Nesterov Accelerated Gradient (NAG)

Improved Momentum with Look-Ahead Gradient Calculation

🔍 Key Innovation

"Look-ahead" mechanism
Calculates gradient at anticipated future position

Update Rules

1. Velocity Update (Look-ahead)

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

2. Parameter Update

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

⚡ **Key Point:** Calculates gradient at $\theta - \beta v$ position for more accurate direction prediction

vs Standard Momentum

Momentum
Calculates gradient at current position

Nesterov
Calculates gradient at future position

✨ Advantages

- ✅ Better convergence characteristics than standard momentum
- ✅ Reduces overshooting through early trajectory correction
- ✅ Particularly effective for convex optimization problems

⚖️ **Trade-off:** Slightly higher computational cost than standard momentum

💡 **Widely Used**

Widely used in major deep learning frameworks like PyTorch and TensorFlow