Assignment 2

Alexander Falk

DD2424, May 2024

i) Table 1 shows the difference between my analytical gradient computations and the numerically computed gradient vectors (for centered difference method). The gradients were computed on a small subset of the data. Difference seems to be sufficiently small, indicating that my analytical computations are correct.

Gradient	Difference
W1	1.865e-10
W2	1.165e-10
b1	1.408e-10
b2	5.063e-11

Table 1. Relative errors for weights and biases between analytic gradient computations and numerically computed gradient vectors using the centered difference method

ii) In the graphs below we can find the replicated figures from Assignment 2. The hyperparameters in the different figures were almost the same for all; eta_min = 1e-5, eta_max = 1e-1 and lambda = .01. The only difference was the value of n_s which was set to 500 for the single cycle and 800 for 3- cycles. The batch size was set to 100. The performance on the training and validation set was measured 10 times per cycle.

Figures 2.1 - 2.3 shows the plots for one cycle of training. Figure 2.4 - shows the plots for 3 cycles of training. For the plots with multiple cycles, we can clearly see that the training and validation curves vary in a similar cyclical nature. Some accuracy improvements were made for the 3- cyclical accuracy plot compared to the accuracy plot for one cycle, but the improvements were not that substantial.

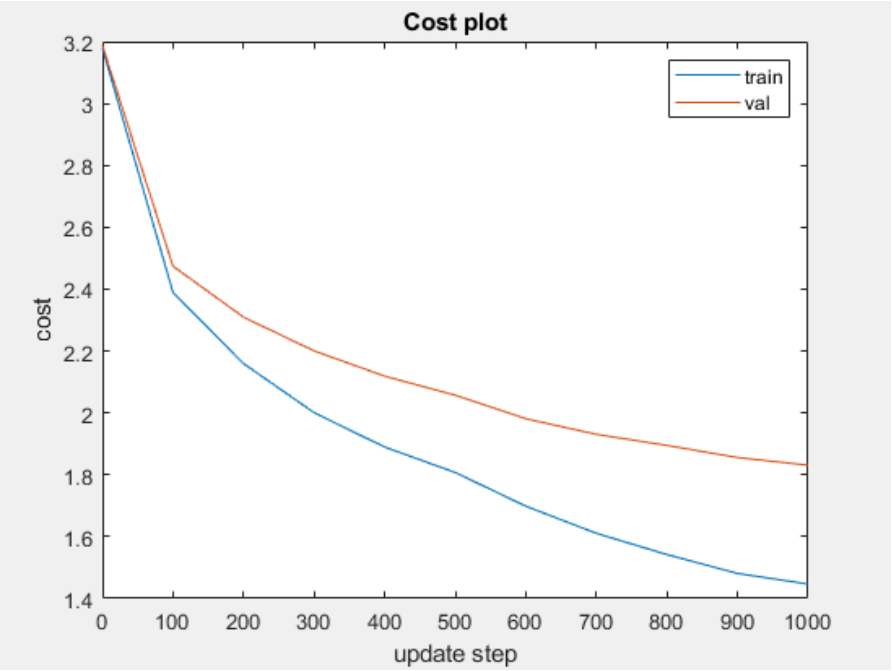

Figure 2.1. Curves for training and validation cost. One cycle.

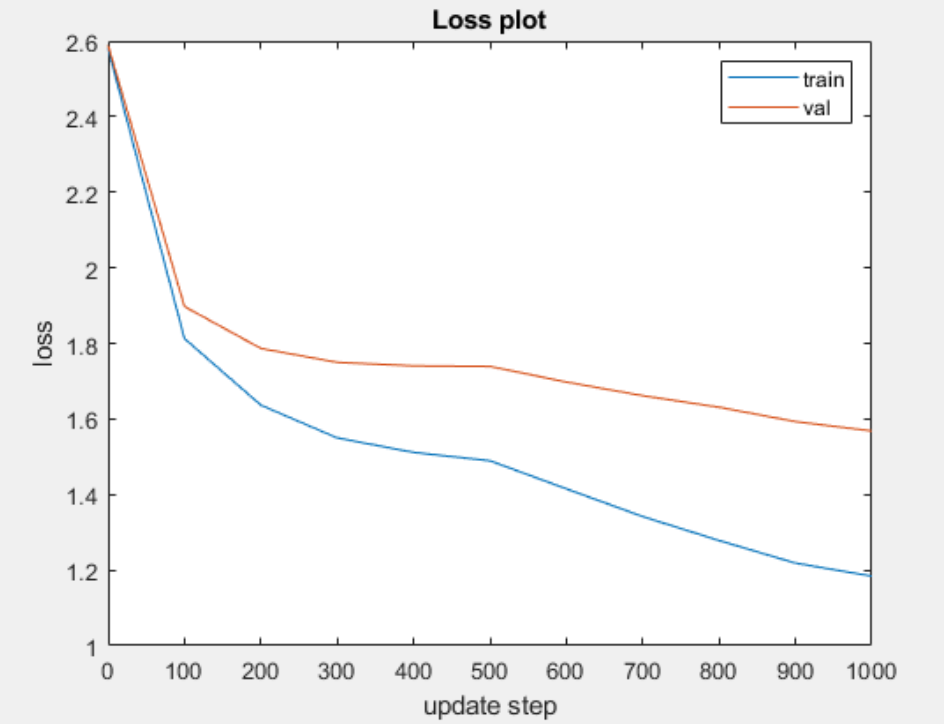

Figure 2.2. Curves for training and validation loss. One cycle.

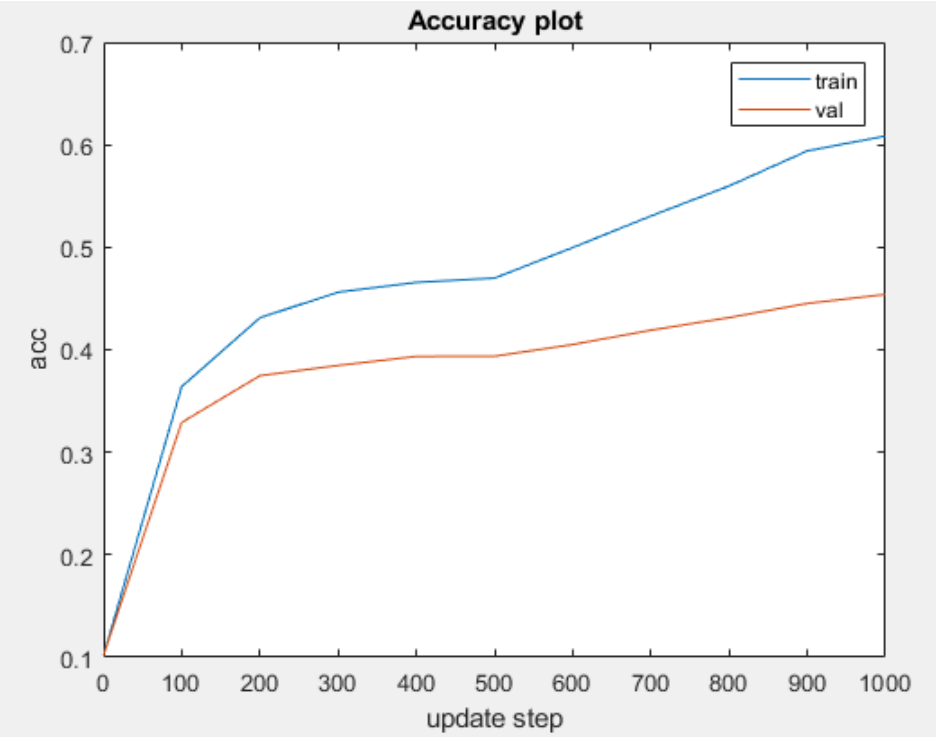

Figure 2.3. Curves for training and validation accuracy. One cycle.

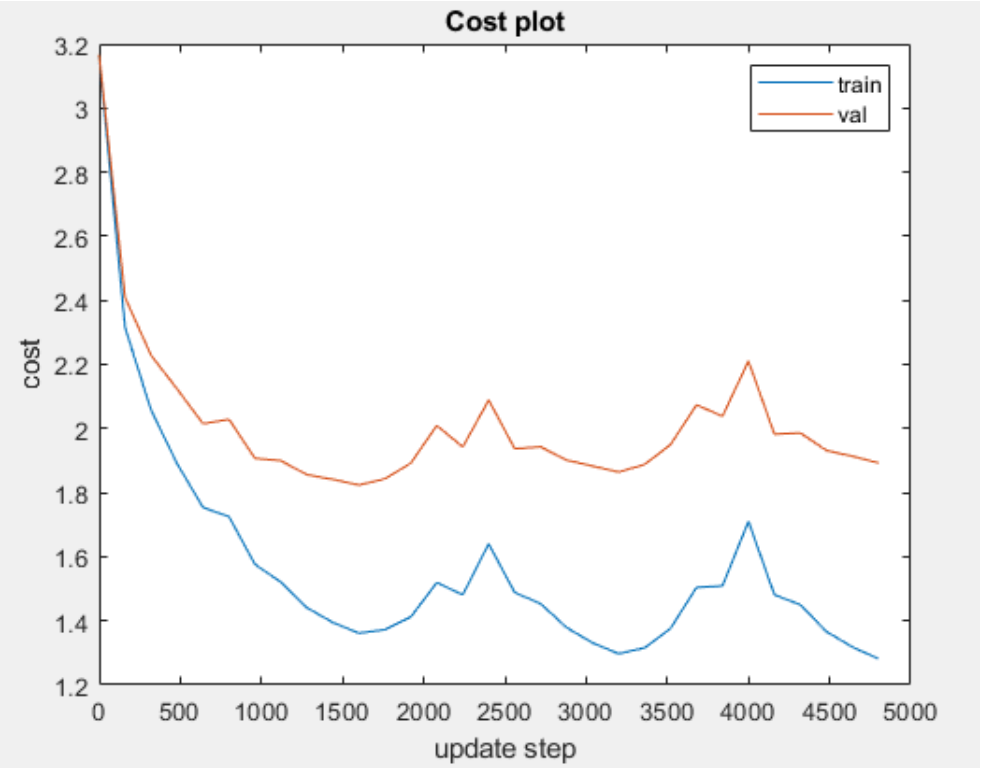

Figure 2.4. Curves for training and validation cost. 3 Cycles.

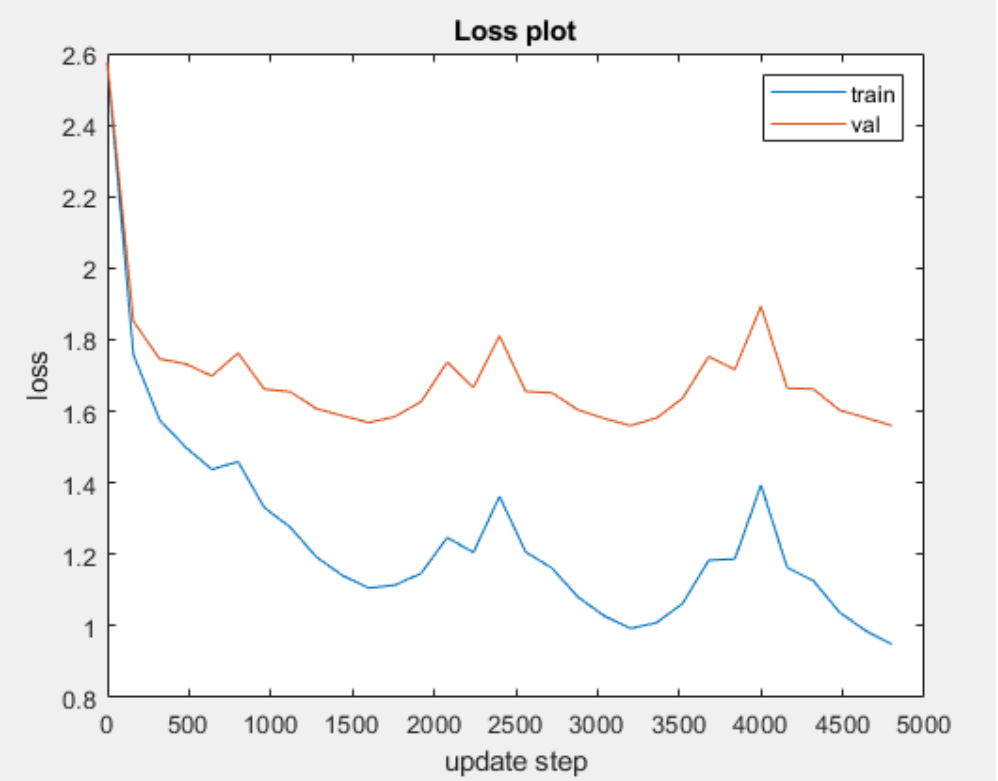

Figure 2.5. Curves for training and validation loss. 3 cycles

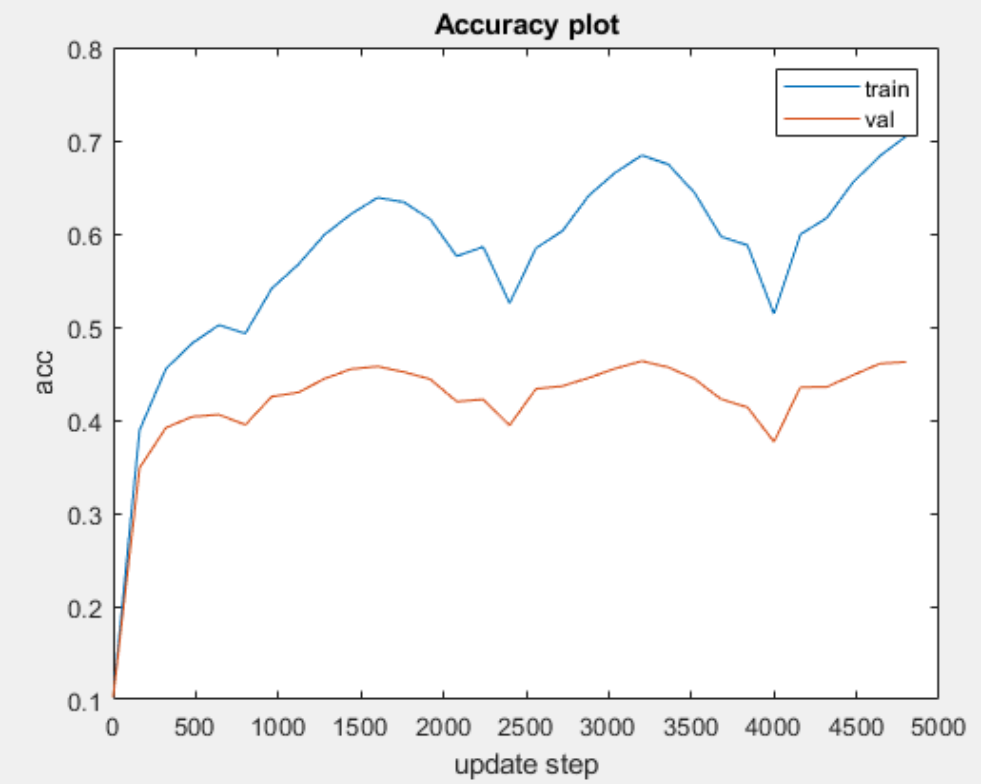

Figure 2.6. Curves for training and validation accuracy. 3 cycles.

iii) For the coarse search of lambda. I set l_min to 1e-5 and l_max to 1e-1. 10 lambdas were sampled and the most accurate are displayed in the table below. During training, I used 45000 images for training and 5000 for validation. Training was run for 2 cycles.

Lambda	Validation accuracy
4.782e-5	52.22 %
0.003	51.88 %
0.00016	51.88 %

Table 2. Top 3 values of lambda found during the coarse search using log-scale. In descending order.

iv) For the fine search of lambda, I used a more narrow range concentrated around the best accuracy from the coarse search. The ranges were from 4.772e-5 to 4.792e-5, i.e. [λ_best - 1e-6, λ_best + 1e-6]. I tried 10 different values in this range, from the lower bound to the upper bound with a uniform step size.

Top 3 performing lambdas are shown in Table 3. I used 45000 images for training and 5000 for validation. Training was run for 2 cycles. Validation accuracy was actually higher for the best lambda found during the coarse search than the best lambda found in the fine search. However, moving forward I used the best performing lambda from the fine search.

Lambda	Validation accuracy
4.792e-5	52.08 %
4.776e-5	51.74 %
4.781e-5	51.54 %

Table 3. Top 3 values of lambda found during the fine search. In descending order

v) I now used my best performing lambda (4.792e-5) to train my network on 49000 examples, saving 1000 for validation. Training was run for 3 cycles. The training and validation loss plots can be found in figure 5.1. The learnt networks performance on the test data was 51.37 %.

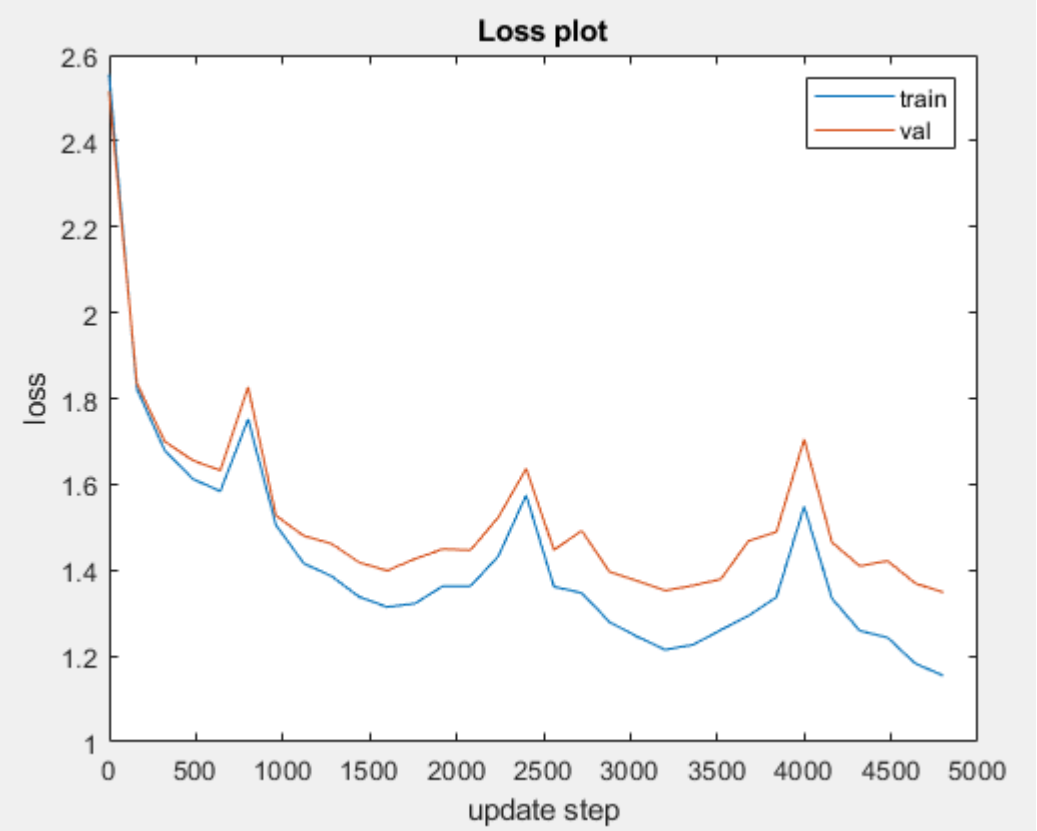

Figure 5.1. Curves for training and validation loss for 3 cycles of training. Using 49000 examples for training and 1000 for validation.