DD2424 Deep Learning Assignment 1

Mikael Ljung (milju@kth.se)
940312–7435

April 2020

1. Introduction

In this paper a two-layer network with multiple outputs was trained to classify images from the CIFAR10-dataset. The CIFAR100 consists of 60 000 32x32 color images in 10 classes, with 6000 images per class.

2. Methods

The two-layer network was built from scratch in Python using mini-batch gradient descent on a cost function that calculated the cross-entropy loss. To make sure that the analytical gradient was calculated correctly it was compared to a numerical calculated gradient. Furthermore, the learning rate was set using a cyclical range of values, and the lambda parameter was found by first doing coarse search and after that a fine search to find its optimal value.

3. Results

3.1. Gradients

The analytically computed gradient was compared to a numerically computed gradient to verify that it was correct. This comparison was conducted using the following formula:

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

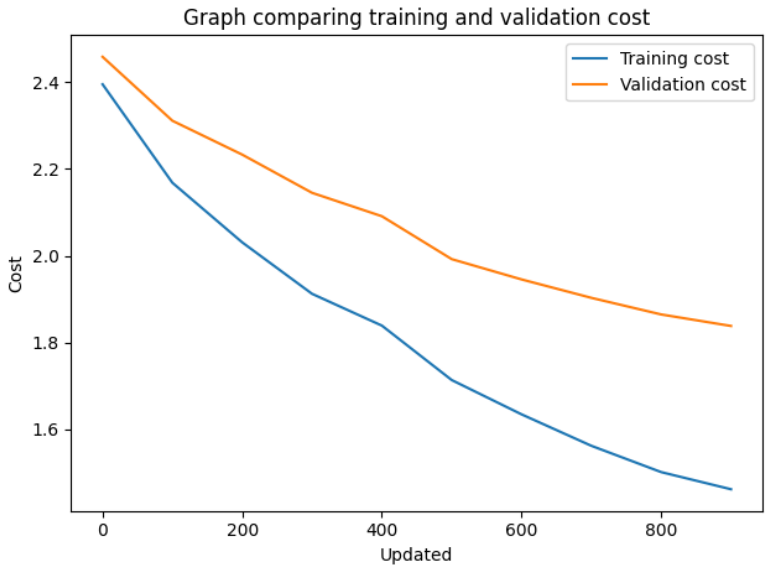
The values for the respective variable are presented in the table below:

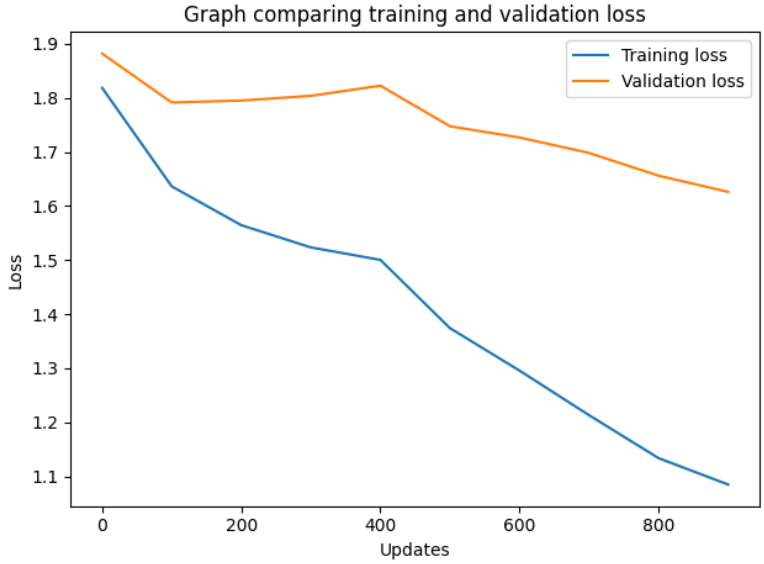
∑ gnW	-244.53249424283283
∑ gnb	0.2458705206098699
∑ gaW	-244.5324943211792
∑ gab	0.2458705185557267
∑ εw	6.999849254668195e-09
∑ εb	6.081649689753396e-09

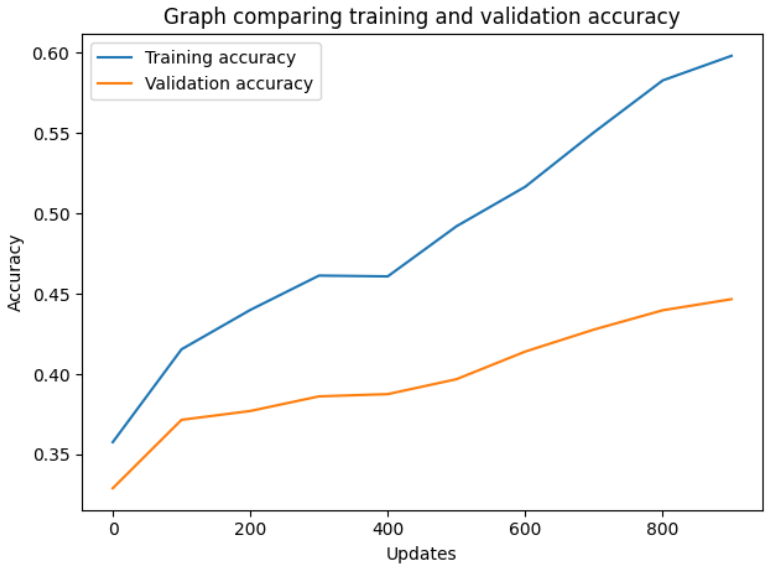
As can be seen in the table, the difference between the numerical and the analytical gradient is small, hence we can conclude that the analytical gradient is computed correctly.

3.2. Costs and losses

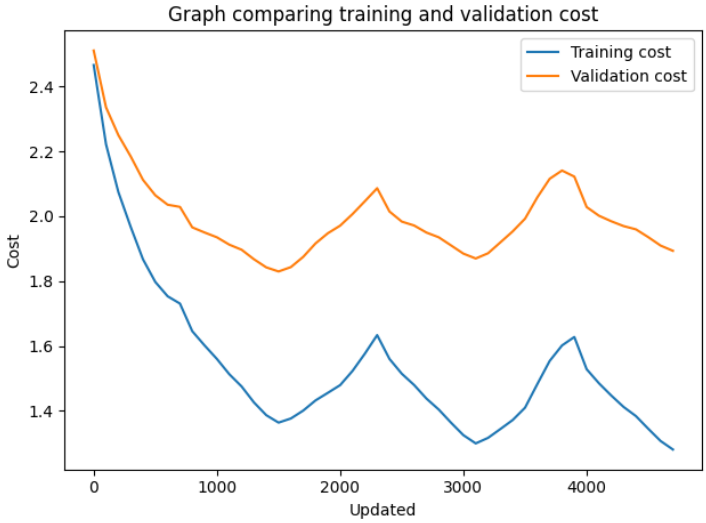
By setting the hyperparameters eta_min = 1e-5, eta_max = 1e-1, lambda=0.01, n_s=500 and training for one cycle the following plots showcasing the cost and loss during training were generated:

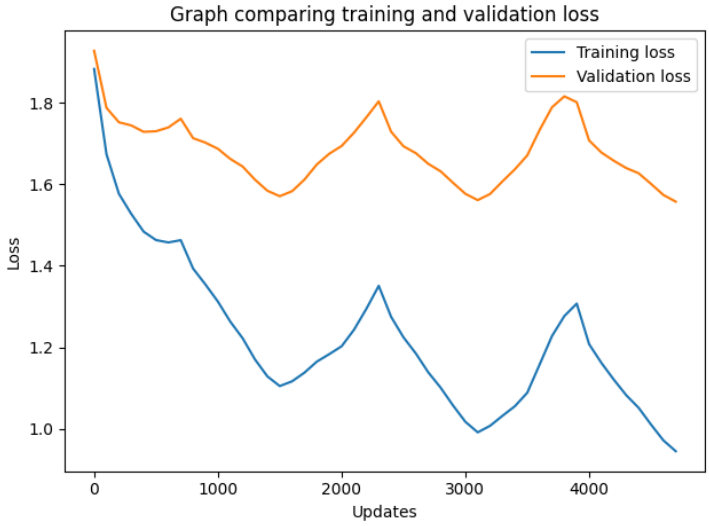

Graph comparing training and validation cost
Training cost
Validation cost
Cost
2.4
2.2
2.0
1.8
1.6
0
200
400
600
800
Updated

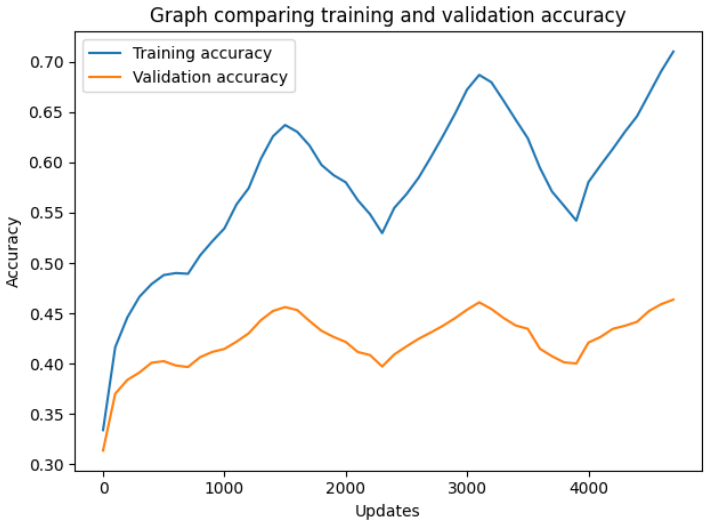

Graph comparing training and validation loss
Training loss
Validation loss
Loss
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
0
200
400
600
800
Updates


Graph comparing training and validation accuracy
Training accuracy
Validation accuracy
Accuracy
0.60
0.55
0.50
0.45
0.40
0.35
0
200
400
600
800
Updates

By setting the hyperparameters eta_min = 1e-5, eta_max = 1e-1, lambda=0.01, n_s=800 and training for three cycles the following plots showcasing the cost and loss during training were generated:







By observing the plots, we can see that the accuracy increases, and the loss decreases in a fast pace when the learning rate is small. During the second cycle, this pattern repeats itself indicating the model benefits from having a small learning rate. Furthermore, the improvements of the model during the third cycle is fairly small – indicating that training on more cycles than 3 might not result in a better model.

3.3. Coarse search

The coarse search was conducted with values ranging from 10^-5 to 10^-1 for the lambda parameter. The training was done with 2 cycles and the best values on lambda was the following:

Lambda = 00013810168311317438	Accuracy=0.4768
Lambda= 0.006551518774129204	Accuracy=0.4732
Lamdda= 0.0005389838990043178	Accuracy=0.4622

3.4. Fine search

Based on the coarse search, the fine search was conducted with values ranging between 10^-3 and 10^-2. The number of cycles were increased to 4 and the best accuracies were achieved with the following values:

Lambda = 009232948748803633	Accuracy=0. 5098
Lambda= 0.005059242510786357	Accuracy=0.515
Lambda= 0.0027095299004427586	Accuracy=0.5102

3.5 Final search

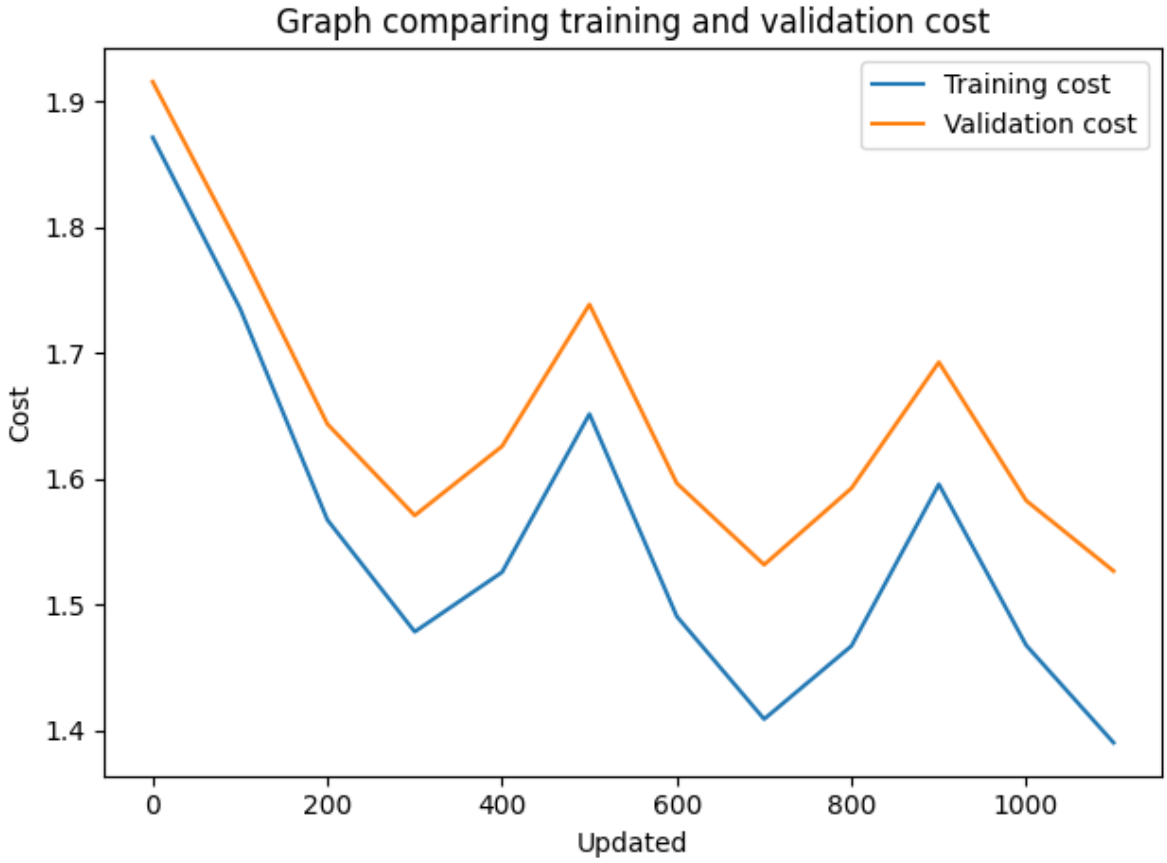
In the final version the two best lambda values from the fine search were used to evaluate the model. In this final round the cycles were increased to 8, and the following results were achieved:

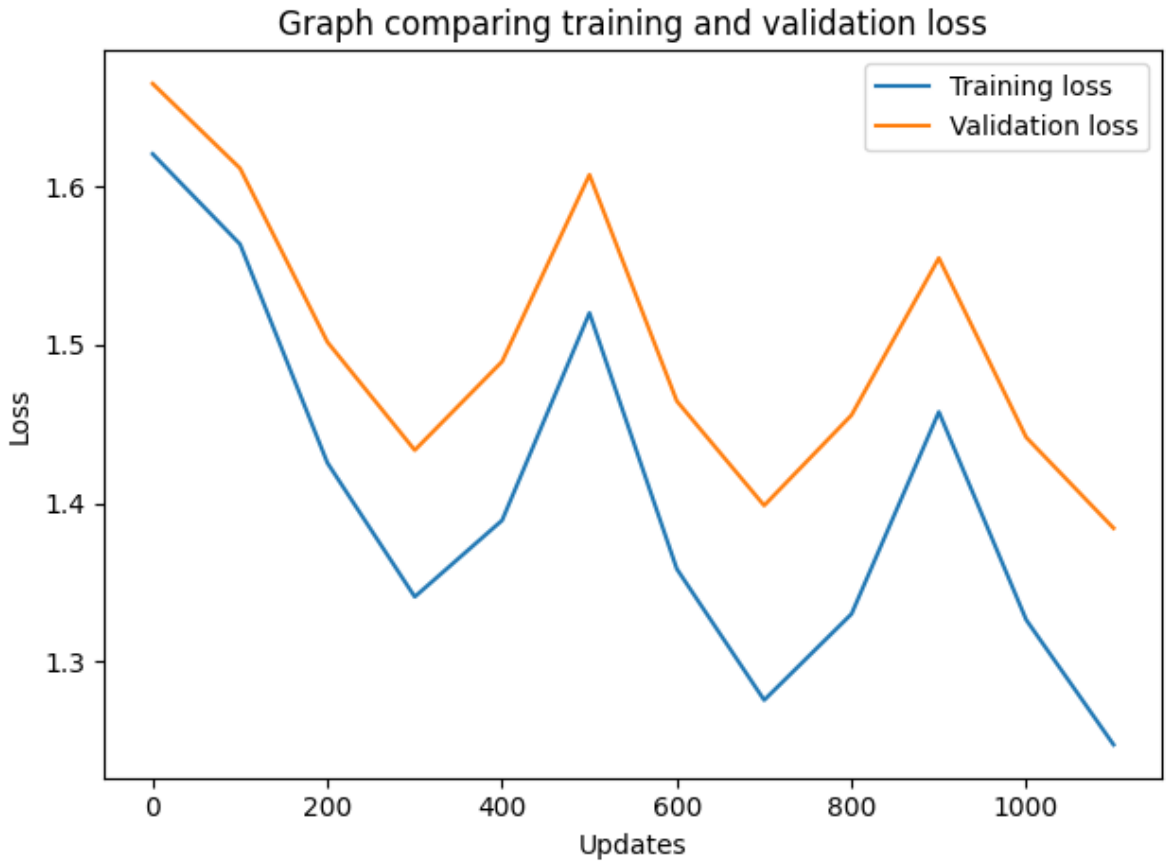
	Validation set	Test set
Lambda=0.005059242510786357	Accuracy=0.507	0.5048
Lambda=0.0027095299004427586	Accuracy=0.5104	0.5013

3.6 Losses and costs on final model

Using lambda = 0.005059242510786357 the final model predicted the test set with an accuracy of 50.48%.

The graph showcasing the cost and losses for training the final training is presented below:


Graph comparing training and validation cost
Training cost
Validation cost
Cost
1.9
1.8
1.7
1.6
1.5
1.4
0
200
400
600
800
1000
Updated


Graph comparing training and validation loss
Training loss
Validation loss
Loss
1.6
1.5
1.4
1.3
0
200
400
600
800
1000
Updates