

Cho phương trình dưới đây, sử dụng chúng để giải các thuật toán về Optimizer:

$$\boxed{f(w_1, w_2) = 0.1w_1^2 + 2w_2^2} \quad (1)$$

1. Gradient Descent

Dựa vào thuật toán **Gradient Descent**, các bạn thực hiện tìm điểm *minimum* của function (1) với các tham số $w_1 = -5$, $w_2 = -2$, $\alpha = 0.4$:

Lời giải:

Gradient của f :

$$\nabla f(w_1, w_2) = \begin{bmatrix} 0.2w_1 \\ 4w_2 \end{bmatrix}.$$

Công thức cập nhật: $W = W - \alpha \cdot \nabla f(W)$.

Khởi tạo: $w_1 = -5$, $w_2 = -2$.

Epoch 1: $\nabla f(w_1, w_2) = \begin{bmatrix} -1 \\ -8 \end{bmatrix}$, cập nhật:

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} -5 \\ -2 \end{bmatrix} - 0.4 \cdot \begin{bmatrix} -1 \\ -8 \end{bmatrix} = \begin{bmatrix} -4.6 \\ 1.2 \end{bmatrix}.$$

Epoch 2: $\nabla f(w_1, w_2) = \begin{bmatrix} -0.92 \\ 4.8 \end{bmatrix}$, cập nhật:

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} -4.6 \\ 1.2 \end{bmatrix} - 0.4 \cdot \begin{bmatrix} -0.92 \\ 4.8 \end{bmatrix} = \begin{bmatrix} -4.232 \\ -0.72 \end{bmatrix}.$$

Kết quả: Sau 2 epoch: $w_1 \approx -4.232$, $w_2 \approx -0.72$.

2. Gradient Descent with Momentum

Dựa vào thuật toán **Gradient Descent**, tìm điểm *minimum* của function (1) với $w_1 = -5$, $w_2 = -2$, $\alpha = 0.4$:

Gradient: $\nabla f(w_1, w_2) = \begin{bmatrix} 0.2w_1 \\ 4w_2 \end{bmatrix}$, $W = W - \alpha \cdot \nabla f(W)$.

Khởi tạo: $w_1 = -5$, $w_2 = -2$.

Epoch 1:

$$\nabla f = \begin{bmatrix} -1 \\ -8 \end{bmatrix}, \quad \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} -5 \\ -2 \end{bmatrix} - 0.4 \cdot \begin{bmatrix} -1 \\ -8 \end{bmatrix} = \begin{bmatrix} -4.6 \\ 1.2 \end{bmatrix}.$$

Epoch 2:

$$\nabla f = \begin{bmatrix} -0.92 \\ 4.8 \end{bmatrix}, \quad \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} -4.6 \\ 1.2 \end{bmatrix} - 0.4 \cdot \begin{bmatrix} -0.92 \\ 4.8 \end{bmatrix} = \begin{bmatrix} -4.232 \\ -0.72 \end{bmatrix}.$$

Kết quả: $w_1 \approx -4.232$, $w_2 \approx -0.72$.

3. RMSProp

Dựa vào thuật toán **RMSProp**, tìm điểm *minimum* của function (1) với $w_1 = -5$, $w_2 = -2$, $s_1 = 0$, $s_2 = 0$, $\alpha = 0.3$, $\gamma = 0.9$, $\epsilon = 10^{-6}$.

Công thức:

$$s_t = \gamma s_{t-1} + (1 - \gamma)g_t^2, \quad w_t = w_{t-1} - \frac{\alpha}{\sqrt{s_t + \epsilon}}g_t$$

Epoch 1: Gradient: $g_t = \begin{bmatrix} -1 \\ -8 \end{bmatrix}$, $s_t = \begin{bmatrix} 0.1 \\ 6.4 \end{bmatrix}$, $w_t = \begin{bmatrix} -4.052 \\ -1.7 \end{bmatrix}$.

Epoch 2: Gradient: $g_t = \begin{bmatrix} -0.8104 \\ -6.8 \end{bmatrix}$, $s_t = \begin{bmatrix} 0.108 \\ 11.648 \end{bmatrix}$, $w_t = \begin{bmatrix} -3.805 \\ -1.416 \end{bmatrix}$.

Kết quả: $w_1 \approx -3.805$, $w_2 \approx -1.416$.

4. Adam

Dựa vào thuật toán **Adam**, tìm điểm *minimum* của function (1) với $w_1 = -5$, $w_2 = -2$, $v_1 = 0$, $v_2 = 0$, $s_1 = 0$, $s_2 = 0$, $\alpha = 0.2$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-6}$.

Công thức:

$$v_t = \beta_1 v_{t-1} + (1 - \beta_1)g_t, \quad s_t = \beta_2 s_{t-1} + (1 - \beta_2)g_t^2, \quad \hat{v}_t = \frac{v_t}{1 - \beta_1^t}, \quad \hat{s}_t = \frac{s_t}{1 - \beta_2^t}, \quad w_t = w_{t-1} - \frac{\alpha}{\sqrt{\hat{s}_t + \epsilon}}\hat{v}_t$$

Epoch 1: Gradient: $g_t = \begin{bmatrix} -1 \\ -8 \end{bmatrix}$, $v_t = \begin{bmatrix} -0.1 \\ -0.8 \end{bmatrix}$, $s_t = \begin{bmatrix} 0.001 \\ 0.064 \end{bmatrix}$, $\hat{v}_t = \begin{bmatrix} -1 \\ -8 \end{bmatrix}$, $\hat{s}_t = \begin{bmatrix} 1 \\ 64 \end{bmatrix}$,
 $w_t = \begin{bmatrix} -4.8 \\ -1.8 \end{bmatrix}$.

Epoch 2: Gradient: $g_t = \begin{bmatrix} -0.96 \\ -7.2 \end{bmatrix}$, $v_t = \begin{bmatrix} -0.186 \\ -1.44 \end{bmatrix}$, $s_t = \begin{bmatrix} 0.00185 \\ 0.103 \end{bmatrix}$, $\hat{v}_t = \begin{bmatrix} -0.979 \\ -7.59 \end{bmatrix}$,
 $\hat{s}_t = \begin{bmatrix} 0.926 \\ 51.65 \end{bmatrix}$, $w_t = \begin{bmatrix} -4.602 \\ -1.662 \end{bmatrix}$.

Kết quả: $w_1 \approx -4.602$, $w_2 \approx -1.662$.