AI VIET NAM – COURSE 2023

Optimization Algorithms in Deep Learning - Solution

Ngày 21 tháng 10 năm 2023

$$f(w_1, w_2) = 0.1w_1^2 + 2w_2^2$$

1. GD: initial point $w1 = -5$, $w2 = -2$, $\alpha = 0.4$, epoch=2
2. GD + momentum: initial point $w1 = -5$, $w2 = -2$, initial velocity $v_1 = 0$, $v_2 = 0$, $\alpha = 0.6$, $\beta = 0.5$, epoch=2
3. RMSProp: initial point $w1 = -5$, $w2 = -2$, initial $s_1 = 0$, $s_2 = 0$, $\alpha = 0.3$, $\gamma = 0.9$, $\epsilon = 10^{-6}$ epoch=2
4. Adam: initial point $w1 = -5$, $w2 = -2$, initial $v_1 = 0$, $v_2 = 0$, $s_1 = 0$, $s_2 = 0$, $\alpha = 0.2$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-6}$ epoch=2

1. **GD:**

$$\boxed{\text{W} = \text{W - } \alpha * dW}$$

- epoch=1:
  - $dw_1 = 0.2w_1 = 0.2 * (-5) = -1$
  - $dw_2 = 4w_2 = 4 * (-2) = -8$
  - $w_1 = w_1 - \alpha * dw_1 = -5 - 0.4 * (-1) = -4.6$
  - $w_2 = w_2 - \alpha * dw_2 = -2 - 0.4 * (-8) = 1.2$
- epoch=2:
  - $dw_1 = 0.2w_1 = 0.2 * (-4.6) = -0.92$
  - $dw_2 = 4w_2 = 4 * 1.2 = 4.8$
  - $w_1 = w_1 - \alpha * dw_1 = -4.6 - 0.4 * (-0.92) \approx -4.232$
  - $w_2 = w_2 - \alpha * dw_2 = 1.2 - 0.4 * 4.8 = -0.72$

2. **GD + momentum**

$$\boxed{\begin{aligned}&\text{V}_t = \beta V_{t-1} + (1-\beta)dW_t \\ &\text{W}_t = W_t - \alpha * V_t\end{aligned}}$$

- epoch=1:
  - $dw_1 = 0.2w_1 = 0.2 * (-5) = -1$
  - $dw_2 = 4w_2 = 4 * (-2) = -8$
  - $v_1 = \beta * v_1 + (1-\beta) * dw_1 = 0.5 * 0 + (1-0.5) * (-1) = -0.5$
  - $v_2 = \beta * v_2 + (1-\beta) * dw_2 = 0.5 * 0 + (1-0.5) * (-8) = -4$
  - $w_1 = w_1 - \alpha * dw_1 = -5 - 0.6 * (-0.5) = -4.7$
  - $w_2 = w_2 - \alpha * dw_2 = -2 - 0.6 * (-4) = 0.4$

- epoch=2:
  - $dw_1 = 0.2w_1 = 0.2 * (-4.7) = -0.94$
  - $dw_2 = 4w_2 = 4 * 0.4 = 1.6$
  - $v_1 = \beta * v_1 + (1 - \beta) * dw_1 = 0.5 * (-0.5) + (1 - 0.5) * (-0.94) = -0.72$
  - $v_2 = \beta * v_2 + (1 - \beta) * dw_2 = 0.5 * (-4) + (1 - 0.5) * 1.6 = -1.2$
  - $w_1 = w_1 - \alpha * dw_1 = -4.7 - 0.6 * (-0.72) = -4.268$
  - $w_2 = w_2 - \alpha * dw_2 = 0.4 - 0.6 * (-1.2) = 1.12$

3. **RMSprop**

$$\boxed{\begin{aligned} &\mathrm{S}_t = \gamma S_{t-1} + (1-\gamma) dW_t^2 \\ &\mathrm{W}_t = W_t - \alpha * \frac{dW}{\sqrt{S_t + \epsilon}} \end{aligned}}$$

- epoch=1:
  - $dw_1 = 0.2w_1 = 0.2 * (-5) = -1$
  - $dw_2 = 4w_2 = 4 * (-2) = -8$
  - $s_1 = \gamma * s_1 + (1 - \gamma) * dw_1^2 = 0.9 * 0 + (1 - 0.9) * (-1)^2 = 0.1$
  - $s_2 = \gamma * s_2 + (1 - \gamma) * dw_2^2 = 0.9 * 0 + (1 - 0.9) * (-8)^2 = 6.4$
  - $w_1 = w_1 - \alpha * \frac{dw_1}{\sqrt{s_1 + \epsilon}} = -5 - 0.3 * \frac{-1}{\sqrt{0.1 + 10^{-6}}} \approx -4.051$
  - $w_2 = w_2 - \alpha * \frac{dw_2}{\sqrt{s_2 + \epsilon}} = -2 - 0.3 * \frac{-8}{\sqrt{6.4 + 10^{-6}}} \approx -1.051$
- epoch=2:
  - $dw_1 = 0.2w_1 = 0.2 * (-4.051) = -0.8102$
  - $dw_2 = 4w_2 = 4 * (-1.051) = -4.204$
  - $s_1 = \gamma * s_1 + (1 - \gamma) * dw_1^2 = 0.9 * 0.1 + (1 - 0.9) * (-0.8102)^2 \approx 0.156$
  - $s_2 = \gamma * s_2 + (1 - \gamma) * dw_2^2 = 0.9 * 6.4 + (1 - 0.9) * (-4.204)^2 \approx 7.527$
  - $w_1 = w_1 - \alpha * \frac{dw_1}{\sqrt{s_1 + \epsilon}} = -4.051 - 0.3 * \frac{-0.8102}{\sqrt{0.156 + 10^{-6}}} \approx -3.436$
  - $w_2 = w_2 - \alpha * \frac{dw_2}{\sqrt{s_2 + \epsilon}} = -1.051 - 0.3 * \frac{-4.204}{\sqrt{7.527 + 10^{-6}}} \approx -0.591$

4. **Adam**

$$\boxed{\begin{aligned} &\mathrm{V}_t = \beta_1 V_{t-1} + (1-\beta_1) dW_t \\ &\mathrm{S}_t = \beta_2 S_{t-1} + (1-\beta_2) dW_t^2 \\ &\mathrm{V}_{corr} = \frac{V_t}{1-\beta_1^t} \\ &\mathrm{S}_{corr} = \frac{S_t}{1-\beta_2^t} \\ &\mathrm{W}_t = W_t - \alpha * \frac{V_{corr}}{\sqrt{S_{corr}} + \epsilon} \end{aligned}}$$

- epoch=1:
  - $dw_1 = 0.2w_1 = 0.2 * (-5) = -1$
  - $dw_2 = 4w_2 = 4 * (-2) = -8$
  - $v_1 = \beta_1 * v_1 + (1 - \beta_1) * dw_1 = 0.9 * 0 + (1 - 0.9) * (-1) = -0.1$
  - $v_2 = \beta_1 * v_2 + (1 - \beta_1) * dw_2 = 0.9 * 0 + (1 - 0.9) * (-8) = -0.8$
  - $s_1 = \beta_2 * s_1 + (1 - \beta_2) * dw_1^2 = 0.999 * 0 + (1 - 0.999) * (-1)^2 = 0.001$
  - $s_2 = \beta_2 * s_2 + (1 - \beta_2) * dw_2^2 = 0.999 * 0 + (1 - 0.999) * (-8)^2 = 0.064$
  - $v_{corr1} = \dfrac{v_1}{1 - \beta_1^t} = \dfrac{-0.1}{1 - 0.9^1} = -1$
  - $v_{corr2} = \dfrac{v_2}{1 - \beta_1^t} = \dfrac{-0.8}{1 - 0.9^1} = -8$
  - $s_{corr1} = \dfrac{s_1}{1 - \beta_2^t} = \dfrac{0.001}{1 - 0.999^1} \approx 1$
  - $s_{corr2} = \dfrac{s_2}{1 - \beta_2^t} = \dfrac{0.064}{1 - 0.999^1} \approx 64$
  - $w_1 = w_1 - \alpha * \dfrac{v_{corr1}}{\sqrt{s_{corr1} + \epsilon}} = -5 - 0.2 * \dfrac{-1}{\sqrt{1} + 10^{-6}} \approx -4.8$
  - $w_2 = w_2 - \alpha * \dfrac{v_{corr2}}{\sqrt{s_{corr2} + \epsilon}} = -2 - 0.2 * \dfrac{-8}{\sqrt{64} + 10^{-6}} \approx -1.8$
- epoch=2:
  - $dw_1 = 0.2w_1 = 0.2 * (-4.8) = -0.96$
  - $dw_2 = 4w_2 = 4 * (-1.8) = -7.2$
  - $v_1 = \beta_1 * v_1 + (1 - \beta_1) * dw_1 = 0.9 * (-0.1) + (1 - 0.9) * (-0.96) = -0.186$
  - $v_2 = \beta_1 * v_2 + (1 - \beta_1) * dw_2 = 0.9 * (-0.8) + (1 - 0.9) * (-7.2) = -1.44$
  - $s_1 = \beta_2 * s_1 + (1 - \beta_2) * dw_1^2 = 0.999 * 0.001 + (1 - 0.999) * (-0.96)^2 \approx 0.0019206$
  - $s_2 = \beta_2 * s_2 + (1 - \beta_2) * dw_2^2 = 0.999 * 0.064 + (1 - 0.999) * (-7.2)^2 \approx 0.115776$
  - $v_{corr1} = \dfrac{v_1}{1 - \beta_1^t} = \dfrac{-0.186}{1 - 0.9^2} \approx -0.9789474$
  - $v_{corr2} = \dfrac{v_2}{1 - \beta_1^t} = \dfrac{-1.44}{1 - 0.9^2} \approx -7.5789474$
  - $s_{corr1} = \dfrac{s_1}{1 - \beta_2^t} = \dfrac{0.0019206}{1 - 0.999^2} \approx 0.9607804$
  - $s_{corr2} = \dfrac{s_2}{1 - \beta_2^t} = \dfrac{0.115776}{1 - 0.999^2} \approx 57.9169585$
  - $w_1 = w_1 - \alpha * \dfrac{v_{corr1}}{\sqrt{s_{corr1} + \epsilon}} = -4.8 - 0.2 * \dfrac{-0.9789474}{\sqrt{0.9607804} + 10^{-6}} \approx -4.6002546$
  - $w_2 = w_2 - \alpha * \dfrac{v_{corr2}}{\sqrt{s_{corr2} + \epsilon}} = -1.8 - 0.2 * \dfrac{-7.5789474}{\sqrt{57.9169585} + 10^{-6}} \approx -1.6008245$