**Bài tập MODULE 5 - [Model Optimazation - 16/11/2024]**

**Function (1)**



**Bài 1 Gradient Descent:**



1. Trình bày chi tiết từng bước thực hiện tìm điểm minimum theo thuật toán Gradient Descent (tìm w1 và w2 sau 2 epoch) với epoch = 2.

Các tham số sau khởi tạo: w1 = −5, w2 = −2, α = 0.4:

* epoch = 1
* STEP1: Tìm giá trị dw1 và dw2 giá trị đạo hàm của hàm (1) theo w1 và w2 (partial derivative) tại điểm khởi tạo [w1, w2]

dw1 = 2\*0.1\*w1 = 2\*0.1\*-5 = -1

dw2 = 2\*2\*w2 = 2\*2\*-2 = -8

* STEP2: Dùng công thức Gradient Descent (1.1) để cập nhật w1 và w2. Hoàn thành epoch = 1

w1 = w1 – α\*dw1 = -5 – 0.4\*(-1) = -4.6

w2 = w2 – α\*dw2 = -2 – 0.4\*(-8) = 1.2

* STEP3: epoch = 2 ta thực hiện tương tự với STEP1 và STEP2 như trên với w1 và w2 đã được cập nhật từ epoch = 1

dw1 = 2\*0.1\*w1 = 2\*0.1\*-4.6 = -0.92

dw2 = 2\*2\*w2 = 2\*2\*1.2 = 4.8

* Cập nhật w1, w2

w1 = w1 – α\*dw1 = -4.6 – 0.4\*(-0.92) = -4.232

w2 = w2 – α\*dw2 = 1.2 – 0.4\*(4.8) = -0.72

**Bài 2 Gradient Descent + Momentum:**

A math equation with numbers and symbols

Description automatically generated

(a) Yêu cầu trình bày chi tiết từng bước thực hiện tìm điểm minimum theo thuật toán Gradient

Descent + Momentum (tìm w1 và w2 sau 2 epoch) với epoch = 2.

Các tham số sau khởi tạo w1 = −5, w2 = −2, v1 = 0, v2 = 0, α = 0.6, β = 0.5:

* **Bắt đầu với epoch = 1**
* STEP1: Tìm giá trị dw1 và dw2 giá trị đạo hàm của hàm (1) theo w1 và w2 (partial derivative) tại điểm khởi tạo [w1, w2]

dw1 = 2\*0.1\*w1 = 2\*0.1\*-5 = -1

dw2 = 2\*2\*w2 = 2\*2\*-2 = -8

* STEP2: Tìm giá trị v1 và v2 dựa vào dw1 và dw2 tìm được ở step 1 (công thức (2.1)).

v1 = β \* v1 + (1 – β)\* dw1 = 0.5 \* 0 + (1-0.5)\*(-1) = -0.5

v2 = β \* v2 + (1 – β)\* dw2 = 0.5 \* 0 + (1-0.5)\*(-8) = -4

* STEP3: Dùng công thức (2.2) Gradient Descent + Momentum để cập nhật w1 và w2.

w1 = w1 - α \* v1 = -5 – 0.6 \* (-0.5) = -4.7

w2 = w2 - α \* v2 = -2 – 0.6 \* (-4) = 0.4

* Hoàn thành epoch = 1
* **STEP4: epoch = 2 ta thực hiện tương tự với STEP1, STEP2 và STEP3 như trên với w1 và w2 đã được cập nhật từ epoch = 1**
* STEP1: Tìm giá trị dw1 và dw2 giá trị đạo hàm của hàm (1) theo w1 và w2 (partial derivative) tại điểm khởi tạo [w1, w2]

dw1 = 2\*0.1\*w1 = 2\*0.1\*-4.7 = -0.94

dw2 = 2\*2\*w2 = 2\*2\*0.4 = 1.6

* STEP2: Tìm giá trị v1 và v2 dựa vào dw1 và dw2 tìm được ở step 1 (công thức (2.1)).

v1 = β \* v1 + (1 – β)\* dw1 = 0.5 \* (-0.5) + (1-0.5)\*(-0.94) = -0.72

v2 = β \* v2 + (1 – β)\* dw2 = 0.5 \* (-4) + (1-0.5)\*(1.6) = -1.2

* STEP3: Dùng công thức (2.2) Gradient Descent + Momentum để cập nhật w1 và w2.

w1 = w1 - α \* v1 = -4.7 – 0.6 \* (0.72) = -4.268

w2 = w2 - α \* v2 = 0.4 – 0.6 \* (-1.2) = 1.12

* Hoàn thành epoch = 2

**Bài 3 RMSProp:**

A math equations with numbers and symbols

Description automatically generated

(a) Yêu cầu trình bày chi tiết từng bước thực hiện tìm điểm minimum theo thuật toán RMSProp

(tìm w1 và w2 sau 2 epoch) với epoch = 2.

Các tham số sau khởi tạo w1 = −5, w2 = −2, s1 = 0, s2 = 0, α = 0.3, γ = 0.9, ε = 10−6

* **Bắt đầu với epoch = 1**
* STEP1: Tìm giá trị dw1 và dw2 giá trị đạo hàm của hàm (1) theo w1 và w2 (partial derivative) tại điểm khởi tạo [w1, w2]

dw1 = 2\*0.1\*w1 = 2\*0.1\*-5 = -1

dw2 = 2\*2\*w2 = 2\*2\*-2 = -8

* STEP2: Tìm giá trị s1 và s2 dựa vào dw1 và dw2 vừa tìm được ở step 1 (công thức (3.1)).

S1 = γ \* S1  + (1- γ) \* dw12 = 0.9\*0 + (1-0.9)\*(-1)\*(-1) = 0.1

S2 = γ \* S2  + (1- γ) \* dw22 = 0.9\*0 + (1-0.9)\*(-8)\*(-8) = 6.4

* STEP3: Dùng công thức (3.2) để cập nhật w1 và w2.

w1 = w1 - α \* (dw1 / sqrt(S1 + ε))= -5 - 0.3 \* (-1)/sqrt(0.1 + 10−6) = -4.05132

w2 = w2 - α \* (dw2 / sqrt(S2 + ε))= -2 - 0.3 \* (-8)/sqrt(6.4 + 10−6) = -1.05132

* Hoàn thành epoch = 1
* **STEP4: epoch = 2 ta thực hiện tương tự với STEP1, STEP2 và STEP3 như trên với w1 và w2 đã được cập nhật từ epoch = 1**
* STEP1: Tìm giá trị dw1 và dw2 giá trị đạo hàm của hàm (1) theo w1 và w2 (partial derivative) tại điểm khởi tạo [w1, w2]

dw1 = 2\*0.1\*w1 = 2\*0.1\*(-4.05132) = -0.81026

dw2 = 2\*2\*w2 = 2\*2\*(-1.05132) = -4.20528

* STEP2: Tìm giá trị s1 và s2 dựa vào dw1 và dw2 vừa tìm được ở step 1 (công thức (3.1)).

S1 = γ \* S1  + (1- γ) \* dw12 = 0.9\*0.1 + (1-0.9)\*( -0.81026)\*( -0.81026) = 0.155652

S2 = γ \* S2  + (1- γ) \* dw22 = 0.9\*6.4 + (1-0.9)\*( -4.20528)\*( -4.20528) = 7.528438

* STEP3: Dùng công thức (3.2) để cập nhật w1 và w2.

w1 = w1 - α \* (dw1 / sqrt(S1 + ε))= -4.05132 - 0.3 \* (-0.81026)/sqrt(0.155652+ 10−6) = -3.4352

w2 = w2 - α \* (dw2 / sqrt(S2 + ε))= -1.05132 - 0.3 \* (-4.20528)/sqrt(7.528438+ 10−6) = -0.59153

* Hoàn thành epoch = 2

**Bài 4 Adam:**

A math equations and formulas

Description automatically generated

(a) Yêu cầu trình bày chi tiết từng bước thực hiện tìm điểm minimum theo thuật toán Adam (tìm w1 và w2 sau 2 epoch) với epoch = 2.

Các tham số sau khởi tạo w1 = −5, w2 = −2,

initial v1 = 0, v2 = 0, s1 = 0, s2 = 0,

α = 0.2, β1 = 0.9, β2 = 0.999, ε = 10−6:

* **Bắt đầu với epoch = 1**
* STEP1: Tìm giá trị dw1 và dw2 giá trị đạo hàm của hàm (1) theo w1 và w2 (partial derivative) tại điểm khởi tạo [w1, w2]

dw1 = 2\*0.1\*w1 = 2\*0.1\*-5 = -1

dw2 = 2\*2\*w2 = 2\*2\*-2 = -8

* STEP2: Tìm giá trị v1 và v2 dựa vào dw1 và dw2 vừa tìm được ở step 1 (công thức ở (4.1)).

V1 = β1 \* V1 + (1 − β1) \* dw1 = 0.9\*0 + (1-0.9)\*(-1) = -0.1

V2 = β1 \* V2 + (1 – β1) \* dw2 = 0.9\*0 + (1-0.9)\*(-8) = -0.8

* STEP3: Tìm giá trị s1 và s2 dựa vào dw1 và dw2 vừa tìm được ở step 1 (công thức (4.2)).

S1 = β2 \* S1  + (1- β2) \* dw12 = 0.999\*0 + (1-0.999)\*( -1)\*( -1) = 0.001

S2 = β2 \* S2  + (1- β2) \* dw22 = 0.999\*0 + (1-0.999)\*( -8)\*( -8) = 0.064

* STEP4: Thực hiện bias-correction cho V và S để thu được V\_coor và S\_coor. Sau khi áp dụng bias\_correction (4.3) và (4.4) ta sẽ thu được vcoor1, vcoor2, scoor1 và scoor2

Vcorr1 = V1 / (1 - β11) = (-0.1) / (1 - 0.9) = -1

Vcorr2 = V2 / (1 - β11) = (-0.8) / (1 - 0.9) = -8

Scorr1 = S1 / (1 – β21) = (0.001) / (1 - 0.999) = 1

Scorr2 = S2 / (1 – β21) = (0.064) / (1 - 0.999) = 64

* STEP5: Dùng công thức (4.5) để cập nhật w1 và w2.

W1 = W1 – α \* Vcorr1 / (sqrt(Scorr1 ) + ε) = -5 – 0.2 \* (-1) / (sqrt(1) + 10-6 ) = -4.8

W2 = W2 – α \* Vcorr2 / (sqrt(Scorr2 ) + ε) = -2 – 0.2 \* (-8) / (sqrt(64) + 10-6 ) = -1.8

* Hoàn thành epoch = 1
* **STEP6: epoch = 2 ta thực hiện tương tự với STEP1 - STEP5 như trên với w1 và w2 đã được cập nhật từ epoch = 1**
* STEP1: Tìm giá trị dw1 và dw2 giá trị đạo hàm của hàm (1) theo w1 và w2 (partial derivative) tại điểm khởi tạo [w1, w2]

dw1 = 2\*0.1\*w1 = 2\*0.1\*-4.8 = -0.96

dw2 = 2\*2\*w2 = 2\*2\*-1.8 = -0.72

* STEP2: Tìm giá trị v1 và v2 dựa vào dw1 và dw2 vừa tìm được ở step 1 (công thức ở (4.1)).

V1 = β1 \* V1 + (1 − β1) \* dw1 = 0.9\*(-0.1) + (1-0.9)\*(-0.96) = -0.186

V2 = β1 \* V2 + (1 – β1) \* dw2 = 0.9\*(-0.8) + (1-0.9)\*( -0.72) = -0.792

* STEP3: Tìm giá trị s1 và s2 dựa vào dw1 và dw2 vừa tìm được ở step 1 (công thức (4.2)).

S1 = β2 \* S1  + (1- β2) \* dw12 = 0.999\*0.001 + (1-0.999)\*( -0.96)\*( -0.96) = 0.0019206

S2 = β2 \* S2  + (1- β2) \* dw22 = 0.999\*0.064 + (1-0.999)\*( -0.72)\*( -0.72) = 0.0644544

* STEP4: Thực hiện bias-correction cho V và S để thu được V\_coor và S\_coor. Sau khi áp dụng bias\_correction (4.3) và (4.4) ta sẽ thu được vcoor1, vcoor2, scoor1 và scoor2

Vcorr1 = V1 / (1 - β12) = (-0.186) / (1 - 0.9\*0.9) = -0.978947368

Vcorr2 = V2 / (1 - β12) = (-0.792) / (1 - 0.9\*0.9) = -4.168421053

Scorr1 = S1 / (1 – β22) = (0.0019206) / (1 - 0.999\*0.999) = 0.96078039

Scorr2 = S2 / (1 – β22) = (0.0644544) / (1 - 0.999\*0.999) = 32.24332166

* STEP5: Dùng công thức (4.5) để cập nhật w1 và w2.

W1 = W1 – α \* Vcorr1 / (sqrt(Scorr1 ) + ε) = -4.8 - 0.2 \* (-0.978947368) / (sqrt(0.96078039) + 10-6 ) = -4.60025458

W2 = W2 – α \* Vcorr2 / (sqrt(Scorr2 ) + ε) = -1.8 - 0.2 \* (-4.168421053) / (sqrt(32.24332166) + 10-6 ) = -1.65318122

* Hoàn thành epoch = 2