Optimization Algorithms for AIO2024

Mai Hanh (ID: AIO_126)

Introduction

This document explains step-by-step applications of Gradient Descent, Gradient Descent + Momentum, RMSProp, and Adam optimizers using the example function:

$$f(w_1, w_2) = w_1^2 + w_2^2$$

We manually compute and update weights w_1 and w_2 over 2 epochs for each optimizer. These optimization techniques are essential for minimizing cost functions in machine learning and are part of Module 5 in the AIO2024 course.

1. Gradient Descent

The gradients for the function are:

$$\frac{\partial f}{\partial w_1} = 2w_1, \quad \frac{\partial f}{\partial w_2} = 2w_2$$

The weight update rule is:

$$w_1 = w_1 - \alpha \frac{\partial f}{\partial w_1}, \quad w_2 = w_2 - \alpha \frac{\partial f}{\partial w_2}$$

Epoch 1:

$$\frac{\partial f}{\partial w_1} = -10, \quad \frac{\partial f}{\partial w_2} = -4$$

$$w_1 = -5 + 0.4 \times 10 = -1, \quad w_2 = -2 + 0.4 \times 4 = -0.4$$

Epoch 2:

$$\frac{\partial f}{\partial w_1} = -2, \quad \frac{\partial f}{\partial w_2} = -0.8$$

$$w_1 = -1 + 0.4 \times 2 = -0.2, \quad w_2 = -0.4 + 0.4 \times 0.8 = -0.08$$

2. Gradient Descent + Momentum

The velocity term is introduced:

$$v_t = \beta v_{t-1} + (1-\beta)\frac{\partial f}{\partial w_t}$$

Weights are updated as:

$$w_t = w_{t-1} - \alpha v_t$$

Epoch 1:

$$v_1 = 0.5 \times 0 + 0.5 \times (-10) = -5, \quad v_2 = 0.5 \times 0 + 0.5 \times (-4) = -2$$

$$w_1 = -5 + 0.4 \times 5 = -3, \quad w_2 = -2 + 0.4 \times 2 = -1.2$$

Epoch 2:

$$v_1 = 0.5 \times (-5) + 0.5 \times (-6) = -5.5, \quad v_2 = 0.5 \times (-2) + 0.5 \times (-2.4) = -2.2$$

$$w_1 = -3 + 0.4 \times 5.5 = -0.8, \quad w_2 = -1.2 + 0.4 \times 2.2 = -0.32$$

3. RMSProp

RMSProp uses a running average of squared gradients:

$$s_t = \gamma s_{t-1} + (1-\gamma)\left(\frac{\partial f}{\partial w_t}\right)^2$$

Weights are updated as:

$$w_t = w_{t-1} - \frac{\alpha}{\sqrt{s_t} + \epsilon}\frac{\partial f}{\partial w_t}$$

Epoch 1:

$$s_1 = 0.9 \times 0 + 0.1 \times (-10)^2 = 10, \quad s_2 = 0.9 \times 0 + 0.1 \times (-4)^2 = 1.6$$

$$w_1 = -5 - \frac{0.4}{\sqrt{10} + 1e - 6} \times (-10), \quad w_2 = -2 - \frac{0.4}{\sqrt{1.6} + 1e - 6} \times (-4)$$

$$w_1 = -4.4, \quad w_2 = -1.76$$

Epoch 2:

$$s_1 = 0.9 \times 10 + 0.1 \times (-4.4)^2 = 13.136, \quad s_2 = 0.9 \times 1.6 + 0.1 \times (-1.76)^2 = 2.1344$$

$$w_1 = -4.4 - \frac{0.4}{\sqrt{13.136} + 1e - 6} \times (-8.8), \quad w_2 = -1.76 - \frac{0.4}{\sqrt{2.1344} + 1e - 6} \times (-3.52)$$

4. Adam Optimizer

Adam combines Momentum and RMSProp:

$$v_t = \beta_1 v_{t-1} + (1 - \beta_1) \frac{\partial f}{\partial w_t}, \quad s_t = \beta_2 s_{t-1} + (1 - \beta_2) \left(\frac{\partial f}{\partial w_t} \right)^2$$

Bias correction is applied:

$$\hat{v}_t = \frac{v_t}{1 - \beta_1^t}, \quad \hat{s}_t = \frac{s_t}{1 - \beta_2^t}$$

Weights are updated as:

$$w_t = w_{t-1} - \frac{\alpha}{\sqrt{\hat{s}_t} + \epsilon} \hat{v}_t$$

Epoch 1:

$$v_t = 0.9 \times 0 + 0.1 \times (-10) = -1, \quad s_t = 0.999 \times 0 + 0.001 \times (-10)^2 = 0.1$$

$$\hat{v}_t = -10, \quad \hat{s}_t = 100$$

$$w_1 = -5 - \frac{0.4}{\sqrt{100} + 1e - 6} \times (-10)$$

Acknowledgment

This document is prepared by Mai Hanh (ID: AIO_126) for the AIO2024 course and reflects the practical application of optimizers in artificial intelligence. Please do not engage in any unauthorized activities such as copying, reproducing, or distributing this document without explicit permission or agreement from the author, Mai Hanh.