

RMSprop (Root Mean Square Propagation)

Adaptive Learning Rate with Exponential Moving Average

Key Improvement

Solves AdaGrad's monotonically decreasing learning rate problem

Innovation

Uses exponential moving average of squared gradients

Update Rules

1. EMA of Squared Gradients

$$E[g^2] = \rho E[g^2] + (1-\rho)g^2$$

2. Parameter Update

$$\theta = \theta - \eta / \sqrt{E[g^2] + \epsilon} \odot g$$

vs AdaGrad

AdaGrad

Monotonically decreasing LR

RMSprop

LR can dynamically increase/decrease

Advantages

- ✓ Learning rate can dynamically increase/decrease
- ✓ Effective for non-stationary objectives
- ✓ Well-suited for Recurrent Neural Networks

Origin

Introduced in Geoff Hinton's Coursera lectures
(Unpublished paper)

 Good default choice for

Decay Rate ρ : Typically 0.9

Maintains moving average of recent gradients

various optimization problems