Descent

$$X, Y \xrightarrow{m \text{ 分割}} X^{(1)} \dots X^{(m)}, Y^{(1)} \dots Y^{(m)}$$

$$\downarrow$$

$$w^{(1)}, b^{(1)}$$

$$J^{(1)}$$

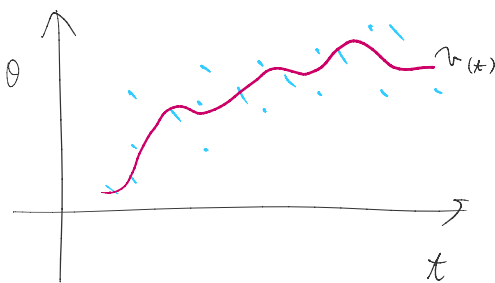
$$dw^{(1)}, db^{(1)}$$

Stochastic Gradient Descent.

Mini-Batch を "m 分割" (たとえば, 4000 の 100 個のバッチ) のみ計算する.

絶対に収束しないか? 速く

EMA

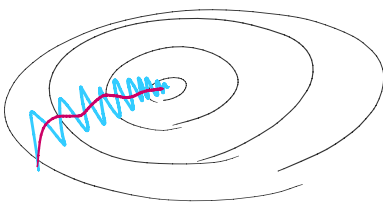


$$r_x = \beta r_{x-1} + (1-\beta) \theta_x$$

$\beta = 0.9 \rightarrow \approx \text{Avg of } 10x.$

$\beta = 0.98 \rightarrow \approx \text{Avg of } 50x$

Gradient Descent



Momentum

\rightarrow EMA を使って学習速度を高める.

RMSprop

$$S_{dw} := \beta S_{dw} + (1-\beta) d_w^2 \quad S_{db} := \beta S_{db} + (1-\beta) d_b^2$$

$$w := w - \alpha \frac{dw}{\sqrt{S_{dw}}} \quad b := b - \alpha \frac{db}{\sqrt{S_{db}}}$$

Adam \rightarrow Momentum + RMSprop