

# MLP Coursework 4

Eskil Joergensen

March 18, 2017

## 1 Introduction

Optimizing the training performance of a convolutional neural network (CNN) on image classification task, CIFAR 10. Building on a baseline set of experiments, done with a deep neural network, the presented () tries to improve the performance of the more advanced methods used here.

In addition to improving the overall performance of the training of the network, the experiments were guided by a goal of 70% accuracy on the test data set. As a result, the methodology and incremental improvements to the model were made on these grounds.

## 2 Methodology

Establishing a baseline model, and choosing the appropriate algorithms to include, can be a daunting task. Even a relatively small CNN has a large number of tunable parameters, and it can be difficult to know where to begin. The section presents the baseline model chose for the experiments, the procedure used to train the network and the parameters used.

### 2.1 The Baseline Model

The default baseline model used has two convolutional layers, two fully-connected layers and finally a softmax classification layer. The network also had two 2x2 pooling layers after each convolutional layer.

Instead of a learning rate scheduler, the Adam adaptive learning rate optimizer was used throughout the investigation. The optimizer was used with varying learning rates, but with  $\beta_1=0.9$ ,  $\beta_2=0.999$  and  $\epsilon = 1e-8$ . (default parameters)

## 2.2 Experimental Procedure

The procedure used to tune the performance of the network involved three stages. During the initial stage the network was trained for few epochs, with a coarse spread of hyper parameter and multiple algorithms were tested during this phase. The overall results from the initial trials informed the finer range of parameters used for additional training and further tuning of hyper parameters. The second stage involved longer training. The third stage, before the final testing, the model was trained for significantly longer, and only three varieties of the model were used.

### 2.2.1 Stage 1: Initial Training and Model Verification

Initially the model was trained using three learning rates, with the four activation functions used in the preliminary experiments: ReLu, Elu, Tahn and Sigmoid. These were tested with a learning rate of 0.0005, 0.005 and 0.1. The relatively large, non-uniform spread of the learning rate was chosen intentionally to observe how the activation functions would react to the rates.

After the initial trials, the best activation function and learning rate was chosen for the rest of the trials in Stage 1.

Next a variety of combinations of kernel sizes and feature maps were tested. The size of kernels were 3, 5 and 7, whereas the number of feature maps were 4, 24 and 32. Here as well, the best combination was chosen for the following trials.

with and without batch normalization and l2 loss

trained for 10 epochs

### 2.2.2 Stage 2: Fine-tuning hyper parameters

changed max pooling <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>

replaced batch norm with lrn

trained for 40 epochs

### 2.2.3 Stage 2: Finding the one and Testing

trained for 100 epochs

**3 Results and Discussion**

**4 Conclusion**

**5 Future Work**

## References