deeplearning.ai

Optimization Algorithms

Adam optimization algorithm

Adam optimization algorithm (momentum + RMSprop)

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

On iteration t:

Compute dW, db using current mini-batch

$$\begin{cases} V_{dW} = \beta_1 V_{dW} + (1-\beta_1) dW, & V_{db} = \beta_1 V_{db} + (1-\beta_1) db \leftarrow \text{"momentum" } \beta_1 \\ S_{dW} = \beta_2 S_{dW} + (1-\beta_2) dW^2, & S_{db} = \beta_2 S_{db} + (1-\beta_2) db^2 \leftarrow \text{"RMSprop" } \beta_2 \end{cases}$$

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

$$\begin{cases} V_{dW}^{corrected} = V_{dW} / (1-\beta_1^t), & V_{db}^{corrected} = V_{db} / (1-\beta_1^t) \\ S_{dW}^{corrected} = S_{dW} / (1-\beta_2^t), & S_{db}^{corrected} = S_{db} / (1-\beta_2^t) \end{cases}$$

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

Andrew Ng

Hyperparameters choice:

→ α : needs to be tuned

→ β_1 : 0.9 → (dw) 第一矩

→ β_2 : 0.999 → (dw^2) 第二矩

→ ε : 10^{-8}

Adam: Adaptive moment estimation

自适应 矩 估计

Adam Coates

Andrew Ng