Optimization
Algorithms

Adam optimization
algorithm

deeplearning.ai

Adam optimization algorithm

$V_{dw} = 0, \quad S_{dw} = 0. \quad V_{db} = 0, \quad S_{db} = 0$

On iteration t:

Compute dW, db using current mini-batch

$V_{dw} = \beta_1 V_{dw} + (1-\beta_1) dW \quad, \quad V_{db} = \beta_1 V_{db} + (1-\beta_1) db \quad \Leftarrow \text{"momentum"} \; \beta_1$

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

yhat = np.array([.9, 0.2, 0.1, .4, .9])

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

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

$W := W - \alpha \dfrac{V_{dw}^{corrected}}{\sqrt{S_{dw}^{corrected}} + \varepsilon} \qquad b := b - \alpha \dfrac{V_{db}^{corrected}}{\sqrt{S_{db}^{corrected}} + \varepsilon}$

Hyperparameters choice:

$\rightarrow \alpha$: needs to be tune

$\rightarrow \beta_1$: 0.9 $\longrightarrow$ ($\underline{dw}$)

$\rightarrow \beta_2$: 0.999 $\longrightarrow$ ($\underline{dw^2}$)

$\rightarrow \varepsilon$: 10^{-8}

Adam : Adaptive moment estimation

Adam Coates

Andrew Ng