Comparison of Optimization Algorithms

Performance Characteristics and Trade-offs

Metric	SGD	Momentum	RMSprop	Adam/AdamW
Training Speed	Slow	Moderate	Fast	**Fastest**
Generalization	Good	**Best**	Moderate	Good
Memory Usage	**Lowest**	Low	Moderate	High
HP Sensitivity	High	Moderate	Low	**Lowest**

SGD + Momentum

Simple, reliable
Good generalization
Requires tuning

Adam/AdamW

Adaptive
Fast convergence
Out-of-the-box

RMSprop

Good for RNNs
Non-stationary problems

Speed Ranking

Adam > RMSprop > Momentum > SGD

Recommendation: Start with Adam/AdamW, then fine-tune with SGD+Momentum

Practice Materials

Gradient Descent Visualization (GitHub)