Gradient Descent Review

Foundation for Modern Neural Network Training

🎯 Core Concept

Iterative optimization algorithm to minimize loss function

Updates parameters in negative gradient direction

Update Rule

$$\theta = \theta - \eta \nabla L(\theta)$$

η: Learning Rate

How It Works

1. Calculate loss gradient for all parameters
2. Move parameters in opposite direction of gradient
3. Repeat until convergence

📊 Key Characteristics

Foundation of all modern neural network training algorithms

Convergence depends on learning rate, loss landscape, and initialization

⚠️ Challenges

- Can get trapped in local minima
- Saddle point problems

⚖️ Trade-off

Convergence speed ↔ Stability