Excercise - Week 4 - Module 5

Tuan Nguyen Huu

November 2024

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

1 Excercise 1

- epoch 1:

$dw_1 = 0.2w_1 = 0.2 \times -5 = -1$

$dw_2 = 4w_2 = 4 \times -2 = -8$

$w_1 = w_1 - \alpha * dw_1 = -5 - 0.4 \times -1 = -4.6$

$w_2 = w_2 - \alpha * dw_2 = -2 - 0.4 \times -8 = 1.2$

- epoch 2:

$dw_1 = 0.2w_1 = 0.2 \times -4.6 = -0.92$

$dw_2 = 4w_2 = 4 \times 1.2 = 4.8$

$w_1 = w_1 - \alpha * dw_1 = -4.6 - 0.4 \times -0.92 = -4.232$

$w_2 = w_2 - \alpha * dw_2 = 1.2 - 0.4 \times 4.8 = -0.72$

2 Excercise 2

- epoch 1:

$dw_1 = 0.2w_1 = 0.2 \times -5 = -1$

$dw_2 = 4w_2 = 4 \times -2 = -8$

$v_1 = \beta v_1 + (1 - \beta)dw_1 = 0.5 \times 0 + 0.5 \times -1 = -0.5$

$v_2 = \beta v_2 + (1 - \beta)dw_2 = 0.5 \times 0 + 0.5 \times -8 = -4$

$w_1 = w_1 - \alpha * v_1 = -5 - 0.6 * -0.5 = -4.7$

$w_2 = w_2 - \alpha * v_2 = -2 - 0.6 * -4 = 0.4$

- epoch 2:

$dw_1 = 0.2w_1 = 0.2 \times -4.7 = -0.94$

$dw_2 = 4w_2 = 4 \times 0.4 = 1.6$

$v_1 = \beta v_1 + (1 - \beta)dw_1 = 0.5 \times -0.5 + 0.5 \times -0.94 = -0.72$

$v_2 = \beta v_2 + (1 - \beta)dw_2 = 0.5 \times -4 + 0.5 \times 1.6 = -1.2$

$w_1 = w_1 - \alpha * v_1 = -4.7 - 0.6 * -0.72 = -4.268$

$w_2 = w_2 - \alpha * v_2 = 0.4 - 0.6 * -1.2 = 1.12$

3 Excercise 3

- epoch 1:

$dw_1 = 0.2w_1 = 0.2 \times -5 = -1$

$dw_2 = 4w_2 = 4 \times -2 = -8$

$s_1 = \gamma s_1 + (1-\gamma)dw_1^2 = 0.9 \times 0 + 0.1 \times (-1)^2 = 0.1$

$s_2 = \gamma s_2 + (1-\gamma)dw_2^2 = 0.9 \times 0 + 0.1 \times (-8)^2 = 6.4$

$w1 = w1 - \alpha\frac{dw_1}{\sqrt{s_1+\epsilon}} = -5 - 0.3 \times \frac{-1}{\sqrt{0.1+10^{-6}}} = -4.051$

$w2 = w2 - \alpha\frac{dw_2}{\sqrt{s_2+\epsilon}} = -2 - 0.3 \times \frac{-8}{\sqrt{6.4+10^{-6}}} = -1.051$

- epoch 2:

$dw_1 = 0.2w_1 = 0.2 \times -4.051 = -0.8102$

$dw_2 = 4w_2 = 4 \times -1.051 = -4.204$

$s_1 = \gamma s_1 + (1-\gamma)dw_1^2 = 0.9 \times 0.1 + 0.1 \times (-0.8102)^2 = 0.156$

$s_2 = \gamma s_2 + (1-\gamma)dw_2^2 = 0.9 \times 6.4 + 0.1 \times (-4.204)^2 = 7.527$

$w1 = w1 - \alpha\frac{dw_1}{\sqrt{s_1+\epsilon}} = -4.051 - 0.3 \times \frac{-0.8102}{\sqrt{0.156+10^{-6}}} = -4.051$

$w2 = w2 - \alpha\frac{dw_2}{\sqrt{s_2+\epsilon}} = -1.051 - 0.3 \times \frac{-4.204}{\sqrt{7.527+10^{-6}}} = -0.591$

4 Excercise 4

- epoch 1:

$dw_1 = 0.2w_1 = 0.2 \times -5 = -1$

$dw_2 = 4w_2 = 4 \times -2 = -8$

$v_1 = \beta_1 V_1 + (1-\beta_1)dw_1 = 0.9 \times 0 + 0.1 \times -1 = -0.1$

$v_2 = \beta_1 V_2 + (1-\beta_1)dw_2 = 0.9 \times 0 + 0.1 \times -8 = -0.8$

$s_1 = \beta_2 s_1 + (1-\beta_1)dw_1^2 = 0.999 \times 0 + 0.001 \times (-1)^2 = 0.001$

$s_2 = \beta_2 s_2 + (1-\beta_1)dw_2^2 = 0.999 \times 0 + 0.001 \times (-8)^2 = 0.064$

$v_{corr1} = \frac{v_1}{1-\beta_1^t} = \frac{-0.1}{1-(0.9)^1} = -1$

$v_{corr2} = \frac{v_2}{1-\beta_2^t} = \frac{-0.9}{1-(0.9)^1} = -8$

$s_{corr1} = \frac{s_1}{1-\beta_2^t} = \frac{0.001}{1-(0.999)^1} = 1$

$s_{corr2} = \frac{s_2}{1-\beta_2^t} = \frac{0.064}{1-(0.999)^1} = 64$

$w_1 = w_1 - \alpha * \frac{v_{corr1}}{\sqrt{s_{corr2}+\epsilon}} = -5 - 0.2 \times \frac{-1}{\sqrt{1+10^{-6}}} = -4.8$

$w_2 = w_2 - \alpha * \frac{v_{corr2}}{\sqrt{s_{corr2}+\epsilon}} = -2 - 0.2 \times \frac{-8}{\sqrt{64+10^{-6}}} = -1.8$

- epoch 2:

$dw_1 = 0.2w_1 = 0.2 \times -4.8 = -0.96$

$dw_2 = 4w_2 = 4 \times -1.8 = -7.2$

$v_1 = \beta_1 V_1 + (1-\beta_1)dw_1 = 0.9 \times -0.1 + 0.1 \times -0.96 = -0.186$

$v_2 = \beta_1 V_2 + (1-\beta_1)dw_2 = 0.9 \times 0.8 + 0.1 \times -7.2 = -1.44$

$s_1 = \beta_2 s_1 + (1-\beta_1)dw_1^2 = 0.999 \times 0.001 + 0.001 \times (-0.96)^2 = 0.0019206$

$s_2 = \beta_2 s_2 + (1-\beta_1)dw_2^2 = 0.999 \times 0.064 + 0.001 \times (-7.2)^2 = 0.115776$

$v_{corr1} = \frac{v_1}{1-\beta_1^t} = \frac{-0.186}{1-(0.9)^2} = -0.9789474$

$$v_{corr2} = \frac{v_2}{1-\beta_2^t} = \frac{-1.44}{1-(0.9)^2} = -7.5789474$$

$$s_{corr1} = \frac{s_1}{1-\beta_2^t} = \frac{0.0019206}{1-(0.999)^2} = 0.9607804$$

$$s_{corr2} = \frac{s_2}{1-\beta_2^t} = \frac{0.115776}{1-(0.999)^1} = 57.9169585$$

$$w_1 = w_1 - \alpha * \frac{v_{corr1}}{\sqrt{s_{corr2}+\epsilon}} = -4.8 - 0.2 \times \frac{-0.9789474}{\sqrt{0.9607804+10^{-6}}} = -4.6002546$$

$$w_2 = w_2 - \alpha * \frac{v_{corr2}}{\sqrt{s_{corr2}+\epsilon}} = -1.8 - 0.2 \times \frac{-7.5789474}{\sqrt{57.9169585+10^{-6}}} = -1.6008245$$