DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
(ARTIFICIAL INTELLIGENCE & MACHINE LEARNING)

RMSProp

RMSProp, which stands for Root Mean Square Propagation, is an adaptive learning rate optimization algorithm designed to address some of the issues encountered with the stochastic gradient descent (SGD) method in training deep neural networks. The RMSProp algorithm was introduced by Geoffrey Hinton in his Coursera class on neural networks and is not formally published, but it has gained popularity due to its effectiveness in various applications.

The RMSProp update adjusts the Adagrad method to reduce its aggressive, monotonically decreasing learning rate. Instead of accumulating all past squared gradients, RMSProp uses an exponential decay that discards history from the extreme past so that it can converge rapidly after finding a convex bowl, as if it were an Adagrad with a fresh start.

RMSProp addresses the issue of a global learning rate by maintaining a moving average of the squares of gradients for each weight and dividing the learning rate by this average. This ensures that the learning rate is adapted for each weight in the model, allowing for more nuanced updates. The general idea is to dampen the oscillations in directions with steep gradients while allowing for faster movement in flat regions of the loss landscape.

$$V_t = \beta V_{t-1} + (1-\beta)(\nabla w_t)^2$$

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{V_t + \varepsilon}} * \nabla w_t$$

- Recent gradients have higher impact as compared to previous one.
- Automatically V_t term won't get that much big as it was becoming in AdaGrad.

Advantages of RMSProp:

- Adaptive Learning Rates: By adjusting the learning rate for each parameter, RMSProp can handle different scales of data and varying curvatures of loss functions.
- Convergence Speed: RMSProp can converge faster than SGD with momentum, especially in scenarios with noisy or sparse gradients.
- Stability: The method avoids the diminishing learning rates found in Adagrad, which can stall the training process in the later stages.