

MSc Project Report

Optimizing learning algorithm of deep neural networks which use
stochastic gradient descent as the optimization technique

Sudheera Withanage

2022

A report submitted as part of the requirement for the degree of
MSc Big Data Analytics at Robert Gordon University, Scotland

Abstract

Neural networks now have become the backbone of most of the businesses and services all around the globe. Training and tuning network parameters until it gives its outputs as accurately as possible is an integral part of developing a well-known product among customers. However, the main drawbacks with using neural networks are the time that takes to train and their generalization ability, i.e., their performance when previously unseen data are presented as inputs. Research shows that minima that network converges shows its generalization ability. Flat and wide minima is responsible for more generalized neural network. Understanding the hyperparameters that contribute mostly to find flat and wide minima are learning rate and batch size. More about studies related to this topic are mentioned in the literature review section. Therefore, the effect of initiation and changing those hyperparameters when the training going on, is unquestionable.

However, most researchers and developers use arbitrary initiations and fluctuations of those parameters while training. This research combines two logical methods that exist for those tasks to obtain better results within a lesser number of epochs. It consists of a training run to find out the maximum learning rate that can be used and minimum is can be set as any value greater than zero. It is required the developer to set a set of parameters as, maximum batch size, minimum batch size and batch size to learning rate ratio. Within the training process the learning rate is fluctuated within the upper and lower bounds of the learning rate and batch size is fluctuated according to the parameters mentioned above. Using this suggested training algorithm, ResNet 20 network was able to obtain 91.22% validation accuracy on Cifar 10 dataset in the expense of 90 epochs.