Cosine Annealing

Smooth Learning Rate Scheduling with Cosine Function

🔍 Core Concept

Learning rate follows a cosine function from maximum to minimum value

Formula

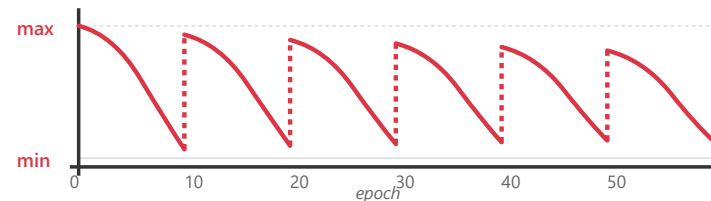$$LR = min\_lr + (max\_lr - min\_lr) \times (1 + \cos(\pi t/T)) / 2$$

t: current epoch, T: total epochs

Decay Behavior

Early Phase
Fast reduction

Late Phase
Slow reduction (Fine-tuning)

Cosine Curve Visualization

✨ Key Advantages

- ✅ Smooth and gradual decay
- ✅ No hyperparameters to tune (only max/min LR)
- ✅ Improves generalization (empirically proven)

🔄 Warm Restarts (SGDR)

Can include restarts to escape local minima

🎯 Popular In

Computer Vision   Transformer Training