Loss Values in Neural Networks

26.03.2021 – Ward Ali Dib

Image classification is the task of assigning a label to an image, seeking to mimic how the human brain recognises an object. The leading algorithms for image classification are convolutional neural networks (CNNs), which utilise various functions and layers to predict the correct outcome. (1)

This report will use the MNIST dataset to explore loss curves, and the effect of parameters on a neural network.

The loss function in a neural network is used to optimize the parameter values, and map them onto a scalar value to indicate how well those parameters accomplish the required task.

First, we will limit the sample size to 3000 images. Then we will use the negative log likelihood loss criterion. It is useful to train a multi-class learning problems where a set of features can be related to one-of-K-classes. (2)

Our first network will use 20 repeat runs with 50 iterations, and compute the mean as a function of iteration numbers. Plotting the mean losses against iterations, we get the following loss curve.

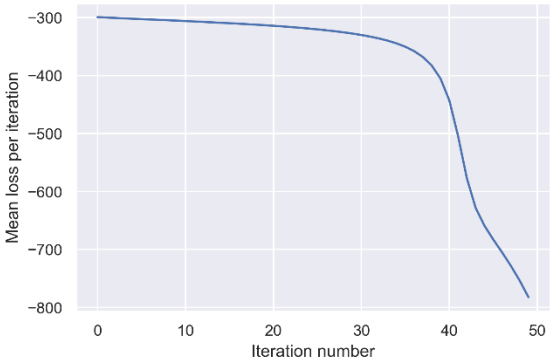

Figure 1 – loss curve of first neural network

We can see losses are still decreasing and did not reach a plateau. Loss is a number indicating how bad the model's prediction was on an example. The goal of training a model is to find a set of weights and biases that have low loss. Thus, we can still improve the network to reach the best results.

There are many factors that we can take into account – number of iterations, learning rate, momentum, sample size, batch number, etc.

Training the model with 100 iterations didn't give a plateau, however, it seems to be an improvement.

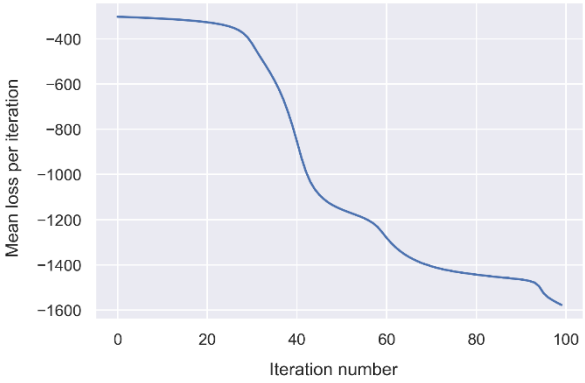

Figure 2 – loss curve with more iterations

Increasing the iterations to 1000+ might work, but that is time consuming and less efficient. Instead, we will try adjusting the optimiser parameters by increasing the learning rate to 1e-3 and decreasing the momentum to 0.4. A learning rate that is too small can cause the algorithm to get stuck. (3)

The new results look much better.

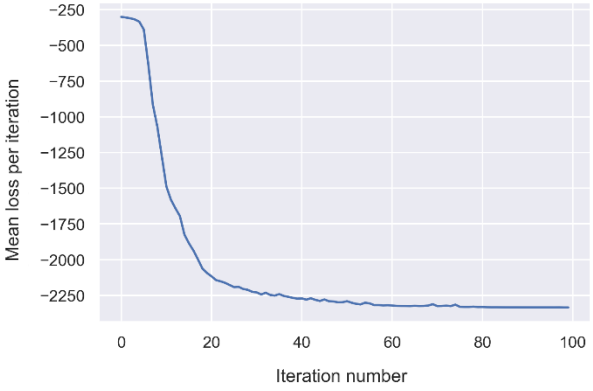

Figure 3 – loss curve with improved optimiser parameters

Thus, the recommendation for this network is changing the learning rate and momentum.

1. Saha S. A Comprehensive Guide to CNNs [Internet]. Towards Data Science. Towards Data Science; 2018. Available from: shorturl.at/ewWZ6
2. MIRANDA L. Understanding softmax and the negative log-likelihood [Internet]. Lj Miranda. 2017. Available from: shorturl.at/mrDP1
3. Brownlee J. Understand the Impact of Learning Rate on Neural Network Performance [Internet]. Available from: shorturl.at/vCFO7