CS 466/566 Assignment 2 Report

SEDA NUR DOĞANAY S004299

Saved trainings of networks : WETRANSFER LINK

Network Designs

Network Layout	
Network 1	***Network 2***
conv 5x5x3x32 with s-1x1x1x1	*conv 5x5x3x32 with s-1x1x1x1*
ReLu - Norm	*ReLu - Norm*
maxpool 1x3x3 with s-1x2x2x1	*maxpool 1x4x4 with s-1x2x2x1*
conv 5x5x32x16	*conv 5x5x32x32*
ReLu - Norm	*ReLu - Norm*
maxpool 1x3x3 with s-1x2x2x1	*maxpool 1x4x4 with s-1x2x2x1*
Fc – 576x192	*Fc – 1152x192*
Fc – 192x96	*Fc – 192x96*
Softmax - 96x10	*Softmax - 96x10*
# of hidden parameters 145530	*# of hidden parameters 268938*
batch size 256	*batch size 128*

Training Description

I have performed exponential decay on my learning rate which was 0.1 initially. Gradient Descent algorithm was used as my optimizer and as my cost function, the mean of the cross entropy per example was used with additional L2 regularization on all the variables i.e. weights and biases. I have trained network_1 with a batch size of 256 items and network_2 with 128 items. Therefore network_1 was taking 0.8 seconds per batch whereas the other network was taking 0.5 seconds. I've trained network_2 until 7049. step and 5143.step for network_1.

Experimental Results

Unfortunately, I couldn't implement McNemar evaluation.

Experimental Results	*Performance in %*	
Data Set	***Network 1***	***Network 2***
CIFAR-10	*73.2%*	*75.59%*

Discussion

If I would wait more and more it would be much more successful but at the same time it could overfit. Test set could be used to determine when to stop training.

The results are pretty similar, that's because I've changed both batch_size and neuron numbers per layer.