Deep2

Ishaq Ezaz

May 2024

1 Gradient Comparison

To ensure the accuracy of the analytic gradients, I compared them with the slow and more accurate version of the numerical gradient computation using the relative error formula:

$$\text{Relative Error} = \frac{\|\text{grad}_\text{a} - \text{grad}_\text{n}\|}{\max(\epsilon, \|\text{grad}_\text{a}\| + \|\text{grad}_\text{n}\|)}$$

where ϵ is a small number set to $1e-7$ to prevent division by zero.

The gradient checking produced the following relative errors between the analytic and numerical gradients:

- $W1 : 9.94281 \times 10^{-11}$
- $b1 : 3.59099 \times 10^{-10}$
- $W2 : 1.00939 \times 10^{-10}$
- $b2 : 1.58156 \times 10^{-10}$

These extremely low values suggest that the analytic gradients are computed correctly. The relative errors are significantly small, which indicates a high level of confidence in the reliability of the gradient computations implemented in the neural network.

2 Analysis of cyclical learning rate variations

The default values for both models are as follows:

Batch Size	Training Size	Learning Rate (η)	lambda
100	$10,000$	1×10^{-5} to 1×10^{-1}	1×10^{-2}

Table 1: Parameters for the model

Model 1: Update step 500, 1 cycle

This model was trained using default settings with update steps set to 500 and a single cycle.

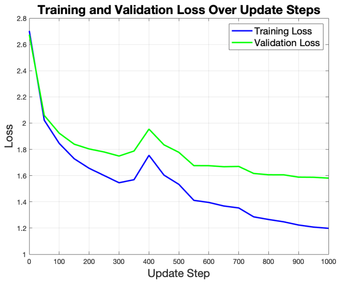

Figure 1: Training and validation loss over steps

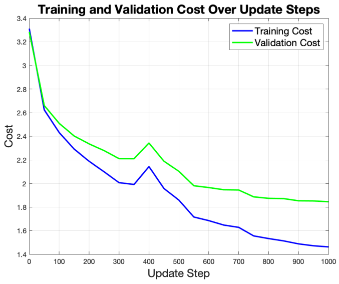

Figure 2: Training and validation cost over steps

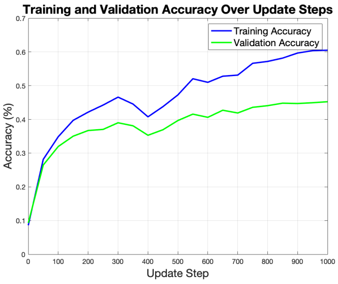

Figure 3: Training and validation accuracy over steps

Total test accuracy: 45.98%

Model 2: Update step 800, 3 cycles

This model was trained using default settings with update steps set to 800 and 3 cycles.

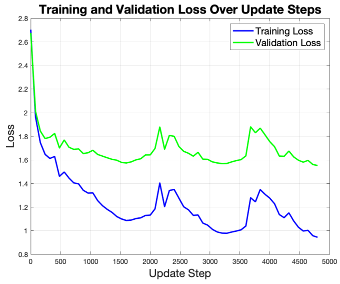

Figure 4: Training and validation loss over steps

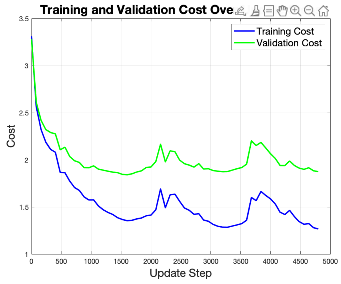

Figure 5: Training and validation cost over steps

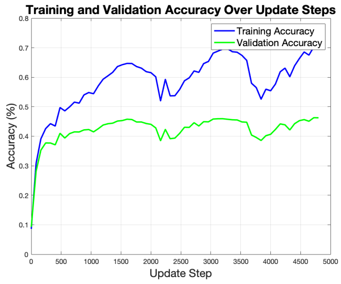

Figure 6: Training and validation accuracy over steps

Total test accuracy: 47.46%.

Analysis

The peaks observed in the graphs correspond to the cyclical learning rates reaching their maximum values. This elevated learning rate enables the model to explore a broader range of the solution space, which can be beneficial for escaping local minima and navigating flat regions. Following each peak, the learning rate decreases, which ideally helps to enhance generalization and smooth the learning curve.

3 Findings from coarse search

Batch Size	Training Size	Learning Rate (η)	Step Size
100	$45,000$	1×10^{-5} to 1×10^{-1}	900

Table 2: Parameters for the model

The lambda range was defined as `lambdas = logspace(-5, -1, 8)`. The training was run for 2 cycles. The results from the 8 lambdas run were as follows:

Lambda	Accuracy
1.00×10^{-5}	0.5088
3.73×10^{-5}	0.5052
1.39×10^{-4}	0.5176
5.18×10^{-4}	0.5194
1.93×10^{-3}	0.5210
7.20×10^{-3}	0.5200
2.68×10^{-2}	0.4692
1.00×10^{-1}	0.3792

Table 3: Accuracy results from the coarse search

4 Findings from fine search

The parameters used in the fine search were the same as in the previous model. The range for the fine-tuned lambda values was determined based on the three best-performing lambdas identified during the coarse search. Logarithmic spacing was used to create a more refined search space ranging from the minimum to the maximum of these top lambdas.

The results of the fine search were as follows:

Lambda	Accuracy
1.000×10^{-5}	0.5122
2.783×10^{-5}	0.5098
7.743×10^{-5}	0.5078
2.154×10^{-4}	0.5112
5.995×10^{-4}	0.5128
1.668×10^{-3}	0.5224
4.642×10^{-3}	0.5174
1.292×10^{-2}	0.5030
3.594×10^{-2}	0.4542
1.000×10^{-1}	0.3822

Table 4: Accuracy results for fine search

The most effective lambda was $\lambda = 1.668 \times 10^{-3}$ with an accuracy of 52,24%.

5 Final model

In this section, the training dataset is increased to 49,000, with the final 1,000 images reserved for the validation set. This adjustment means that the step size increases to 980. The rest of the parameters remain the same, except for the number of cycles, which has been increased to 3 from 2. The best-fitting lambda found in the previous step will be used, which is $\lambda = 1.668 \times 10^{-3}$.
Results

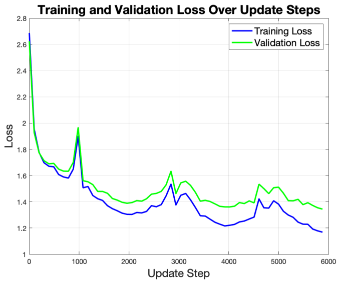

Figure 7: Training and validation loss over steps

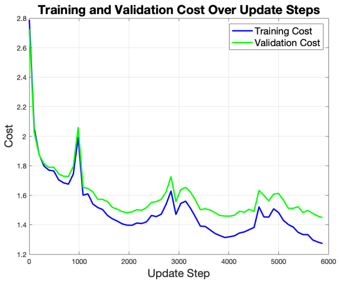

Figure 8: Training and validation cost over steps

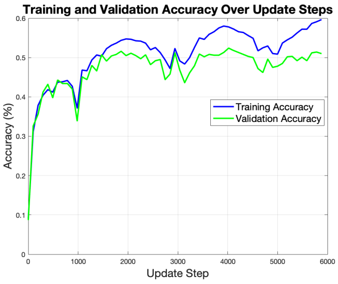

Figure 9: Training and validation accuracy over steps

The learnt network's performance on the test data resulted in: 51,2%.