Optimization Algorithms

Adam optimization algorithm

deeplearning.ai

Adam은 모멘텀이랑 RMSProp이랑 합친거다.

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

$V_{dw} = \beta_1 V_{dw} + (1-\beta_1) dW$, $V_{db} = \beta_1 V_{db} + (1-\beta_1) db$ ← "momentum" β_1

$S_{dw} = \beta_2 S_{dw} + (1-\beta_2) dW^2$, $S_{db} = \beta_2 S_{db} + (1-\beta_2) db^2$ ← "RMSprop" β_2

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

bias corection

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

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

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

$b := b - \alpha \dfrac{V_{db}^{corrected}}{\sqrt{S_{db}^{corrected}} + \varepsilon}$ → 모멘텀 부분

→ RMSProp 부분

Andrew Ng

Hyperparameters choice:

$$\rightarrow \alpha : \text{needs to be tune}$$

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

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

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

Adam: Adaptive moment estimation

Adam Coates

Andrew Ng