DD2424 Deep Learning in Data Science
Assignment 2

Rui Shi
srui@kth.se

April 25, 2022

1 Introduction

In this assignment, all the functions and codes are implemented on Python. In this section, I trained and tested a two layer network with multiple outputs to classify images from the CIFAR-10 datasets.

2 Check gradients

In this section, I compared the difference between the analytical gradient and numerical gradient. Here are part of the results. From the results, we can see that the difference is tiny.

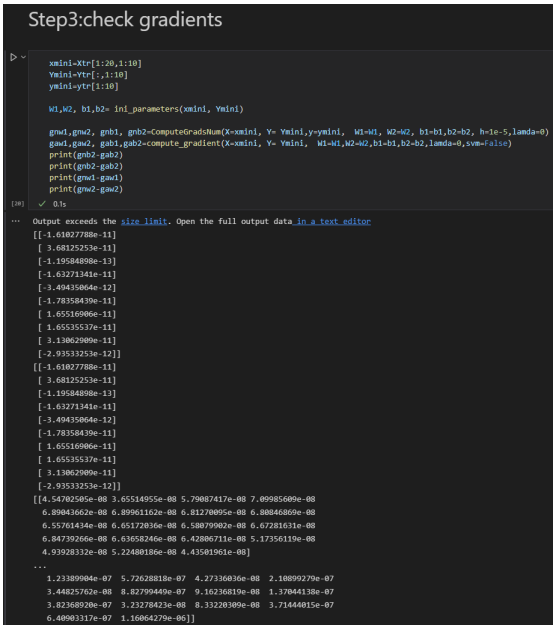


Figure 1: Difference between analytical gradient and numerical gradient

After that, I move to sanity check. I tried and trained my network on 100 examples with regularization turned off. From the results, I can overfit to the training data and get a very low loss on the training data after training for a sufficient number of epochs. Here are the graphs representing the training and validation loss. For the training data, the loss could be less that 0.003, and the accuracy could be around 0.99.

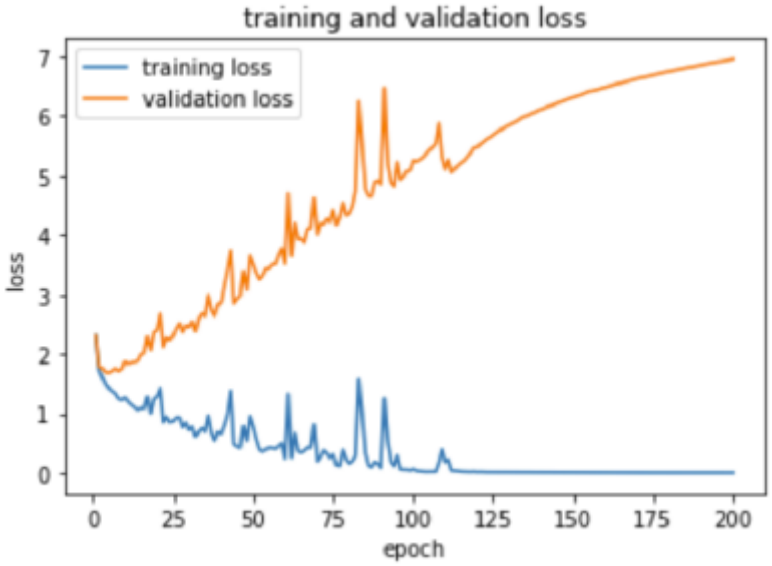


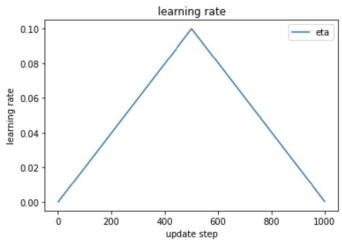
Figure 2: Training and validation loss

3 Cycling Learning Rate

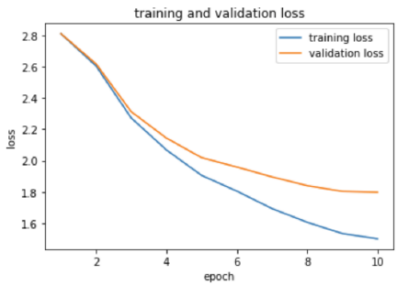
3.1 One cycle training

The parameters are defined as below:

eta_min = 1e-5, eta_max = 1e-1, lamda = 0.01, ns = 500. Here are the graphs representing learning rate, training and validation loss, training and validation accuracy.



(a) Learning rate



(b) Training and validation loss

(c) Training and validation accuracy

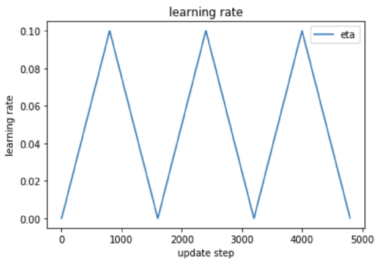
Figure 3: One cycle training

Finally, the accuracy could be up to 44.97%.

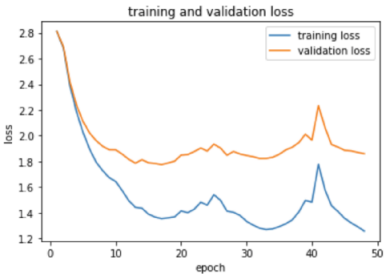
3.2 Three cycle training

The parameters are defined as below:

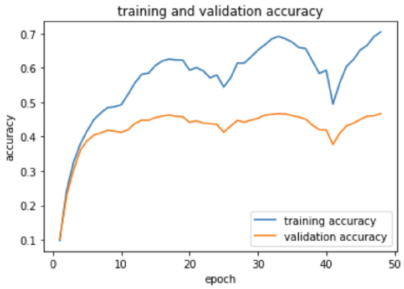
eta_min = 1e-5, eta_max = 1e-1, lamda = 0.01, ns = 800. Here are the graphs representing learning rate, training and validation loss, training and validation accuracy.



(a) Learning rate



(b) Training and validation loss



(c) Training and validation accuracy

Figure 4: One cycle training

Finally, the accuracy could be up to 46.75%.

4 Lambda searching

4.1 Coarse search

The parameters are defined as below:

l_min = -5, l_max = -1, ns = 200, eta_min = 1e-5, eta_max = 1e-1, n_cycle = 2. After coarse searching for lambda, the parameters of the three best networks are:

lambda	Test accuracy
10^(-3.196)	51.42%
10^(-2.325)	51.31%
10^(-2.826)	51.48%

4.2 Fine search

The parameters are defined as below:

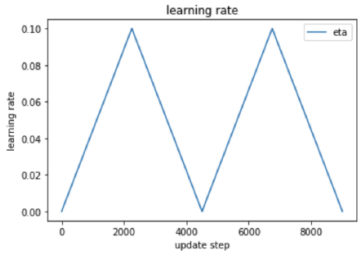
l_min = -4, l_max = -2, ns = 200, eta_min = 1e-5, eta_max = 1e-1, n_cycle = 2. After searching for lambda, the parameters of the best networks are:

lambda = 10^(-2.826380)

5 Best setting training

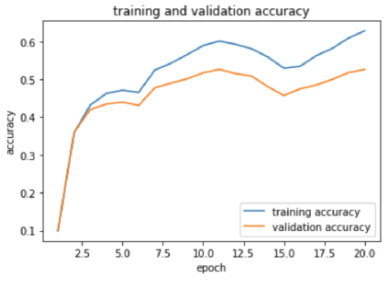
The parameters are defined as below:

lambda = 10^(-2.826380), ns = 500,n_epochs = 20, n_batch = 100 eta_min = 1e-5, eta_max = 1e-1.



(a) Learning rate

(b) Training and validation loss



(c) Training and validation accuracy

Figure 5: One cycle training

The final accuracy on test data could be 51.43%.