

# Importance of Learning Rate

The Most Critical Hyperparameter



The most important hyperparameter affecting training dynamics and convergence

## Learning Rate Impact

### ⚠ Too Large

- Training divergence
- Loss oscillation
- Loss explosion

### ⚠ Too Small

- Slow convergence
- Stuck in poor local minima

### ✓ Optimal (Dynamic)

Optimal learning rate changes throughout training

## Training Phase Strategy



### Early Training

Large learning rate for fast initial progress



### Late Training

Small learning rate for fine-tuning and convergence

## 💡 Key Insights

- 1 Different layers may benefit from different learning rates
- 2 Learning rate scheduling improves final model performance