

Gradient Descent:

$$W := W - \alpha dw$$

$$b := b - \alpha db$$

Momentum: 通常情况下: α (手动调整)

初始化 $V_{dw} = 0, V_{db} = 0$ $\beta_1 = 0.9$ (默认)

$$V_{dw} = \beta_1 V_{dw} + (1 - \beta_1) dw$$

$$V_{db} = \beta_1 V_{db} + (1 - \beta_1) db$$

$$W := W - \alpha V_{dw}$$

$$b := b - \alpha V_{db}$$

如果使用了 mini-batch, 则每个 batch
更新 V_{dw}, V_{db}, W, b

RMSprop:

初始化 $S_{dw} = 0$, $S_{db} = 0$

$$S_{dw} = \beta_2 S_{dw} + (1 - \beta_2) dw^2$$

$$S_{db} = \beta_2 S_{db} + (1 - \beta_2) db^2$$

$$w := w - \alpha \frac{dw}{\sqrt{S_{dw} + \epsilon}}$$

$$b := b - \alpha \frac{db}{\sqrt{S_{db} + \epsilon}}$$

通常:

α (手动调整)

$$\beta_2 = 0.999$$

$$\epsilon = 10^{-8}$$

默认

如果使用了 mini-batch, 则每个 batch 更新 S_{dw} , S_{db} , w , b

Adam:

按 Momentum 和 RMSprop 方法计算得到

$V_{dw}, V_{db}, S_{dw}, S_{db}$

对 $V_{dw}, V_{db}, S_{dw}, S_{db}$ 进行修正

$$V_{dw}^{corrected} = \frac{V_{dw}}{1 - \beta_1^t}, \quad V_{db}^{corrected} = \frac{V_{db}}{1 - \beta_1^t}$$

$$S_{dw}^{corrected} = \frac{S_{dw}}{1 - \beta_2^t}, \quad S_{db}^{corrected} = \frac{S_{db}}{1 - \beta_2^t}$$

$$W := W - 2 \frac{V_{dw}^{corrected}}{\sqrt{S_{dw}^{corrected}} + \epsilon}$$

$$b := b - 2 \frac{V_{db}^{corrected}}{\sqrt{S_{db}^{corrected}} + \epsilon}$$

通常: λ (手动调整)

$$\beta_1 = 0.9$$

$$\beta_2 = 0.999 \quad (\text{默认})$$

$$\varepsilon = 10^{-8}$$

如果使用了 mini-batch, 则每个 batch
更新 $V_{dw}^{\text{corrected}}$, $V_{db}^{\text{corrected}}$, t , $S_{dw}^{\text{corrected}}$,
 $S_{db}^{\text{corrected}}$, w , b