

Adam and AdamW

Most Popular and Improved Optimizers

★ Adam

Most popular optimization algorithm combining Momentum and RMSprop

Moment Estimates

First Moment

Mean

Second Moment

Variance

🎛 Default Parameters

$\beta_1 = 0.9$ (Momentum)

$\beta_2 = 0.999$ (RMSprop)

$\epsilon = 1e-8$ (Numerical stability)

🔧 AdamW

Improved Adam using decoupled weight decay regularization

✨ Key Improvements

- 1 Fixes weight decay implementation bug in original Adam
- 2 Decouples L2 regularization from gradient-based optimization
- 3 Generally better generalization than Adam

Key Difference

Adam

Weight decay in gradient

AdamW

Weight decay decoupled

Includes **Bias Correction**

for moment estimates in early iterations



AdamW is the preferred choice
in modern deep learning