Hannes Kindbom

DD2424 Deep Learning – Assignment 1

Testing Gradient Implementation

To test and confirm that the analytical computations of the gradient were correct, the analytical gradient was compared to the numerical version (computed with the centered difference method). The relative error was computed, according to the formula in figure 1, for four different number of images and image dimensions. The results of the tests are shown in table 1.

$$\frac{|g_a - g_n|}{\max(\text{eps}, |g_a| + |g_n|)}$$

Figure 1. The relative error between a numerically computed gradient value g_n and an analytically computed gradient value g_a

Gradient test nr	Nr of images	Image dimension	λ	Relative error (W)	Relative error (b)
1	10	10	0	8.8074e-11	9.5178e-10
2	100	100	0	4.1719e-10	1.6369e-09
3	1000	1000	0.5	5.6111e-10	8.6363e-09
4	2000	20	0.5	1.9403e-10	1.5934e-08

Table 1. The conducted gradient tests. Note that step length of h=0.0001 was used in the numerical computations.

One may observe that the relative error is < 1e-7 in all test cases, which indicate that the gradient implementation is correct.

Experiments

Four experiments with the following parameter settings were thereafter conducted:

1. lambda=0, n_epochs=40, n_batch=100, eta=.1
2. lambda=0, n_epochs=40, n_batch=100, eta=.001
3. lambda=.1, n_epochs=40, n_batch=100, eta=.001
4. lambda=1, n_epochs=40, n_batch=100, eta=.001

The random number seed generator was fixed to rng(400) in all experiments.

Model Performance

The four pre-defined experiments yielded the following results:

Experiment	Accuracy (%)
1	28.16
2	38.69
3	39.08
4	37.57

Table 2. The test accuracy for each experiment

Experiment 1 – plots

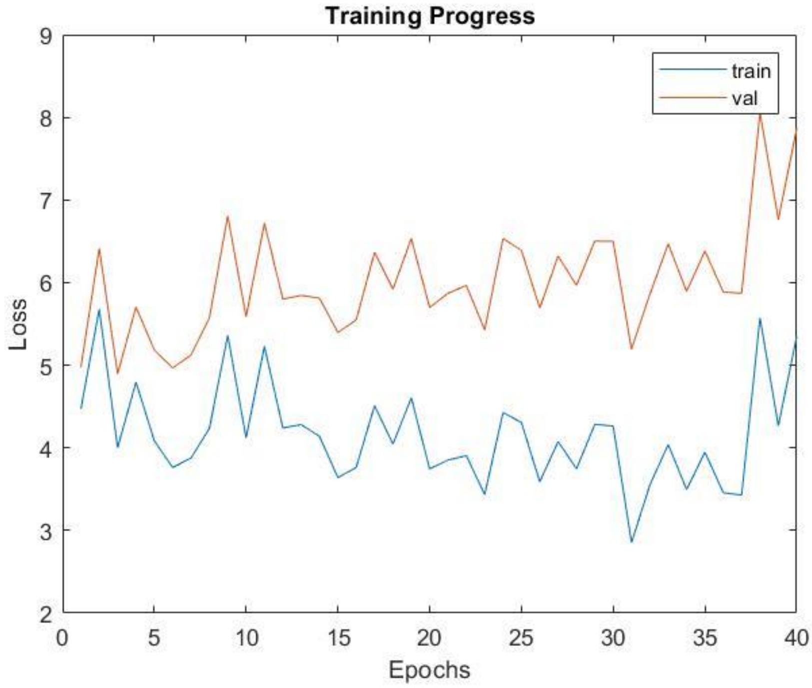

Figure 2. Training and validation loss in experiment 1 for each epoch

Figure 3. Images representing the learnt weight matrix after the completion of training in experiment 1

Experiment 2 – plots

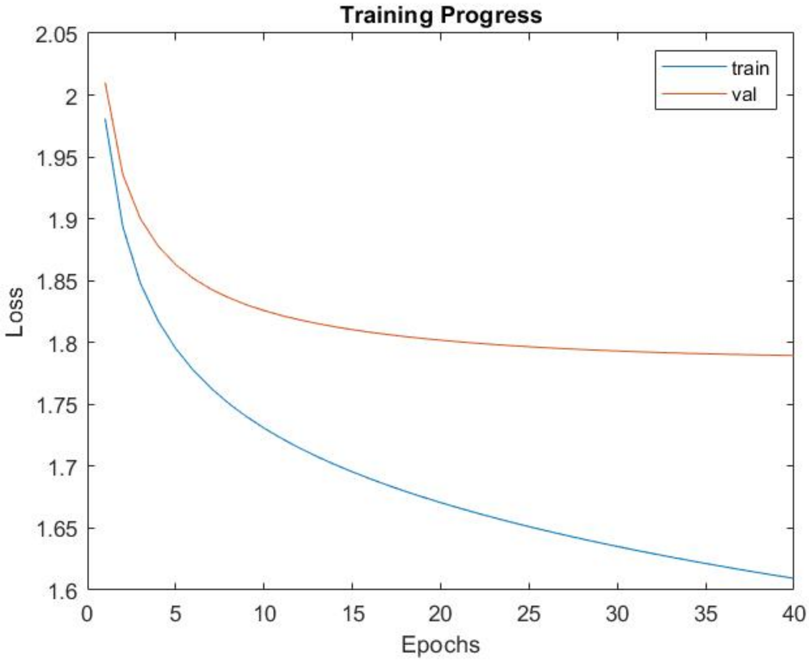

Figure 4. Training and validation loss in experiment 2 for each epoch

Figure 5. Images representing the learnt weight matrix after the completion of training in experiment 2

Experiment 3 – plots

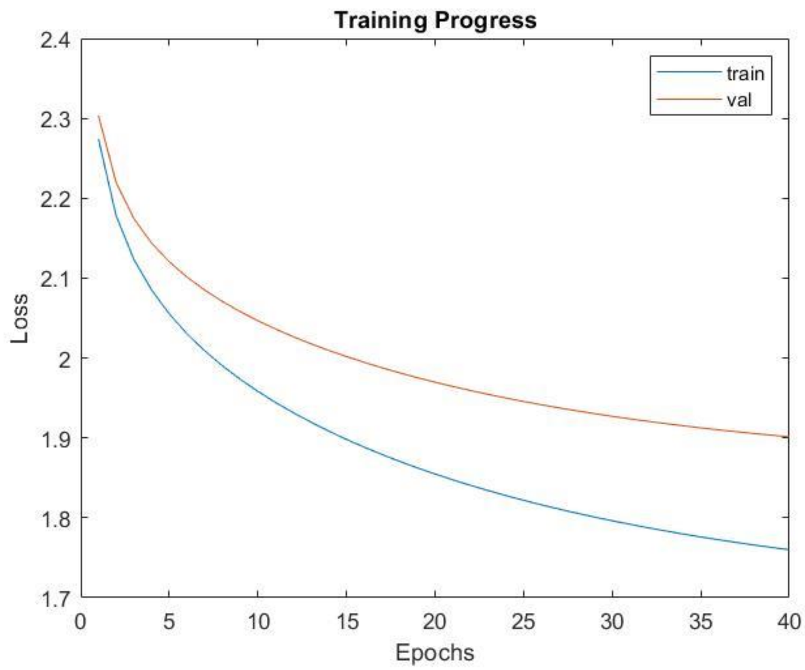

Figure 6. Training and validation loss in experiment 3 for each epoch

Figure 7. Images representing the learnt weight matrix after the completion of training in experiment 3

Experiment 4 – plots

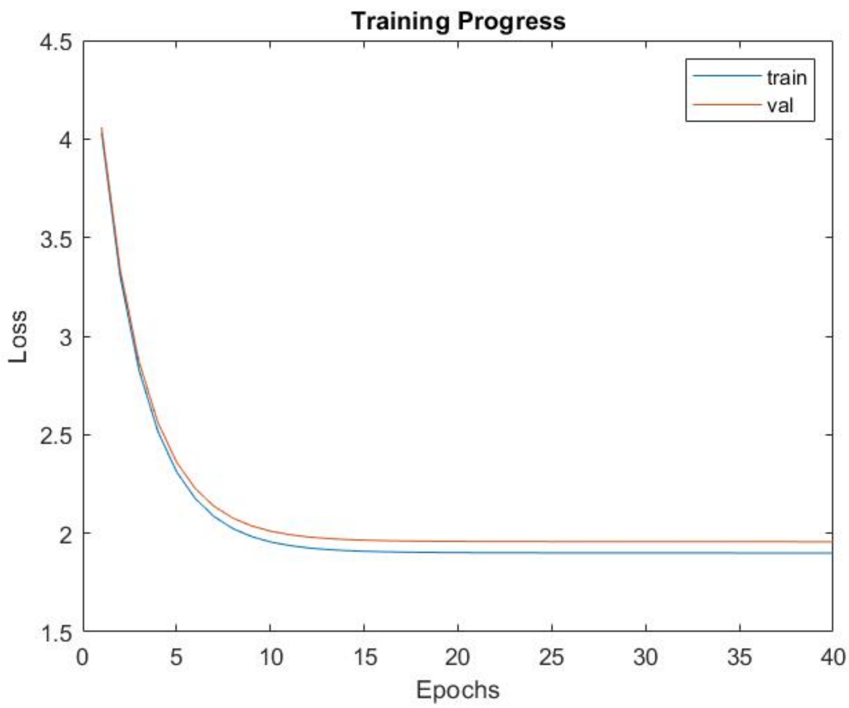

Figure 8. Training and validation loss in experiment 4 for each epoch

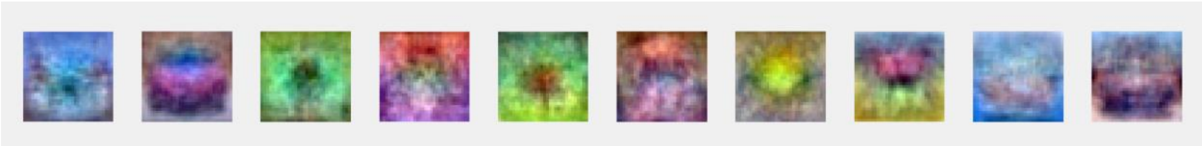

Figure 9. Images representing the learnt weight matrix after the completion of training in experiment 4

Comments on Result

Regarding the regularization, it can be observed that the model shows signs of overfitting when λ is small. This can be seen from the fact that the training loss decreases faster than the validation loss when λ is small (compare experiment 2 to 4). The regularization term penalizes large weights and thus decorrelates the neural network and mitigates overfitting. Regarding learning rate, it can be said that a too small learning rate increases time for learning, since only small updates are made in each iteration. However, a too large learning rate will stop the algorithm from converging to a loss minimum. A large learning rate will in many cases cause each update to "step over" the local minimum. This can be observed in experiment 1, where the loss oscillates around 5-6.