



deeplearning.ai

# Optimization Algorithms

---

## Adam optimization algorithm

RMSprop and the Adam optimization algorithm, is one of those rare algorithms that has really stood up, and has been shown to work well across a wide range of deep learning architectures. And the Adam optimization algorithm is basically taking momentum and RMSprop and putting them together

# Adam optimization algorithm

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

On iteration  $t$ :

Compute  $dw, db$  using current mini-batch

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

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

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

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

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

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

# Hyperparameters choice:

→  $\alpha$  : needs to be tune

→  $\beta_1$  : 0.9 → ( $dw$ )

moving average, weighted average. momentum light term

→  $\beta_2$  : 0.999 → ( $dw^2$ )

→  $\epsilon$  :  $10^{-8}$

Adam : Adaptive moment estimation

So beta1 is computing the mean of the derivatives, this called the first moment.

And beta2 is used to compute exponentially weighted average of the squares, and that's called the second moment.

But everyone just calls it the Adam authorization algorithm.



Adam Coates