Momentum Method

Accelerating Gradient Descent with Velocity Accumulation

🚀 Core Concept

Accelerates gradient descent by accumulating velocity from past gradients

Update Rules

1. Velocity Update

$$v = \beta v + \nabla L(\theta)$$

2. Parameter Update

$$\theta = \theta - \eta v$$

Momentum Coefficient β: Typically 0.9
(Retains 90% of previous velocity)

✨ Key Benefits

1. Helps overcome local minima
2. Navigates ravines in loss landscape
3. Reduces oscillations in high curvature directions
4. Accelerates progress in consistent gradient directions
5. Dampens oscillations in narrow valleys

Physical Analogy

Similar to a ball rolling down a hill gaining momentum

Effect Visualization


Without Momentum
Zigzag movement
With Momentum
Smooth progress