Họ và tên: Trần Ngọc Đại

MSSV: 056

Module 05 - Week 03
Optimizer

1 Giới thiệu

Trong bài tập này, tôi sẽ trình bày lại chi tiết các bước thực hiện của một số thuật toán optimization cơ bản: Gradient Descent, Gradient Descent + Momentum, RMSProp, Adam. Đồng thời trình bày về vấn đề Vanishing cũng như thay đổi các optimizer để sự giảm thiểu vấn đề vanishing của từng loại thuật toán.
Các thuật toán sẽ được áp dụng trên hàm sau:

$$f(\theta_1, \theta_2) = 0.1\theta_1^2 + 2\theta_2^2 \tag{1}$$

2 Gradient Descent

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta J(\theta_t) \tag{2}$$

Trong đó:

- θ_t: Tham số tại bước t
- η: Tốc độ học (learning rate)
- $\nabla_\theta J(\theta_t)$: Gradient của hàm mất mát $J(\theta)$ tại θ_t

Hàm mục tiêu:

$$f(\theta_1, \theta_2) = 0.1\theta_1^2 + 2\theta_2^2$$

Tốc độ học $\eta = 0.4$, khởi tạo: $\theta_1^{(0)} = -5$, $\theta_2^{(0)} = -2$, số lượng epoch = 2.

Bước 1: Tính gradient của f
Gradient của $f(\theta_1, \theta_2)$:

$$\frac{\partial f}{\partial \theta_1} = 0.2\theta_1$$
$$\frac{\partial f}{\partial \theta_2} = 4\theta_2$$

Vậy gradient:

$$\nabla f = (0.2\theta_1, 4\theta_2)$$

Bước 2: Epoch 1

$$\theta_1^{(0)} = -5, \quad \theta_2^{(0)} = -2$$

Tính gradient tại $\theta_1^{(0)}$ và $\theta_2^{(0)}$:

$$\frac{\partial f}{\partial \theta_1} = 0.2(-5) = -1$$

$$\frac{\partial f}{\partial \theta_2} = 4(-2) = -8$$

Vậy gradient tại epoch 1 là:

$$\nabla f = (-1, -8)$$

Cập nhật tham số:

$$\theta_1^{(1)} = \theta_1^{(0)} - \eta \cdot \frac{\partial f}{\partial \theta_1} = -5 - 0.4(-1) = -5 + 0.4 = -4.6$$

$$\theta_2^{(1)} = \theta_2^{(0)} - \eta \cdot \frac{\partial f}{\partial \theta_2} = -2 - 0.4(-8) = -2 + 3.2 = 1.2$$

Kết quả sau epoch 1:

$$\theta_1^{(1)} = -4.6, \quad \theta_2^{(1)} = 1.2$$

Bước 3: Epoch 2

Tính gradient tại $\theta_1^{(1)} = -4.6$ và $\theta_2^{(1)} = 1.2$:

$$\frac{\partial f}{\partial \theta_1} = 0.2(-4.6) = -0.92$$

$$\frac{\partial f}{\partial \theta_2} = 4(1.2) = 4.8$$

Vậy gradient tại epoch 2 là:

$$\nabla f = (-0.92, 4.8)$$

Cập nhật tham số:

$$\theta_1^{(2)} = \theta_1^{(1)} - \eta \cdot \frac{\partial f}{\partial \theta_1} = -4.6 - 0.4(-0.92) = -4.6 + 0.368 = -4.232$$

$$\theta_2^{(2)} = \theta_2^{(1)} - \eta \cdot \frac{\partial f}{\partial \theta_2} = 1.2 - 0.4(4.8) = 1.2 - 1.92 = -0.72$$

Kết quả sau epoch 2:

$$\theta_1^{(2)} = -4.232, \quad \theta_2^{(2)} = -0.72$$

Tóm tắt kết quả

Epoch	θ_1	θ_2
0	-5.0	-2.0
1	-4.6	1.2
2	-4.232	-0.72

3 Gradient Descent with Momentum

$$v_t = \gamma v_{t-1} + \eta \nabla_\theta J(\theta_t) \tag{3}$$

$$\theta_{t+1} = \theta_t - v_t \tag{4}$$

Trong đó:

- v_t: Vector động lượng tại bước t
- γ: Hệ số động lượng (momentum coefficient)

Hàm mục tiêu:

$$f(\theta_1, \theta_2) = 0.1\theta_1^2 + 2\theta_2^2$$

Tốc độ học $\eta = 0.6$, Hệ số động lượng $\beta = 0.5$, Khởi tạo: $\theta_1^{(0)} = -5$, $\theta_2^{(0)} = -2$, $v_1^{(0)} = 0$, $v_2^{(0)} = 0$.

Bước 1: Tính gradient của f

Gradient của $f(\theta_1, \theta_2)$:

$$\frac{\partial f}{\partial \theta_1} = 0.2\theta_1$$

$$\frac{\partial f}{\partial \theta_2} = 4\theta_2$$

Bước 2: Epoch 1

Khởi tạo:

$$\theta_1^{(0)} = -5, \quad \theta_2^{(0)} = -2$$

$$v_1^{(0)} = 0, \quad v_2^{(0)} = 0$$

Tính gradient tại $\theta_1^{(0)}$ và $\theta_2^{(0)}$:

$$\frac{\partial f}{\partial \theta_1} = 0.2(-5) = -1$$

$$\frac{\partial f}{\partial \theta_2} = 4(-2) = -8$$

Cập nhật velocity:

$$v_1^{(1)} = 0.5 \cdot 0 + 0.5 \cdot (-1) = -0.5$$

$$v_2^{(1)} = 0.5 \cdot 0 + 0.5 \cdot (-8) = -4$$

Cập nhật tham số:

$$\theta_1^{(1)} = -5 - 0.6 \cdot (-0.5) = -4.7$$

$$\theta_2^{(1)} = -2 - 0.6 \cdot (-4) = 0.4$$

Bước 3: Epoch 2

Tính gradient tại $\theta_1^{(1)}$ và $\theta_2^{(1)}$:

$$\frac{\partial f}{\partial \theta_1} = 0.2(-4.7) = -0.94$$

$$\frac{\partial f}{\partial \theta_2} = 4(0.4) = 1.6$$

Cập nhật velocity:

$$v_1^{(2)} = 0.5 \cdot (-0.5) + 0.5 \cdot (-0.94) = -0.72$$

$$v_2^{(2)} = 0.5 \cdot (-4) + 0.5 \cdot 1.6 = -1.2$$

Cập nhật tham số:

$$\theta_1^{(2)} = -4.7 - 0.6 \cdot (-0.72) = -4.268$$

$$\theta_2^{(2)} = 0.4 - 0.6 \cdot (-1.2) = 1.12$$

Tóm tắt kết quả

Epoch	θ_1	θ_2
0	-5.0	-2.0
1	-4.7	0.4
2	-4.268	1.12

4 RMSProp

$$s_t = \beta s_{t-1} + (1-\beta) \left(\nabla_\theta J(\theta_t)\right)^2 \tag{5}$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{s_t + \epsilon}} \nabla_\theta J(\theta_t) \tag{6}$$

Trong đó:

- s_t: Trung bình bình phương trọng số
- β: Hệ số giảm trung bình động (decay rate)
- ϵ: Giá trị nhỏ để tránh chia cho 0

Cho hàm mục tiêu $f(\theta_1, \theta_2) = 0.1\theta_1^2 + 2\theta_2^2$, ta có các tham số khởi tạo như sau:

$$\theta_1^{(0)} = -5, \quad \theta_2^{(0)} = -2, \quad v_1^{(0)} = 0, \quad v_2^{(0)} = 0, \quad \eta = 0.3, \quad \beta = 0.9, \quad \epsilon = 10^{-6}$$

Bước 1: Tính toán tại Epoch 1 z **Tính gradient:**

$$\frac{\partial f}{\partial \theta_1} = 0.2 \cdot \theta_1 = 0.2 \times (-5) = -1$$

$$\frac{\partial f}{\partial \theta_2} = 4 \cdot \theta_2 = 4 \times (-2) = -8$$

Cập nhật bình phương gradient:

$$v_1^{(1)} = \beta \cdot v_1^{(0)} + (1-\beta) \cdot (\frac{\partial f}{\partial \theta_1})^2 = 0.9 \cdot 0 + 0.1 \cdot (-1)^2 = 0.1$$

$$v_2^{(1)} = \beta \cdot v_2^{(0)} + (1-\beta) \cdot (\frac{\partial f}{\partial \theta_2})^2 = 0.9 \cdot 0 + 0.1 \cdot (-8)^2 = 6.4$$

Cập nhật tham số:

$$\theta_1^{(1)} = \theta_1^{(0)} - \frac{\eta}{\sqrt{v_1^{(1)} + \epsilon}} \cdot \frac{\partial f}{\partial \theta_1} = -5 - \frac{0.3}{\sqrt{0.1 + 10^{-6}}} \times (-1) = -4.0513$$

$$\theta_2^{(1)} = \theta_2^{(0)} - \frac{\eta}{\sqrt{v_2^{(1)} + \epsilon}} \cdot \frac{\partial f}{\partial \theta_2} = -2 - \frac{0.3}{\sqrt{6.4 + 10^{-6}}} \times (-8) = -1.0513$$

Bước 2: Tính toán tại Epoch 2
Tính gradient:

$$\frac{\partial f}{\partial \theta_1} = 0.2 \cdot \theta_1 = 0.2 \times (-4.0513) = -0.81026$$

$$\frac{\partial f}{\partial \theta_2} = 4 \cdot \theta_2 = 4 \times (-1.0513) = -4.2052$$

Cập nhật bình phương gradient:

$$v_1^{(2)} = \beta \cdot v_1^{(1)} + (1-\beta) \cdot (\frac{\partial f}{\partial \theta_1})^2 = 0.9 \cdot 0.1 + 0.1 \cdot (-0.81026)^2 = 0.1 + 0.1 \cdot 0.6565 = 0.16565$$

$$v_2^{(2)} = \beta \cdot v_2^{(1)} + (1-\beta) \cdot (\frac{\partial f}{\partial \theta_2})^2 = 0.9 \cdot 6.4 + 0.1 \cdot (-4.2052)^2 = 5.76 + 0.1 \cdot 17.67 = 7.4367$$

Cập nhật tham số:

$$\theta_1^{(2)} = \theta_1^{(1)} - \frac{\eta}{\sqrt{v_1^{(2)} + \epsilon}} \cdot \frac{\partial f}{\partial \theta_1} = -4.0513 - \frac{0.3}{\sqrt{0.16565 + 10^{-6}}} \times (-0.81026) = -3.4531$$

$$\theta_2^{(2)} = \theta_2^{(1)} - \frac{\eta}{\sqrt{v_2^{(2)} + \epsilon}} \cdot \frac{\partial f}{\partial \theta_2} = -1.0513 - \frac{0.3}{\sqrt{7.4367 + 10^{-6}}} \times (-4.2052) = -0.5915$$

Kết quả

Epoch	θ_1	θ_2	v_1	v_2
0	-5.0	-2.0	0	0
1	-4.0513	-1.0513	0.1	6.4
2	-3.4531	-0.5915	0.16565	7.4367

5 Adam

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\nabla_\theta J(\theta_t) \tag{7}$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2)\left(\nabla_\theta J(\theta_t)\right)^2 \tag{8}$$

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1-\beta_2^t} \tag{9}$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon}\hat{m}_t \tag{10}$$

Trong đó:

- m_t: Trung bình động bậc nhất của gradient
- v_t: Trung bình động bậc hai của gradient
- $\hat{m}_t$, $\hat{v}_t$: Ước tính không thiên lệch của m_t và v_t

- β_1, β_2: Hệ số giảm trung bình động

Khởi tạo và thông số Khởi tạo:

$$\theta_1 = -5, \quad \theta_2 = -2$$

Momen:

$$v_1 = 0, \quad v_2 = 0, \quad s_1 = 0, \quad s_2 = 0$$

Các tham số động:

$$\eta = 0.2, \quad \beta_1 = 0.9, \quad \beta_2 = 0.999, \quad \epsilon = 10^{-6}$$

Hàm mục tiêu:

$$f(\theta_1, \theta_2) = 0.1\theta_1^2 + 2\theta_2^2$$

Gradient:

$$\frac{\partial f}{\partial \theta_1} = 0.2\theta_1, \quad \frac{\partial f}{\partial \theta_2} = 4\theta_2$$

Bước 1: Tính gradient Tính gradient tại $\theta_1 = -5$, $\theta_2 = -2$:

$$\nabla_{\theta_1} = 0.2 \times (-5) = -1, \quad \nabla_{\theta_2} = 4 \times (-2) = -8$$

Bước 2: Cập nhật momen m_t Momen được tính theo công thức:

$$m_1^{(t)} = \beta_1 m_1^{(t-1)} + (1 - \beta_1)\nabla_{\theta_1}$$

$$m_2^{(t)} = \beta_1 m_2^{(t-1)} + (1 - \beta_1)\nabla_{\theta_2}$$

Với $m_1^{(0)} = 0$ và $m_2^{(0)} = 0$, ta có:

$$m_1^{(1)} = -0.1, \quad m_2^{(1)} = -0.8$$

Bước 3: Cập nhật độ biến thiên v_t Độ biến thiên được tính bằng cách sử dụng bình phương gradient:

$$v_1^{(t)} = \beta_2 v_1^{(t-1)} + (1 - \beta_2)(\nabla_{\theta_1})^2$$

$$v_2^{(t)} = \beta_2 v_2^{(t-1)} + (1 - \beta_2)(\nabla_{\theta_2})^2$$

Với $v_1^{(0)} = 0$ và $v_2^{(0)} = 0$, ta có:

$$v_1^{(1)} = 0.001, \quad v_2^{(1)} = 0.064$$

Bước 4: Bias correction

$$\hat{m_1}^{(1)} = \frac{m_1^{(1)}}{1 - \beta_1^1} = -1, \quad \hat{m_2}^{(1)} = \frac{m_2^{(1)}}{1 - \beta_1^1} = -8$$

$$\hat{v_1}^{(1)} = \frac{v_1^{(1)}}{1 - \beta_2^1} = 1, \quad \hat{v_2}^{(1)} = \frac{v_2^{(1)}}{1 - \beta_2^1} = 64$$

Bước 5: Cập nhật tham số

$$\theta_1^{(1)} = -4.8000001, \quad \theta_2^{(1)} = -1.8$$

Epoch 2: Tiếp tục cập nhật tham số

Tính gradient

$$\nabla_{\theta_1} = 0.2 \times (-4.8) = -0.96, \quad \nabla_{\theta_2} = 4 \times (-1.8) = -7.2$$

Cập nhật momen và độ biến thiên

$$m_1^{(2)} = -0.196, \quad m_2^{(2)} = -1.36$$
$$v_1^{(2)} = 0.0019232, \quad v_2^{(2)} = 0.12144$$

Bias correction

$$\hat{m_1}^{(2)} = -1.96, \quad \hat{m_2}^{(2)} = -13.6$$
$$\hat{v_1}^{(2)} = 1.923, \quad \hat{v_2}^{(2)} = 121.44$$

Cập nhật tham số

$$\theta_1^{(2)} = -4.60025458, \quad \theta_2^{(2)} = -1.60082446$$

Bảng tóm tắt kết quả

Epoch	θ_1	θ_2
1	-4.8000001	-1.8
2	-4.60025458	-1.60082446