

RMSProp

RMSProp, or Root Mean Squared Prop, is an adaptive learning rate optimization algorithm commonly used in deep learning. It addresses some shortcomings of standard gradient descent and aims to achieve faster convergence.

Steps

1. **Initialize Parameters:** Begin with initial values for the parameters that you want to optimize.
2. **Set Hyperparameters:** Choose a learning rate and decay rate, typically around 0.9.
3. **Initialize Accumulated Gradient:** Set an accumulated gradient variable (G) to zero. This will store the sum of the squares of the gradients.
4. **Compute Gradient:** For each iteration, calculate the gradient (g_t) of the loss function with respect to the parameters.
5. **Update Accumulated Gradient:** Update (G) by adding the square of the current gradient, discounted by the decay rate:

$$G(t) = \gamma G(t - 1) + (1 - \gamma)g^2(t)$$

6. **Adjust Learning Rate:** Modify the learning rate for each parameter based on the accumulated gradient:

$$\lambda(t) = \lambda / \sqrt{G(t) + \epsilon}$$

where (ϵ) is a small constant added to improve numerical stability.

7. **Update Parameters:** Adjust the parameters in the opposite direction of the gradient, scaled by the adjusted learning rate:

$$\theta(t + 1) = \theta(t) - \lambda(t) \cdot g(t)$$

9. **Repeat:** Continue the process for a specified number of iterations or until convergence.

Advantages of RMSProp:

- Faster convergence compared to standard gradient descent, especially for problems with non-convex loss functions.
- Adapts learning rates to individual parameters, addressing issues with features having vastly different scales.