DD2424 Deep Learning in Data Science
Assignment 2 - Bonus

Gianluigi Silvestri
giasil@kth.se 940106-4614

April 18, 2018

1 Optimize the performance of the network

Early Stopping

All the improvements will be added individually and compared tho the simple network with RELU activation and early stopping. The early stopping stops the training procedure if the accuracy on the validation set does not increase for more than n epochs (for example n = 10). If this happens, the weights that gave the best accuracy on the validation set are returned. The training as been done on the whole CFAIR dataset, apart from 1000 images used as validation set. The parameters were: $\lambda \approx 0.00262$, $\eta \approx 0.084$, $\rho = 0.9$, weight decay=0.92 per epoch, 50 hidden nodes. The obtained accuracy is 57.56% for the training set and 53.1% for the validation set, while on the test set the accuracy is 53.11%. The algorithm was stopped after 99 iteration, after 30 epochs without improvement.

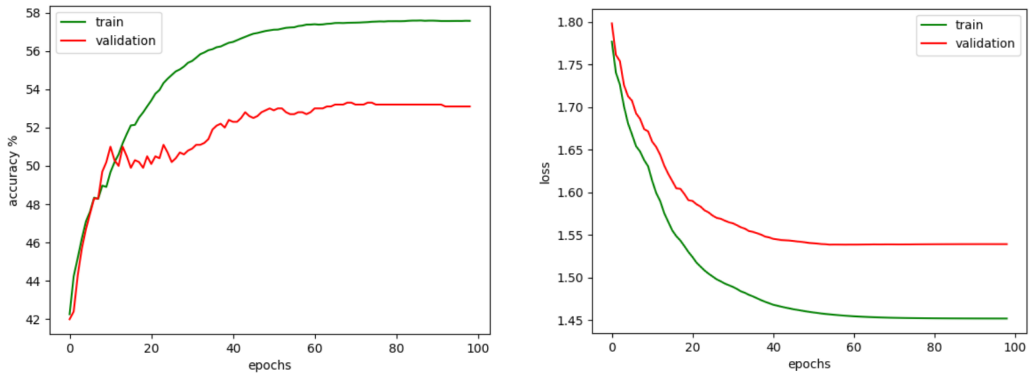

Figure 1: Left: Accuracy for the original network with early stopping. Right: Loss for the original network with early stopping.

He initialization

To implement the He initialization, the weights W_1 were initialized with standard deviation $\sigma = \sqrt{\frac{2}{3072}}$, while the weights W_2 were initialized with standard deviation $\sigma = \sqrt{\frac{2}{50}}$, where 50 is the number of hidden nodes and 3072 the dimension of each input vector. The obtained accuracy is 57.72% for the training set and 52.2% for the validation set, while on the test set the accuracy is 53.02%. The algorithm was stopped after 66 iterations, and the performances on the test set are slightly decreased.

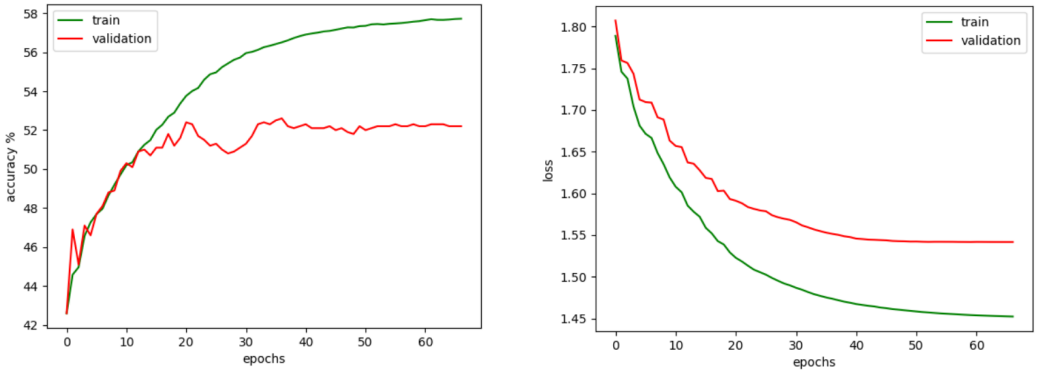

Figure 2: Left: Accuracy for the network with He initialization and early stopping. Right: Loss for the network with He initialization and early stopping.

More Hidden Nodes

Increasing the number of hidden nodes to 80 I have obtained slightly better results. The obtained accuracy is 60.65% for the training set and 53.9% for the validation set, while on the test set the accuracy is 54.02%. The algorithm was stopped after 90 iterations, and the performances on the test set are slightly increased.

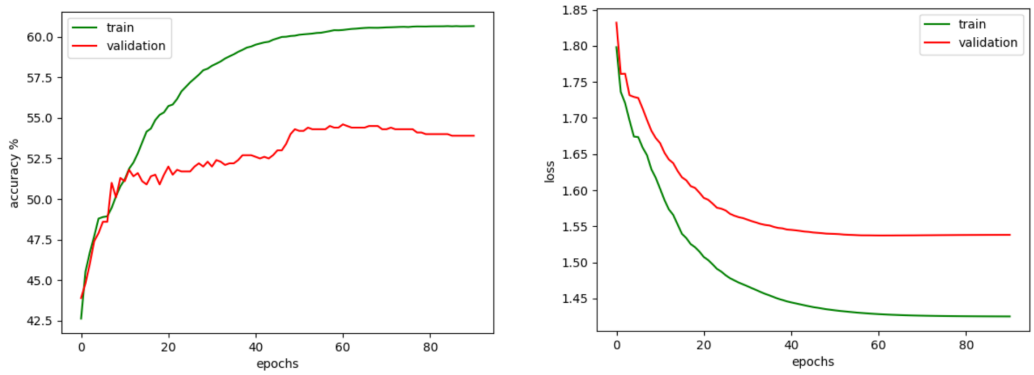

Figure 3: Left: Accuracy for the network with 100 hidden nodes and early stopping. Right: Loss for the network with 100 hidden nodes and early stopping.

Random Jitter

In each batch, at each epoch, some of the input images are flipped along the y axes with a certain probability. This is almost the same as increasing the size of the training data set. The obtained accuracy is 55.43% for the training set and 53.7% for the validation set, while on the test set the accuracy is 53.15%. The algorithm was stopped after 66 iterations, and the performances on the test set are slightly increased.

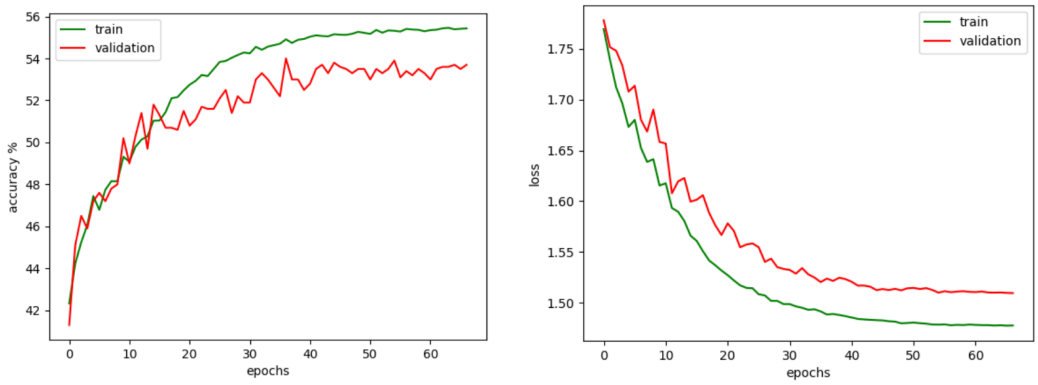

Figure 4: Left: Accuracy for the network with random image flipping and early stopping. Right: Loss for the network with random image flipping and early stopping.

Dropout

I have implemented the Inverted Dropout, so the only thing changed was the forward pass in the training set. The accuracy and cost were still computed on the whole network. The dropout is applied after computing the RELU, and the output of each hidden unit is active or not based on a Bernoulli sampling outcome. The dropout probability was 0.5, and the outcome of each active unit was then divided by this value. In addition, in the backprop the error is propagated only for the activated units. The obtained accuracy is 48.17% for the training set and 48% for the validation set, while on the test set the accuracy is 46.23%. The algorithm was stopped after 99 iterations. The performances decreased. This may be due to the fact that the network learns only with approximately 25 neurons at the time, and may not be enough to learn a good classification (Or maybe I did some mistake).

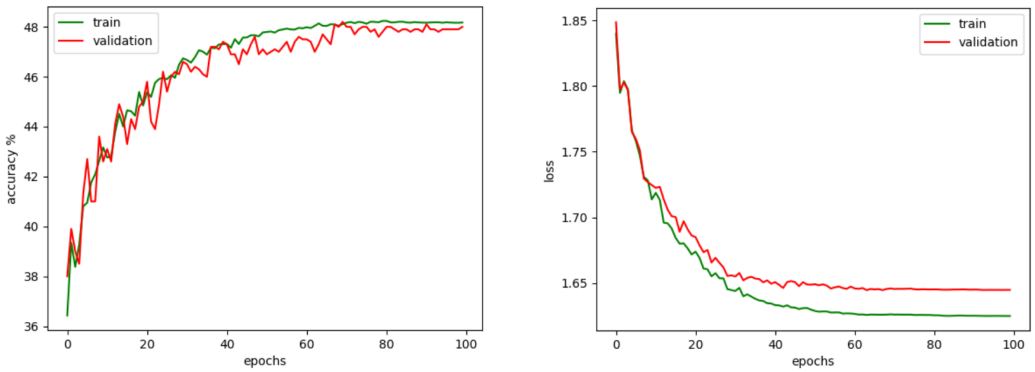

Figure 5: Left: Accuracy for the network with dropout and early stopping. Right: Loss for the network with dropout and early stopping.

2 Final best result

The best result was obtained with the following parameters: $\lambda \approx 0.00262$, $\eta \approx 0.084$, $\rho = 0.9$, weight decay=0.92 per epoch, 100 hidden nodes, He initialization, early stopping after 40 iterations with no improvement, and random jitter with probability 0.5. The obtained accuracy is 57.91% for the training set and 55.3% for the validation set, while on the test set the accuracy is 54.78%. The algorithm was stopped after 111 iterations.

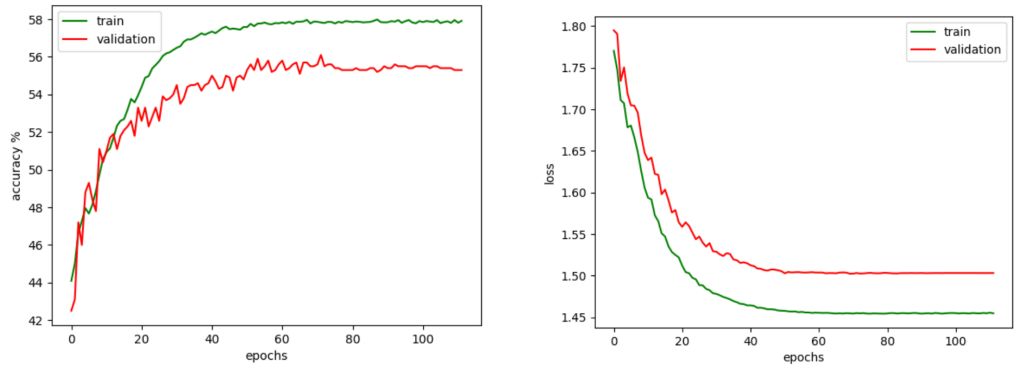

Figure 6: Best result with RELU

3 Leaky RELU

The leaky RELU was implemented and tested with the exact parameters as the best result, and the correctness has been checked with the same methods as for the RELU (both comparison with numerical gradient and overfitting test). The obtained accuracy is 57.02% for the training set and 54% for the validation set, while on the test set the accuracy is 54.38%. The algorithm was stopped after 85 iterations.

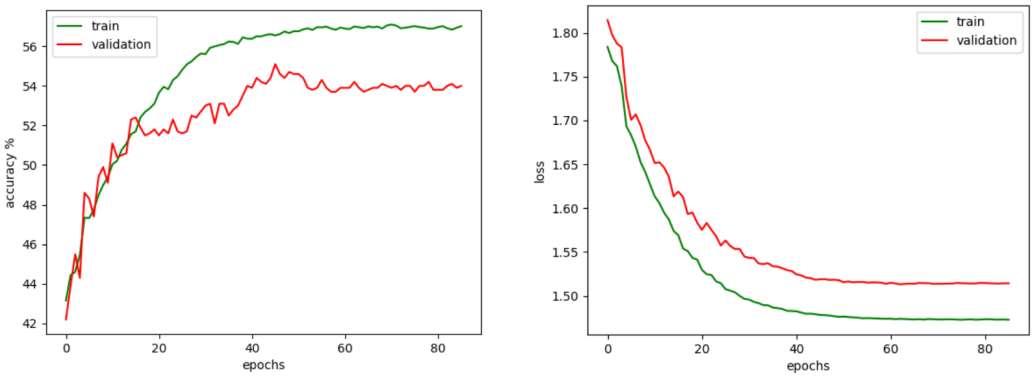

Figure 7: Best result with Leaky RELU