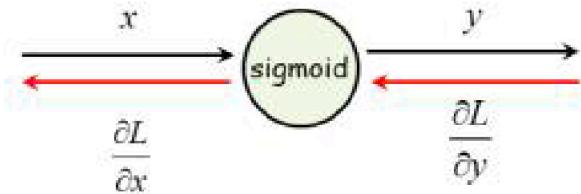


문제 1. Sigmoid 함수의 Backpropagation

sigmoid 층으로 data batch 를 $X = \begin{pmatrix} \log 2 & \log 3 & \log 4 \\ \log 5 & \log 6 & \log 7 \end{pmatrix}$ 가 입력되고, $\frac{\partial L}{\partial Y} = \begin{pmatrix} 3^2 & 4^2 & 5^2 \\ 6^2 & 7^2 & 8^2 \end{pmatrix}$

일 때 $\frac{\partial L}{\partial X}$ 를 구하세요.



주어진 정보

- 입력:

$$X = \begin{pmatrix} \log 2 & \log 3 & \log 4 \\ \log 5 & \log 6 & \log 7 \end{pmatrix}$$

$$\frac{\partial L}{\partial Y} = \begin{pmatrix} 3^2 & 4^2 & 5^2 \\ 6^2 & 7^2 & 8^2 \end{pmatrix} = \begin{pmatrix} 9 & 16 & 25 \\ 36 & 49 & 64 \end{pmatrix}$$

- 구해야 것:

$$\frac{\partial L}{\partial X}$$

풀이 과정

Step 1: Sigmoid 함수의 정의와 미분

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

$$\frac{\partial \sigma}{\partial x} = \sigma(x)(1 - \sigma(x))$$

Step 2: 역전파 공식 적용

$$\frac{\partial L}{\partial X} = \frac{\partial L}{\partial Y} \odot \frac{\partial Y}{\partial X} = \frac{\partial L}{\partial Y} \odot Y \odot (1 - Y)$$

여기서 \odot 은 element-wise 곱셈(Hadamard product)

Step 3: Forward pass - Y 계산

$$Y = \sigma(X)$$

각 원소별로 계산:

- $y_{11} = \sigma(\log 2) = \frac{1}{1+e^{-\log 2}} = \frac{1}{1+\frac{1}{2}} = \frac{2}{3}$
- $y_{12} = \sigma(\log 3) = \frac{1}{1+e^{-\log 3}} = \frac{1}{1+\frac{1}{3}} = \frac{3}{4}$
- $y_{13} = \sigma(\log 4) = \frac{1}{1+e^{-\log 4}} = \frac{1}{1+\frac{1}{4}} = \frac{4}{5}$
- $y_{21} = \sigma(\log 5) = \frac{1}{1+e^{-\log 5}} = \frac{1}{1+\frac{1}{5}} = \frac{5}{6}$
- $y_{22} = \sigma(\log 6) = \frac{1}{1+e^{-\log 6}} = \frac{1}{1+\frac{1}{6}} = \frac{6}{7}$

- $y_{23} = \sigma(\log 7) = \frac{1}{1+e^{-\log 7}} = \frac{1}{1+\frac{1}{7}} = \frac{7}{8}$

따라서:

$$Y = \begin{pmatrix} \frac{2}{3} & \frac{3}{4} & \frac{4}{5} \\ \frac{5}{6} & \frac{6}{7} & \frac{7}{8} \end{pmatrix}$$

Step 4: $1 - Y$ 계산

$$1 - Y = \begin{pmatrix} \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{6} & \frac{1}{7} & \frac{1}{8} \end{pmatrix}$$

Step 5: $Y \odot (1 - Y)$ 계산

$$Y \odot (1 - Y) = \begin{pmatrix} \frac{2}{3} \cdot \frac{1}{3} & \frac{3}{4} \cdot \frac{1}{4} & \frac{4}{5} \cdot \frac{1}{5} \\ \frac{5}{6} \cdot \frac{1}{6} & \frac{6}{7} \cdot \frac{1}{7} & \frac{7}{8} \cdot \frac{1}{8} \end{pmatrix} = \begin{pmatrix} \frac{2}{9} & \frac{3}{16} & \frac{4}{25} \\ \frac{5}{36} & \frac{6}{49} & \frac{7}{64} \end{pmatrix}$$

Step 6: 최종 계산

$$\begin{aligned} \frac{\partial L}{\partial X} &= \frac{\partial L}{\partial Y} \odot Y \odot (1 - Y) \\ &= \begin{pmatrix} 9 & 16 & 25 \\ 36 & 49 & 64 \end{pmatrix} \odot \begin{pmatrix} \frac{2}{9} & \frac{3}{16} & \frac{4}{25} \\ \frac{5}{36} & \frac{6}{49} & \frac{7}{64} \end{pmatrix} \\ &= \begin{pmatrix} 9 \cdot \frac{2}{9} & 16 \cdot \frac{3}{16} & 25 \cdot \frac{4}{25} \\ 36 \cdot \frac{5}{36} & 49 \cdot \frac{6}{49} & 64 \cdot \frac{7}{64} \end{pmatrix} \end{aligned}$$

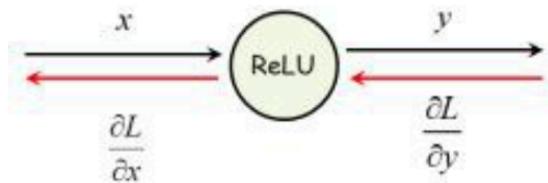
답

$$\frac{\partial L}{\partial X} = \begin{pmatrix} 2 & 3 & 4 \\ 5 & 6 & 7 \end{pmatrix}$$

문제 2. ReLU 함수의 Backpropagation

ReLU 층으로 data batch 묶음 $X = \begin{pmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \end{pmatrix}$ 가 입력되고, $\frac{\partial L}{\partial Y} = \begin{pmatrix} 1 & 2 & -3 \\ 4 & 5 & -6 \end{pmatrix}$ 일 때

$\frac{\partial L}{\partial X}$ 를 구하세요.



주어진 정보

- 입력: $X = \begin{pmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \end{pmatrix}$
- $\frac{\partial L}{\partial Y} = \begin{pmatrix} 1 & -2 & -3 \\ 4 & 5 & -6 \end{pmatrix}$
- 구하는 것: $\frac{\partial L}{\partial X}$

풀이 과정

Step 1: ReLU 함수의 정의

$$\text{ReLU}(x) = \max(0, x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

Step 2: ReLU의 미분

$$\frac{\partial \text{ReLU}(x)}{\partial x} = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

Step 3: Forward pass - Y 계산

$$Y = \text{ReLU}(X)$$

각 원소별로:

- $y_{11} = \text{ReLU}(1) = 1$
- $y_{12} = \text{ReLU}(-2) = 0$
- $y_{13} = \text{ReLU}(3) = 3$
- $y_{21} = \text{ReLU}(-4) = 0$
- $y_{22} = \text{ReLU}(5) = 5$
- $y_{23} = \text{ReLU}(-6) = 0$

따라서:

$$Y = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 5 & 0 \end{pmatrix}$$

Step 4: 미분 마스크 계산

$$\frac{\partial Y}{\partial X} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

(입력 X 에서 양수인 위치는 1, 음수 또는 0인 위치는 0)

Step 5: 역전파 계산

$$\begin{aligned} \frac{\partial L}{\partial X} &= \frac{\partial L}{\partial Y} \odot \frac{\partial Y}{\partial X} \\ &= \begin{pmatrix} 1 & -2 & -3 \\ 4 & 5 & -6 \end{pmatrix} \odot \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 \cdot 1 & -2 \cdot 0 & -3 \cdot 1 \\ 4 \cdot 0 & 5 \cdot 1 & -6 \cdot 0 \end{pmatrix} \end{aligned}$$

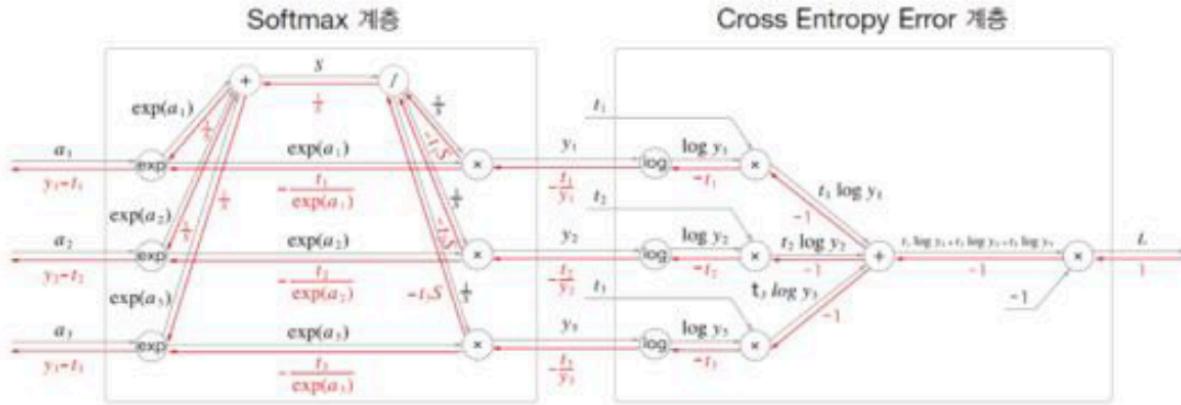
답

$$\boxed{\frac{\partial L}{\partial X} = \begin{pmatrix} 1 & 0 & -3 \\ 0 & 5 & 0 \end{pmatrix}}$$

문제 3. SoftmaxWithLoss의 Backpropagation

softmaxwithloss 층으로 data batch 묶음 $X = (\log 2 \quad \log 3)$ 가 입력되고, target label t 이

(1 0) 일 때 $\frac{\partial L}{\partial X}$ 를 구하세요.



주어진 정보

- 입력: $X = (\log 2 \quad \log 3)$ (단일 샘플로 해석)
- Target label: $t = (1, 0)$ (첫 번째 클래스가 정답)
- 구하는 것:

$$\frac{\partial L}{\partial X}$$

풀이 과정

Step 1: Softmax 함수 정의

$$\text{softmax}(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

Step 2: Forward pass - Softmax 출력 계산

입력이 $x_1 = \log 2, x_2 = \log 3$ 일 때:

$$y_1 = \frac{e^{\log 2}}{e^{\log 2} + e^{\log 3}} = \frac{2}{2+3} = \frac{2}{5}$$

$$y_2 = \frac{e^{\log 3}}{e^{\log 2} + e^{\log 3}} = \frac{3}{2+3} = \frac{3}{5}$$

따라서: $Y = \left(\frac{2}{5} \quad \frac{3}{5} \right)$

Step 3: Cross Entropy Loss 계산

$$L = -\sum_i t_i \log y_i = -\left(1 \cdot \log \frac{2}{5} + 0 \cdot \log \frac{3}{5}\right) = -\log \frac{2}{5} = \log \frac{5}{2}$$

Step 4: SoftmaxWithLoss의 역전파 공식

SoftmaxWithLoss 층의 역전파는 다음과 같이 간단하게 표현됩니다:

$$\frac{\partial L}{\partial X} = Y - T$$

여기서:

- Y : Softmax 출력
- T : Target one-hot vector

Step 5: 최종 계산

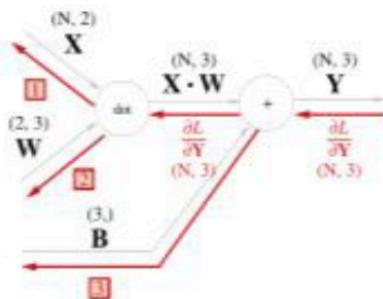
$$\begin{aligned}\frac{\partial L}{\partial X} &= Y - T = \left(\frac{2}{5} \quad \frac{3}{5}\right) - (1 \quad 0) \\ &= \left(\frac{2}{5} - 1 \quad \frac{3}{5} - 0\right) = \left(-\frac{3}{5} \quad \frac{3}{5}\right)\end{aligned}$$

답

$$\boxed{\frac{\partial L}{\partial X} = \left(-\frac{3}{5} \quad \frac{3}{5}\right)}$$

문제 4. Affine Layer의 Backpropagation

4. (affine layer 의 backpropagation) $X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$, $W = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$, $B = (7 \ 8 \ 9)$ 으로 주어져 있고, $\frac{\partial L}{\partial Y} = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ 일 때, Affine 층의 계산 그래프를 이용하여 $\frac{\partial L}{\partial X}$, $\frac{\partial L}{\partial W}$, $\frac{\partial L}{\partial B}$ 을 구하세요.



주어진 정보

- $X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$ (3×2 행렬, batch size=3)
- $W = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ (2×3 행렬)
- $B = (7 \ 8 \ 9)$ (1×3 행렬)
- $\frac{\partial L}{\partial Y} = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ (3×3 행렬)
- 구하는 것:

$$\frac{\partial L}{\partial X}, \frac{\partial L}{\partial W}, \frac{\partial L}{\partial B}$$

풀이 과정

Step 1: Affine 층의 정의

$$Y = XW + B$$

여기서:

- X : (N, D) 입력
- W : (D, M) 가중치
- B : (1, M) 편향 (브로드캐스팅)
- Y : (N, M) 출력

Step 2: Forward pass 확인

$$\begin{aligned}
 Y &= XW + B \\
 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 7 & 8 & 9 \end{pmatrix} \\
 &= \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 5 & 7 & 9 \end{pmatrix} + \begin{pmatrix} 7 & 8 & 9 \\ 7 & 8 & 9 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 8 & 10 & 12 \\ 11 & 13 & 15 \\ 12 & 15 & 18 \end{pmatrix}
 \end{aligned}$$

Step 3: $\frac{\partial L}{\partial X}$ 계산

역전파 공식:

$$\begin{aligned}
 \frac{\partial L}{\partial X} &= \frac{\partial L}{\partial Y} \cdot W^T \\
 &= \begin{pmatrix} 2 & 1 & -1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}
 \end{aligned}$$

각 원소 계산:

- 행 1: $(2 \cdot 1 + 1 \cdot 2 + (-1) \cdot 3, 2 \cdot 4 + 1 \cdot 5 + (-1) \cdot 6) = (1, 7)$
- 행 2: $(1 \cdot 1 + 0 \cdot 2 + 0 \cdot 3, 1 \cdot 4 + 0 \cdot 5 + 0 \cdot 6) = (1, 4)$
- 행 3: $(0 \cdot 1 + 0 \cdot 2 + 1 \cdot 3, 0 \cdot 4 + 0 \cdot 5 + 1 \cdot 6) = (3, 6)$

$$\frac{\partial L}{\partial X} = \begin{pmatrix} 1 & 7 \\ 1 & 4 \\ 3 & 6 \end{pmatrix}$$

Step 4: $\frac{\partial L}{\partial W}$ 계산

역전파 공식:

$$\begin{aligned}
 \frac{\partial L}{\partial W} &= X^T \cdot \frac{\partial L}{\partial Y} \\
 &= \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & -1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}
 \end{aligned}$$

각 원소 계산:

- 행 1: $(1 \cdot 2 + 0 \cdot 1 + 1 \cdot 0, 1 \cdot 1 + 0 \cdot 0 + 1 \cdot 0, 1 \cdot (-1) + 0 \cdot 0 + 1 \cdot 1) = (2, 1, 0)$
- 행 2: $(0 \cdot 2 + 1 \cdot 1 + 1 \cdot 0, 0 \cdot 1 + 1 \cdot 0 + 1 \cdot 0, 0 \cdot (-1) + 1 \cdot 0 + 1 \cdot 1) = (1, 0, 1)$

$$\frac{\partial L}{\partial W} = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

Step 5: $\frac{\partial L}{\partial B}$ 계산

역전파 공식 (배치 차원에 대한 합):

$$\begin{aligned}
 \frac{\partial L}{\partial B} &= \sum_{i=1}^N \frac{\partial L}{\partial Y_i} \\
 &= \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}^T = (3 \ 1 \ 0)
 \end{aligned}$$

열별로 합산:

- 첫 번째 열: $2 + 1 + 0 = 3$

- 두 번째 열: $1 + 0 + 0 = 1$
- 세 번째 열: $-1 + 0 + 1 = 0$

$$\frac{\partial L}{\partial B} = (3 \quad 1 \quad 0)$$

답

$$\frac{\partial L}{\partial X} = \begin{pmatrix} 1 & 7 \\ 1 & 4 \\ 3 & 6 \end{pmatrix}, \quad \frac{\partial L}{\partial W} = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad \frac{\partial L}{\partial B} = (3 \quad 1 \quad 0)$$

문제 5. 대칭변환의 Backpropagation

5. (대칭변환) $X=(x_1, x_2, x_3)$ 가 $Y=(x_2, x_3, x_1)$ 으로 변환되는 layer가 있습니다. $\frac{\partial L}{\partial Y} = (d_1, d_2, d_3)$

일 때 $\frac{\partial L}{\partial X}$ 를 구하세요. (힌트: $X \times \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = Y$)

주어진 정보

- 변환: $X = (x_1, x_2, x_3) \rightarrow Y = (x_2, x_3, x_1)$
- $\frac{\partial L}{\partial Y} = (d_1, d_2, d_3)$
- 힌트: $Y = X \times \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$
- 구하는 것: $\frac{\partial L}{\partial X}$

풀이 과정

Step 1: 변환 행렬의 이해

변환 행렬을 P 라 하면:

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

이는 순환 순열(cyclic permutation) 행렬입니다.

Step 2: Forward pass 확인

$$Y = XP$$

$$(x_1 \quad x_2 \quad x_3) \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = (x_2 \quad x_3 \quad x_1)$$

확인:

- $y_1 = 0 \cdot x_1 + 1 \cdot x_2 + 0 \cdot x_3 = x_2 \checkmark$
- $y_2 = 0 \cdot x_1 + 0 \cdot x_2 + 1 \cdot x_3 = x_3 \checkmark$
- $y_3 = 1 \cdot x_1 + 0 \cdot x_2 + 0 \cdot x_3 = x_1 \checkmark$

Step 3: 역전파 공식 적용

행렬 곱의 역전파:

$$\frac{\partial L}{\partial X} = \frac{\partial L}{\partial Y} \cdot P^T$$

Step 4: P^T 계산

$$P^T = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

Step 5: 최종 계산

$$\frac{\partial L}{\partial X} = (d_1 \quad d_2 \quad d_3) \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

각 원소 계산:

- $\frac{\partial L}{\partial x_1} = 0 \cdot d_1 + 0 \cdot d_2 + 1 \cdot d_3 = d_3$
- $\frac{\partial L}{\partial x_2} = 1 \cdot d_1 + 0 \cdot d_2 + 0 \cdot d_3 = d_1$
- $\frac{\partial L}{\partial x_3} = 0 \cdot d_1 + 1 \cdot d_2 + 0 \cdot d_3 = d_2$

Step 6: 직관적 해석

역변환을 생각하면:

- $Y = (x_2, x_3, x_1)$ 이므로
- $y_1 = x_2$ 에서 $\frac{\partial L}{\partial x_2} = \frac{\partial L}{\partial y_1} = d_1$
- $y_2 = x_3$ 에서 $\frac{\partial L}{\partial x_3} = \frac{\partial L}{\partial y_2} = d_2$
- $y_3 = x_1$ 에서 $\frac{\partial L}{\partial x_1} = \frac{\partial L}{\partial y_3} = d_3$

답

$$\frac{\partial L}{\partial X} = (d_3, d_1, d_2)$$

문제 6. Momentum Algorithm

6. (Momentum algorithm) 이변수 함수 $f(x, y) = x^2 + xy$ 를 momentum 방법으로 최적화하고자 합니다. 초기 위치 $x_0 = (1, 1)$ 에서 출발하여 3 step 진행할 때, x_1, x_2, x_3 를 구하세요. (단 learning rate $\eta = 1$, momentum 계수 $\alpha = 1$)

주어진 정보

- 목적 함수: $f(x, y) = x^2 + xy$
- 초기 위치: $x_0 = (1, 1)$
- Learning rate: $\eta = 1$
- Momentum 계수: $\alpha = 1$
- 구하는 것: 3 step 진행 후 x_1, x_2, x_3

풀이 과정

Step 1: Momentum 알고리즘 공식

$$v_{t+1} = \alpha v_t - \eta \nabla f(x_t)$$

$$x_{t+1} = x_t + v_{t+1}$$

초기값: $v_0 = (0, 0)$

Step 2: 그래디언트 계산

$$\nabla f(x, y) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} 2x + y \\ x \end{pmatrix}$$

Step 3: Step 1 계산 ($x_0 = (1, 1)$)

그래디언트:

$$\nabla f(1, 1) = \begin{pmatrix} 2(1) + 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

Velocity 업데이트:

$$v_1 = \alpha v_0 - \eta \nabla f(x_0) = 1 \cdot \begin{pmatrix} 0 \\ 0 \end{pmatrix} - 1 \cdot \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} -3 \\ -1 \end{pmatrix}$$

위치 업데이트:

$$x_1 = x_0 + v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} -3 \\ -1 \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

Step 4: Step 2 계산 ($x_1 = (-2, 0)$)

그래디언트:

$$\nabla f(-2, 0) = \begin{pmatrix} 2(-2) + 0 \\ -2 \end{pmatrix} = \begin{pmatrix} -4 \\ -2 \end{pmatrix}$$

Velocity 업데이트:

$$v_2 = \alpha v_1 - \eta \nabla f(x_1) = 1 \cdot \begin{pmatrix} -3 \\ -1 \end{pmatrix} - 1 \cdot \begin{pmatrix} -4 \\ -2 \end{pmatrix} = \begin{pmatrix} -3 + 4 \\ -1 + 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

위치 업데이트:

$$x_2 = x_1 + v_2 = \begin{pmatrix} -2 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Step 5: Step 3 계산 ($x_2 = (-1, 1)$)

그래디언트:

$$\nabla f(-1, 1) = \begin{pmatrix} 2(-1) + 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

Velocity 업데이트:

$$v_3 = \alpha v_2 - \eta \nabla f(x_2) = 1 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 1 \cdot \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 + 1 \\ 1 + 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

위치 업데이트:

$$x_3 = x_2 + v_3 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

답

$$x_1 = (-2, 0), \quad x_2 = (-1, 1), \quad x_3 = (1, 3)$$

문제 7. NAG (Nesterov Accelerated Gradient) Algorithm

7. (NAG algorithm) 이변수 함수 $f(x, y) = x^2 + xy$ 를 NAG 방법으로 최적화하고자 합니다. 초기 위치 $x_0 = (1, 1)$ 에서 출발하여 3 step 진행할 때, x_1, x_2, x_3 를 구하세요. (단 learning rate $\eta = 1$, momentum 계수 $\alpha = 1$)

주어진 정보

- 목적 함수: $f(x, y) = x^2 + xy$
- 초기 위치: $x_0 = (1, 1)$
- Learning rate: $\eta = 1$
- Momentum 계수: $\alpha = 1$
- 구하는 것: 3 step 진행 후 x_1, x_2, x_3

풀이 과정

Step 1: NAG 알고리즘 공식

NAG는 look-ahead 위치에서 그래디언트를 계산합니다:

$$v_{t+1} = \alpha v_t - \eta \nabla f(x_t + \alpha v_t)$$

$$x_{t+1} = x_t + v_{t+1}$$

초기값: $v_0 = (0, 0)$

Step 2: 그래디언트 함수 (재확인)

$$\nabla f(x, y) = \begin{pmatrix} 2x + y \\ x \end{pmatrix}$$

Step 3: Step 1 계산 ($x_0 = (1, 1), v_0 = (0, 0)$)

Look-ahead 위치:

$$x_0 + \alpha v_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 1 \cdot \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

그래디언트 (look-ahead 위치에서):

$$\nabla f(1, 1) = \begin{pmatrix} 2(1) + 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

Velocity 업데이트:

$$v_1 = \alpha v_0 - \eta \nabla f(x_0 + \alpha v_0) = 1 \cdot \begin{pmatrix} 0 \\ 0 \end{pmatrix} - 1 \cdot \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} -3 \\ -1 \end{pmatrix}$$

위치 업데이트:

$$x_1 = x_0 + v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} -3 \\ -1 \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

Step 4: Step 2 계산 ($x_1 = (-2, 0), v_1 = (-3, -1)$)

Look-ahead 위치:

$$x_1 + \alpha v_1 = \begin{pmatrix} -2 \\ 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} -3 \\ -1 \end{pmatrix} = \begin{pmatrix} -5 \\ -1 \end{pmatrix}$$

그래디언트 (look-ahead 위치에서):

$$\nabla f(-5, -1) = \begin{pmatrix} 2(-5) + (-1) \\ -5 \end{pmatrix} = \begin{pmatrix} -11 \\ -5 \end{pmatrix}$$

Velocity 업데이트:

$$v_2 = \alpha v_1 - \eta \nabla f(x_1 + \alpha v_1) = 1 \cdot \begin{pmatrix} -3 \\ -1 \end{pmatrix} - 1 \cdot \begin{pmatrix} -11 \\ -5 \end{pmatrix} = \begin{pmatrix} -3 + 11 \\ -1 + 5 \end{pmatrix} = \begin{pmatrix} 8 \\ 4 \end{pmatrix}$$

위치 업데이트:

$$x_2 = x_1 + v_2 = \begin{pmatrix} -2 \\ 0 \end{pmatrix} + \begin{pmatrix} 8 \\ 4 \end{pmatrix} = \begin{pmatrix} 6 \\ 4 \end{pmatrix}$$

Step 5: Step 3 계산 ($x_2 = (6, 4)$, $v_2 = (8, 4)$)

Look-ahead 위치:

$$x_2 + \alpha v_2 = \begin{pmatrix} 6 \\ 4 \end{pmatrix} + 1 \cdot \begin{pmatrix} 8 \\ 4 \end{pmatrix} = \begin{pmatrix} 14 \\ 8 \end{pmatrix}$$

그래디언트 (look-ahead 위치에서):

$$\nabla f(14, 8) = \begin{pmatrix} 2(14) + 8 \\ 14 \end{pmatrix} = \begin{pmatrix} 36 \\ 14 \end{pmatrix}$$

Velocity 업데이트:

$$v_3 = \alpha v_2 - \eta \nabla f(x_2 + \alpha v_2) = 1 \cdot \begin{pmatrix} 8 \\ 4 \end{pmatrix} - 1 \cdot \begin{pmatrix} 36 \\ 14 \end{pmatrix} = \begin{pmatrix} 8 - 36 \\ 4 - 14 \end{pmatrix} = \begin{pmatrix} -28 \\ -10 \end{pmatrix}$$

위치 업데이트:

$$x_3 = x_2 + v_3 = \begin{pmatrix} 6 \\ 4 \end{pmatrix} + \begin{pmatrix} -28 \\ -10 \end{pmatrix} = \begin{pmatrix} -22 \\ -6 \end{pmatrix}$$

답

$x_1 = (-2, 0), \quad x_2 = (6, 4), \quad x_3 = (-22, -6)$

문제 8. AdaGrad Algorithm

8. (AdaGrad algorithm) 이변수 함수 $f(x, y) = x^2 + xy$ 를 AdaGrad 방법으로 최적화하고자 합니다. 초기 위치 $x_0 = (1, 1)$ 에서 출발하여 2 step 진행할 때, x_1 , x_2 를 구하세요. (단 learning rate $\eta = \frac{1}{2}$)

주어진 정보

- 목적 함수: $f(x, y) = x^2 + xy$
- 초기 위치: $x_0 = (1, 1)$
- Learning rate: $\eta = \frac{1}{2}$
- 구하는 것: 2 step 진행 후 x_1, x_2

풀이 과정

Step 1: AdaGrad 알고리즘 공식

$$h_{t+1} = h_t + (\nabla f(x_t))^2$$

$$x_{t+1} = x_t - \eta \frac{\nabla f(x_t)}{\sqrt{h_{t+1}} + \epsilon}$$

여기서:

- $h_0 = (0, 0)$ (초기 누적 제곱 그래디언트)

- ϵ 은 매우 작은 값 (보통 10^{-7} , 여기서는 0으로 가정)
- 제곱과 나눗셈은 element-wise 연산

Step 2: 그래디언트 함수

$$\nabla f(x, y) = \begin{pmatrix} 2x + y \\ x \end{pmatrix}$$

Step 3: Step 1 계산 ($x_0 = (1, 1)$)

그래디언트:

$$\nabla f(1, 1) = \begin{pmatrix} 2(1) + 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

누적 제곱 그래디언트 업데이트:

$$h_1 = h_0 + (\nabla f(x_0))^2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 3^2 \\ 1^2 \end{pmatrix} = \begin{pmatrix} 9 \\ 1 \end{pmatrix}$$

위치 업데이트:

$$\begin{aligned} x_1 &= x_0 - \eta \frac{\nabla f(x_0)}{\sqrt{h_1}} \\ &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{2} \cdot \frac{\begin{pmatrix} 3 \\ 1 \end{pmatrix}}{\begin{pmatrix} \sqrt{9} \\ \sqrt{1} \end{pmatrix}} \\ &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{2} \cdot \begin{pmatrix} \frac{3}{3} \\ \frac{1}{1} \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{2} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 - 0.5 \\ 1 - 0.5 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} \end{aligned}$$

Step 4: Step 2 계산 ($x_1 = (0.5, 0.5)$)

그래디언트:

$$\nabla f(0.5, 0.5) = \begin{pmatrix} 2(0.5) + 0.5 \\ 0.5 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 0.5 \end{pmatrix}$$

누적 제곱 그래디언트 업데이트:

$$h_2 = h_1 + (\nabla f(x_1))^2 = \begin{pmatrix} 9 \\ 1 \end{pmatrix} + \begin{pmatrix} 1.5^2 \\ 0.5^2 \end{pmatrix} = \begin{pmatrix} 9 + 2.25 \\ 1 + 0.25 \end{pmatrix} = \begin{pmatrix} 11.25 \\ 1.25 \end{pmatrix}$$

위치 업데이트:

$$\begin{aligned} x_2 &= x_1 - \eta \frac{\nabla f(x_1)}{\sqrt{h_2}} \\ &= \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} - \frac{1}{2} \cdot \frac{\begin{pmatrix} 1.5 \\ 0.5 \end{pmatrix}}{\begin{pmatrix} \sqrt{11.25} \\ \sqrt{1.25} \end{pmatrix}} \end{aligned}$$

계산:

- $\sqrt{11.25} = \sqrt{\frac{45}{4}} = \frac{3\sqrt{5}}{2}$
- $\sqrt{1.25} = \sqrt{\frac{5}{4}} = \frac{\sqrt{5}}{2}$

$$\begin{aligned}
&= \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} - \frac{1}{2} \cdot \begin{pmatrix} \frac{1.5}{2} \\ \frac{0.5}{\frac{\sqrt{5}}{2}} \end{pmatrix} \\
&= \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} - \frac{1}{2} \cdot \begin{pmatrix} \frac{1.5 \cdot 2}{3\sqrt{5}} \\ \frac{0.5 \cdot 2}{\sqrt{5}} \end{pmatrix} \\
&= \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} - \frac{1}{2} \cdot \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} \\
&= \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} - \begin{pmatrix} \frac{1}{2\sqrt{5}} \\ \frac{1}{2\sqrt{5}} \end{pmatrix} \\
&= \begin{pmatrix} 0.5 - \frac{\sqrt{5}}{10} \\ 0.5 - \frac{\sqrt{5}}{10} \end{pmatrix} = \begin{pmatrix} \frac{5-\sqrt{5}}{10} \\ \frac{5-\sqrt{5}}{10} \end{pmatrix}
\end{aligned}$$

답

$$x_1 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \quad x_2 = \begin{pmatrix} \frac{5-\sqrt{5}}{10} \\ \frac{5-\sqrt{5}}{10} \end{pmatrix}$$

문제 9. RMSProp Algorithm

9. (RMSProp algorithm) 이변수 함수 $f(x, y) = x^2 + xy$ 를 RMSProp 방법으로 최적화하고자 합니다. 초기 위치 $x_0 = (1, 1)$ 에서 출발하여 2 step 진행할 때, x_1, x_2 를 구하세요. (단 learning rate $\eta = \frac{1}{2}$, forgetting factor $\gamma = \frac{8}{9}$)

주어진 정보

- 목적 함수: $f(x, y) = x^2 + xy$
- 초기 위치: $x_0 = (1, 1)$
- Learning rate: $\eta = \frac{1}{2}$
- Forgetting factor: $\gamma = \frac{8}{9}$
- 구하는 것: 2 step 진행 후 x_1, x_2

풀이 과정

Step 1: RMSProp 알고리즘 공식

$$h_{t+1} = \gamma h_t + (1 - \gamma)(\nabla f(x_t))^2$$

$$x_{t+1} = x_t - \eta \frac{\nabla f(x_t)}{\sqrt{h_{t+1}} + \epsilon}$$

여기서:

- $h_0 = (0, 0)$ (초기값)
- $\gamma = \frac{8}{9}$ (forgetting factor, decay rate)
- ϵ 은 매우 작은 값 (여기서는 0으로 가정)

Step 2: 그래디언트 함수

$$\nabla f(x, y) = \begin{pmatrix} 2x + y \\ x \end{pmatrix}$$

Step 3: Step 1 계산 ($x_0 = (1, 1)$)

그래디언트:

$$\nabla f(1, 1) = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

이동 평균 제곱 그래디언트 업데이트:

$$\begin{aligned} h_1 &= \gamma h_0 + (1 - \gamma)(\nabla f(x_0))^2 \\ &= \frac{8}{9} \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \left(1 - \frac{8}{9}\right) \begin{pmatrix} 9 \\ 1 \end{pmatrix} \\ &= \frac{1}{9} \begin{pmatrix} 9 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{1}{9} \end{pmatrix} \end{aligned}$$

위치 업데이트:

$$\begin{aligned} x_1 &= x_0 - \eta \frac{\nabla f(x_0)}{\sqrt{h_1}} \\ &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{2} \cdot \frac{\begin{pmatrix} 3 \\ 1 \end{pmatrix}}{\begin{pmatrix} \sqrt{1} \\ \sqrt{\frac{1}{9}} \end{pmatrix}} \\ &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{2} \cdot \begin{pmatrix} \frac{3}{1} \\ \frac{1}{\frac{1}{3}} \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{2} \cdot \begin{pmatrix} 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 - 1.5 \\ 1 - 1.5 \end{pmatrix} = \begin{pmatrix} -0.5 \\ -0.5 \end{pmatrix} \end{aligned}$$

Step 4: Step 2 계산 ($x_1 = (-0.5, -0.5)$)

그래디언트:

$$\nabla f(-0.5, -0.5) = \begin{pmatrix} 2(-0.5) + (-0.5) \\ -0.5 \end{pmatrix} = \begin{pmatrix} -1.5 \\ -0.5 \end{pmatrix}$$

이동 평균 제곱 그래디언트 업데이트:

$$\begin{aligned} h_2 &= \gamma h_1 + (1 - \gamma)(\nabla f(x_1))^2 \\ &= \frac{8}{9} \begin{pmatrix} 1 \\ \frac{1}{9} \end{pmatrix} + \frac{1}{9} \begin{pmatrix} 2.25 \\ 0.25 \end{pmatrix} \\ &= \begin{pmatrix} \frac{8}{9} + \frac{2.25}{9} \\ \frac{8}{81} + \frac{0.25}{9} \end{pmatrix} = \begin{pmatrix} \frac{8+2.25}{9} \\ \frac{8}{81} + \frac{0.25}{9} \end{pmatrix} \\ &= \begin{pmatrix} \frac{10.25}{9} \\ \frac{8+2.25}{81} \end{pmatrix} = \begin{pmatrix} \frac{41}{36} \\ \frac{41}{324} \end{pmatrix} \end{aligned}$$

위치 업데이트:

$$\begin{aligned} x_2 &= x_1 - \eta \frac{\nabla f(x_1)}{\sqrt{h_2}} \\ &= \begin{pmatrix} -0.5 \\ -0.5 \end{pmatrix} - \frac{1}{2} \cdot \frac{\begin{pmatrix} -1.5 \\ -0.5 \end{pmatrix}}{\begin{pmatrix} \sqrt{\frac{41}{36}} \\ \sqrt{\frac{41}{324}} \end{pmatrix}} \end{aligned}$$

$$\begin{aligned}
&= \begin{pmatrix} -0.5 \\ -0.5 \end{pmatrix} - \frac{1}{2} \cdot \begin{pmatrix} \frac{-1.5}{\frac{\sqrt{41}}{6}} \\ \frac{-0.5}{\frac{\sqrt{41}}{18}} \end{pmatrix} \\
&= \begin{pmatrix} -0.5 \\ -0.5 \end{pmatrix} - \frac{1}{2} \cdot \begin{pmatrix} \frac{-9}{\sqrt{41}} \\ \frac{-9}{\sqrt{41}} \end{pmatrix} \\
&= \begin{pmatrix} -0.5 \\ -0.5 \end{pmatrix} + \frac{9}{2\sqrt{41}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} -0.5 + \frac{9}{2\sqrt{41}} \\ -0.5 + \frac{9}{2\sqrt{41}} \end{pmatrix} = \begin{pmatrix} \frac{9\sqrt{41}-\sqrt{41}}{2\sqrt{41}} \\ \frac{9\sqrt{41}-\sqrt{41}}{2\sqrt{41}} \end{pmatrix} \\
&= \begin{pmatrix} \frac{9-\sqrt{41}}{2\sqrt{41}} \cdot \sqrt{41} \\ \frac{9-\sqrt{41}}{2\sqrt{41}} \cdot \sqrt{41} \end{pmatrix}
\end{aligned}$$

더 간단하게:

$$\begin{aligned}
&= \begin{pmatrix} -\frac{1}{2} + \frac{9}{2\sqrt{41}} \\ -\frac{1}{2} + \frac{9}{2\sqrt{41}} \end{pmatrix} = \begin{pmatrix} \frac{9-\sqrt{41}}{2\sqrt{41}} \\ \frac{9-\sqrt{41}}{2\sqrt{41}} \end{pmatrix} \cdot \frac{\sqrt{41}}{\sqrt{41}} \\
&= \begin{pmatrix} \frac{9\sqrt{41}-41}{82} \\ \frac{9\sqrt{41}-41}{82} \end{pmatrix}
\end{aligned}$$

답

$$x_1 = \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}, \quad x_2 = \begin{pmatrix} \frac{9\sqrt{41}-41}{82} \\ \frac{9\sqrt{41}-41}{82} \end{pmatrix}$$

문제 10. Adam Algorithm

10. (Adam algorithm) 이변수 함수 $f(x, y) = x^2 + xy$ 를 Adam 방법으로 최적화하고자 합니다. 초기 위치 $x_0 = (1, 1)$ 에서 출발하여 2 step 진행할 때, x_1, x_2 를 구하세요. (단 learning rate $\eta = 1, \beta_1 = \beta_2 = \frac{1}{2}$)

주어진 정보

- 목적 함수: $f(x, y) = x^2 + xy$
- 초기 위치: $x_0 = (1, 1)$
- Learning rate: $\eta = 1$
- $\beta_1 = \beta_2 = \frac{1}{2}$
- 구하는 것: 2 step 진행 후 x_1, x_2

풀이 과정

Step 1: Adam 알고리즘 공식

Adam은 momentum과 RMSProp를 결합한 알고리즘입니다:

$$\begin{aligned}
m_{t+1} &= \beta_1 m_t + (1 - \beta_1) \nabla f(x_t) \\
v_{t+1} &= \beta_2 v_t + (1 - \beta_2) (\nabla f(x_t))^2
\end{aligned}$$

$$\hat{m}_{t+1} = \frac{m_{t+1}}{1 - \beta_1^{t+1}}$$

$$\hat{v}_{t+1} = \frac{v_{t+1}}{1 - \beta_2^{t+1}}$$

$$x_{t+1} = x_t - \eta \frac{\hat{m}_{t+1}}{\sqrt{\hat{v}_{t+1}} + \epsilon}$$

초기값:

- $m_0 = (0, 0)$ (1차 모멘트)
- $v_0 = (0, 0)$ (2차 모멘트)
- $\epsilon \approx 0$ (여기서는 무시)

Step 2: 그래디언트 함수

$$\nabla f(x, y) = \begin{pmatrix} 2x + y \\ x \end{pmatrix}$$

Step 3: Step 1 계산 ($x_0 = (1, 1)$, $t = 0$)

그래디언트:

$$\nabla f(1, 1) = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

1차 모멘트 업데이트:

$$m_1 = \beta_1 m_0 + (1 - \beta_1) \nabla f(x_0) = \frac{1}{2} \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 0.5 \end{pmatrix}$$

2차 모멘트 업데이트:

$$v_1 = \beta_2 v_0 + (1 - \beta_2) (\nabla f(x_0))^2 = \frac{1}{2} \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 9 \\ 1 \end{pmatrix} = \begin{pmatrix} 4.5 \\ 0.5 \end{pmatrix}$$

Bias correction ($t=1$):

$$\hat{m}_1 = \frac{m_1}{1 - \beta_1^1} = \frac{\begin{pmatrix} 1.5 \\ 0.5 \end{pmatrix}}{1 - \frac{1}{2}} = \frac{\begin{pmatrix} 1.5 \\ 0.5 \end{pmatrix}}{\frac{1}{2}} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

$$\hat{v}_1 = \frac{v_1}{1 - \beta_2^1} = \frac{\begin{pmatrix} 4.5 \\ 0.5 \end{pmatrix}}{1 - \frac{1}{2}} = \frac{\begin{pmatrix} 4.5 \\ 0.5 \end{pmatrix}}{\frac{1}{2}} = \begin{pmatrix} 9 \\ 1 \end{pmatrix}$$

위치 업데이트:

$$x_1 = x_0 - \eta \frac{\hat{m}_1}{\sqrt{\hat{v}_1}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 1 \cdot \frac{\begin{pmatrix} 3 \\ 1 \end{pmatrix}}{\sqrt{\begin{pmatrix} 9 \\ 1 \end{pmatrix}}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Step 4: Step 2 계산 ($x_1 = (0, 0)$, $t = 1$)

그래디언트:

$$\nabla f(0, 0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

1차 모멘트 업데이트:

$$m_2 = \beta_1 m_1 + (1 - \beta_1) \nabla f(x_1) = \frac{1}{2} \begin{pmatrix} 1.5 \\ 0.5 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}$$

2차 모멘트 업데이트:

$$v_2 = \beta_2 v_1 + (1 - \beta_2)(\nabla f(x_1))^2 = \frac{1}{2} \begin{pmatrix} 4.5 \\ 0.5 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2.25 \\ 0.25 \end{pmatrix}$$

Bias correction (t=2):

$$\hat{m}_2 = \frac{m_2}{1 - \beta_1^2} = \frac{\begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}}{1 - \frac{1}{4}} = \frac{\begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}}{\frac{3}{4}} = \begin{pmatrix} 1 \\ \frac{1}{3} \end{pmatrix}$$

$$\hat{v}_2 = \frac{v_2}{1 - \beta_2^2} = \frac{\begin{pmatrix} 2.25 \\ 0.25 \end{pmatrix}}{1 - \frac{1}{4}} = \frac{\begin{pmatrix} 2.25 \\ 0.25 \end{pmatrix}}{\frac{3}{4}} = \begin{pmatrix} 3 \\ \frac{1}{3} \end{pmatrix}$$

위치 업데이트:

$$\begin{aligned} x_2 &= x_1 - \eta \frac{\hat{m}_2}{\sqrt{\hat{v}_2}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - 1 \cdot \frac{\begin{pmatrix} \frac{1}{3} \\ \frac{1}{\sqrt{3}} \end{pmatrix}}{\begin{pmatrix} \sqrt{3} \\ \frac{1}{\sqrt{3}} \end{pmatrix}} \\ &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{3} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{\sqrt{3}}{3} \end{pmatrix} \\ &= \begin{pmatrix} -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{pmatrix} = \begin{pmatrix} -\frac{\sqrt{3}}{3} \\ -\frac{\sqrt{3}}{3} \end{pmatrix} \end{aligned}$$

답

$$x_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad x_2 = \begin{pmatrix} -\frac{\sqrt{3}}{3} \\ -\frac{\sqrt{3}}{3} \end{pmatrix}$$
