

Comparison of first order optimization algorithms for neural networks

Xristos Krallis, Nikolaos Ploskas

In this day and age more and more applications tend to be based in some form of machine learning algorithms. One of the most widely used machine learning algorithms is neural networks. Even though neural networks tend to become more and more common, some problems still preserve. For example, the training of a neural network is a cumbersome optimization task that still requires fine tuning. The aim of this work is to compare various optimization methods used in neural network training. The algorithms that will be included in the computational comparison to train these neural networks are mainly first order optimization algorithms; these are Stochastic Gradient Descent, Adagrad, Adadelata, RMSprop, Adam, Adamax, Nadam, and FTRL. The experimental process will be composed of typical neural network structures like convolutional and recurrent ones. The experiments will be also performed for shallow and wide, deep and narrow as well as balanced neural networks. Moreover, another parameter that we will take into consideration is the density of the neural networks, meaning that the experiments will be performed for both dense and sparse neural networks. The optimization methods will be compared according to their execution time required to train the neural network at an accuracy threshold, the accuracy of the model after a certain number of epochs, and finally, the generalization of each model.