deeplearning.ai

Optimization Algorithms

Adam optimization algorithm

Adam optimization algorithm

– Momentum + RMS Prop (has param β_1) (has param β_2)

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

on iteration t:

compute dw, db using current minibatch

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

→ momentum part with β_1

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

→ RMS prop part β_2

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

// Here we use bias correction for momentum/RMS prop

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

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

// Update

$$w := w - \alpha \frac{V_{dw}^{corrected}}{\sqrt{S_{dw}^{corrected}} + \varepsilon}, \quad b = b - \alpha \cdot \frac{V_{db}^{corrected}}{\sqrt{S_{db}^{corrected}} + \varepsilon}$$

ε → prevent $\sqrt{0}$

Andrew Ng

Hyperparameters choice:

α = needs to be tuned

$\beta_1 = 0.9$ (dw)

$\beta_2 = 0.999$ (dw^2)

$\varepsilon = 10^{-8}$

dw = first moment

dw^2 = second moment

Adam : Adaptive Moment Estimation

Adam Coates

Andrew Ng