

Parameter Selection Strategies

Learning Rate Scheduling & Layer-wise Optimization

Critical hyperparameters for successful fine-tuning

Learning Rate Guidelines

- **PEFT:** $1e-3$ to $5e-3$ (higher than full fine-tuning)
- **Full fine-tuning:** $1e-5$ to $5e-5$
- **Warmup:** 5-10% of total steps
- **Decay:** Linear or cosine to 10% of peak

Layer-wise Learning Rates

- Lower layers: Smaller LR (preserve general features)
- Upper layers: Higher LR (adapt to domain)
- Ratio: 1:10 (bottom:top layers)
- Gradual unfreezing strategy

Regularization Techniques

- Weight decay: 0.01-0.1
- Dropout: 0.1-0.3 in adapters
- Early stopping: patience 3-5 epochs

- Gradient clipping: max norm 1.0