Gradient Descent Variations

**Gradient Descent**

It is an iterative optimization algorithm, used to find the minimum value for a function. The general idea is to initialize the parameters to random values, and then take small steps in the direction of the “slope” at each iteration. Gradient descent is highly used in supervised learning to minimize the error function and find the optimal values for the parameters.

Numerous extensions have been designed for gradient descent algorithm. Some of them are discussed below:

**Momentum method**

This method is used to accelerate the gradient descent algorithm by taking into consideration the exponentially weighted average of the gradients. Using averages makes the algorithm converge towards the minima in a faster way, as the gradients towards the uncommon directions are canceled out. The pseudocode for momentum method is given below.

**RMSprop**

RMSprop was proposed by University of Toronto’s Geoffrey Hinton. The intuition is to apply an exponentially weighted average method to the second moment of the gradients (dW2).

**Adam Optimization**

Adam optimization algorithm incorporates the momentum method and RMSprop, along with bias correction.