

# Adam Optimizer

Adam = RMSProp + Momentum (Adaptive Moment Estimation)

- Start by initializing  $V_{dw}$ ,  $S_{dw}$  and  $V_{db}$ ,  $S_{db}$  to all zero
- Then do the Momentum calculations for  $V_{dw}$ ,  $V_{db}$
- Then do the RMS calculations for  $S_{dw}$ ,  $S_{db}$
- Then do bias corrections for all
- Then just combine Momentum update with RMS update 😊

$$V_{dw} = 0, S_{dw} = 0, V_{db} = 0, S_{db} = 0$$

On iteration  $t$ :

Compute  $dw, db$  using current mini-batch

$$V_{dw} = \beta_1 V_{dw} + (1 - \beta_1) dw, V_{db} = \beta_1 V_{db} + (1 - \beta_1) db \leftarrow \text{"momentum"} \beta_1$$

$$S_{dw} = \beta_2 S_{dw} + (1 - \beta_2) dw^2, S_{db} = \beta_2 S_{db} + (1 - \beta_2) db^2 \leftarrow \text{"RMSprop"} \beta_2$$

$$V_{dw}^{\text{corrected}} = V_{dw} / (1 - \beta_1^t), V_{db}^{\text{corrected}} = V_{db} / (1 - \beta_1^t)$$

$$S_{dw}^{\text{corrected}} = S_{dw} / (1 - \beta_2^t), S_{db}^{\text{corrected}} = S_{db} / (1 - \beta_2^t)$$

$$W := W - \alpha \frac{V_{dw}^{\text{corrected}}}{\sqrt{S_{dw}^{\text{corrected}} + \epsilon}}, b := b - \alpha \frac{V_{db}^{\text{corrected}}}{\sqrt{S_{db}^{\text{corrected}} + \epsilon}}$$

# Hyperparameters choice:

- $\alpha$  : needs to be tune
- $\beta_1$  : 0.9 (dw)
- $\beta_2$  : 0.999 ( $dw^2$ )
- $\epsilon$  :  $10^{-8}$

Only need to tune alpha usually.  $\beta_1$  and  $\beta_2$  were the default suggestions by the authors of the Adam optimizer paper