

Optimization Algorithms in Deep Learning

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

Bài 1 Gradient Descent

$$W = W - \alpha * dW$$

$$w_1 = -5, w_2 = -2, \alpha = 0.4$$

$$dw_1 = 0.2w_1$$

$$dw_2 = 4w_2$$

Epoch 1:

$$dw_1 = 0.2 * (-5) = -1$$

$$dw_2 = 4 * (-2) = -8$$

$$w_1 = -5 - 0.4 * (-1) = -4.6$$

$$w_2 = -2 - 0.4 * (-8) = 1.2$$

Epoch 2:

$$dw_1 = 0.2 * (-4.6) = -0.92$$

$$dw_2 = 4 * (1.2) = 4.8$$

$$w_1 = -4.6 - 0.4 * (-0.92) = -4.232$$

$$w_2 = 1.2 - 0.4 * (4.8) = -0.72$$

Bài 2 Gradient Descent + Momentum

$$V_t = \beta V_{t-1} + (1 - \beta) dW_t$$

$$W_t = W_t - \alpha * V_t$$

$$w_1 = -5, w_2 = -2, v_1 = 0, v_2 = 0, \alpha = 0.6, \beta = 0.5$$

$$dw_1 = 0.2w_1$$

$$dw_2 = 4w_2$$

Epoch 1:

$$dw_1 = 0.2 * (-5) = -1$$

$$dw_2 = 4 * (-2) = -8$$

$$v_1 = 0.5 * 0 + 0.5 * (-1) = -0.5$$

$$v_2 = 0.5 * 0 + 0.5 * (-8) = -4$$

$$w_1 = -5 - 0.6 * (-0.5) = -4.7$$

$$w_2 = -2 - 0.6 * (-4) = 0.4$$

Epoch 2:

$$dw_1 = 0.2 * (-4.7) = -0.94$$

$$dw_2 = 4 * (0.4) = 1.6$$

$$v_1 = 0.5 * (-0.5) + 0.5 * (-0.94) = -0.72$$

$$v_2 = 0.5 * (-4) + 0.5 * (1.6) = -1.2$$

$$w_1 = -4.7 - 0.6 * (-0.72) = -4.268$$

$$w_2 = 0.4 - 0.6 * (-1.2) = 1.12$$

Bài 3 RMSProp

$$S_t = \gamma S_{t-1} + (1 - \gamma) dW_t^2$$

$$W_t = W_t - \alpha * \frac{dW}{\sqrt{S_t + \epsilon}}$$

$$w1 = -5, w2 = -2, s1 = 0, s2 = 0, \alpha = 0.3, \gamma = 0.9, \epsilon = 10^{-6}$$

Epoch 1:

$$dw_1 = 0.2 * (-5) = -1$$

$$dw_2 = 4 * (-2) = -8$$

$$s_1 = 0.9 * 0 + 0.1 * (-1)^2 = 0.1$$

$$s_2 = 0.9 * 0 + 0.1 * (-8)^2 = 6.4$$

$$w_1 = -5 - 0.3 * \frac{(-1)}{\sqrt{0.1 + 10^{-6}}} = -4.051$$

$$w_2 = -2 - 0.3 * \frac{(-8)}{\sqrt{6.4 + 10^{-6}}} = -1.051$$

Epoch 2:

$$dw_1 = 0.2 * (-4.051) = -0.8102$$

$$dw_2 = 4 * (-1.051) = -4.204$$

$$s_1 = 0.9 * 0.1 + 0.1 * (-0.8102)^2 = 0.156$$

$$s_2 = 0.9 * 6.4 + 0.1 * (-4.204)^2 = 7.527$$

$$w_1 = -4.051 - 0.3 * \frac{(-0.8102)}{\sqrt{0.156 + 10^{-6}}} = -3.436$$

$$w_2 = -1.051 - 0.3 * \frac{(-4.204)}{\sqrt{7.527 + 10^{-6}}} = -0.591$$

Bài 4 Adam

$$w1 = -5, w2 = -2, v1 = 0, v2 = 0, s1 = 0, s2 = 0, \alpha = 0.2, \beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-6}$$

$$V_t = \beta_1 V_{t-1} + (1 - \beta_1) dW_t$$

$$S_t = \beta_2 S_{t-1} + (1 - \beta_2) dW_t^2$$

$$V_{corr} = \frac{V_t}{1 - \beta_1^t}$$

$$S_{corr} = \frac{S_t}{1 - \beta_2^t}$$

$$W_t = W_t - \alpha * \frac{V_{corr}}{\sqrt{S_{corr} + \epsilon}}$$

Epoch 1:

$$dw_1 = 0.2 * (-5) = -1$$

$$dw_2 = 4 * (-2) = -8$$

$$v_1 = 0.9 * 0 + 0.1 * (-1) = -0.1$$

$$v_2 = 0.9 * 0 + 0.1 * (-8) = -0.8$$

$$s_1 = 0.999 * 0 + 0.001 * (-1)^2 = 0.001$$

$$s_2 = 0.999 * 0 + 0.001 * (-8)^2 = 0.064$$

$$v_{coord1} = \frac{v_1}{1 - \beta_1^1} = \frac{-0.1}{1 - 0.9^1} = -1$$

$$v_{coord2} = \frac{v_2}{1 - \beta_1^1} = \frac{-0.8}{1 - 0.9^1} = -8$$

$$s_{coord1} = \frac{s_1}{1 - \beta_2^1} = \frac{0.001}{1 - 0.999^1} = 1$$

$$s_{coord2} = \frac{s_2}{1 - \beta_2^1} = \frac{0.064}{1 - 0.999^1} = 64$$

$$w_1 = -5 - 0.2 * \frac{-1}{\sqrt{1 + 10^{-6}}} = -4.8$$

$$w_2 = -2 - 0.2 * \frac{-8}{\sqrt{64 + 10^{-6}}} = -1.8$$

Epoch 2:

$$dw_1 = 0.2 * (-4.8) = -0.96$$

$$dw_2 = 4 * (-1.8) = -7.2$$

$$v_1 = 0.9 * (-0.1) + 0.1 * (-0.96) = -0.186$$

$$v_2 = 0.9 * (-0.8) + 0.1 * (-7.2) = -1.44$$

$$s_1 = 0.999 * 0.001 + 0.001 * (-0.96)^2 = 0.0019206$$

$$s_2 = 0.999 * 0.064 + 0.001 * (-7.2)^2 = 0.115776$$

$$v_{coord1} = \frac{v_1}{1 - \beta_1^2} = \frac{-0.186}{1 - 0.9^2} = -0.9789$$

$$v_{coord2} = \frac{v_2}{1 - \beta_1^2} = \frac{-1.44}{1 - 0.9^2} = -7.5789$$

$$s_{coord1} = \frac{s_1}{1 - \beta_2^2} = \frac{0.0019206}{1 - 0.999^2} = 0.9608$$

$$s_{coord2} = \frac{s_2}{1 - \beta_2^2} = \frac{0.115776}{1 - 0.999^2} = 57.9170$$

$$w_1 = -4.8 - 0.2 * \frac{-0.9789}{\sqrt{0.9608} + 10^{-6}} = -4.6$$

$$w_2 = -1.8 - 0.2 * \frac{-7.5789}{\sqrt{57.9170} + 10^{-6}} = -0.6$$