

Abstract

Neural network has become a core part of machine learning in recent years and conventional neural network although successful, gets bottle-necked due to slow technology and lack of optimization potential in the common methods used to train the neural networks.

Some common methods used include, SGD(stochastic gradient descent), ADAM and ADA-GRAD etc. These all algorithms are based on Gradient Descent, a first order root-finding algorithm. These are great for small scale neural network computation, but at industry level, where millions of data-points are produced, the neural networks required for the learning from the data-points either require large number of nodes or lots of layers, so it lacks the performance. The problem with Gradient Descent is that it becomes immensely slow as layers or nodes in the neural network increases, as well as the lack of optimal-direction finding potential on pathological curves makes it even harder to train networks using Gradient descent based algorithms.

2nd order optimization method, Newton's method, has been known to converge to the root faster than Gradient Descent and since it is a 2nd order algorithm, it has the curvature data, so we can modify the Newton's method to compensate for the problems that Gradient Descent faces. The thesis research deals with one such modification, which in optimization terminology is called Hessian-free approach. Further in this document, we will suggest ways to modify the Newton's method. The modified Hessian-free Newton's method will deal with the problem of optimization of cost function of the neural network efficiently.