

Root mean square propagation

Root mean square propagation is another method (like momentum) to improve the speed of gradient descent.

The RMS propagation optimizer is similar to the gradient descent algorithm with momentum. The RMS propagation optimizer restricts the oscillations in the vertical direction. Therefore, we can increase our learning rate and our algorithm could take larger steps in the horizontal direction converging faster. The difference between RMS propagation and gradient descent is on how the gradients are calculated. The following equations show how the gradients are calculated for the RMS propagation and gradient descent with momentum. The value of momentum is denoted by beta and is usually set to 0.9.

$$s(0) = 0$$

At iteration t

$$s(t) := \beta s(t) + (1 - \beta) \frac{\delta C^2}{\delta w}$$

the square indicates element wise square

$$w := w - \alpha \frac{\frac{\delta C}{\delta w}}{\sqrt{s(t) + \epsilon}}$$

here we add ϵ because s approaches zero at times and ϵ is taken to be around 10^{-8} . α is learning rate.