DD2424 Deep Learning in Data Science

Yuchen Gao

April 3, 2022

Introduction

This is the report for assignment 1, the implementation of basic feed forward network and backward propagation to update the weights.

Gradient comparison

Given the value of g_a, we can compute the relative error between the numerical gradient and the gradient we concluded basing on the equation:

$$\frac{|\ g_a\ | - |\ g_n\ |}{\max(\varepsilon, |\ g_a\ | + |\ g_n\ |)} \tag{1}$$

The result is listed below:

errW=4.504498528367488e-10	errb=2.949356382437707e-09

The error is acceptable, so the gradient descent will be carried on.

Graphs for the loss and cost function

0.1 Training condition

0.1.1 Parameters:

lambda=0,n_epochs=40,_batches=100,eta=0.1:

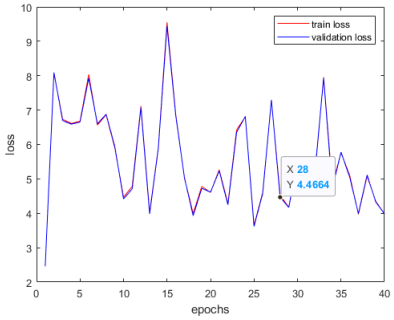

Figure 1: Loss function for 1 parameter setting.

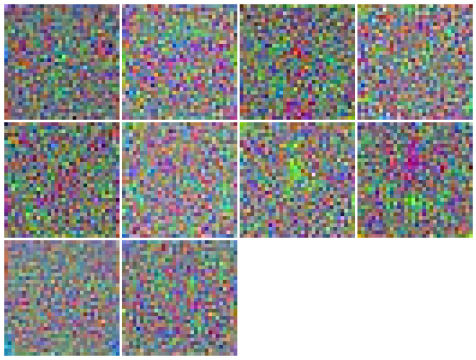

Figure 2: Template for 1 parameter setting.

0.1.2 Parameters:

lambda=0,n_epochs=40,_batches=100,eta=0.001:

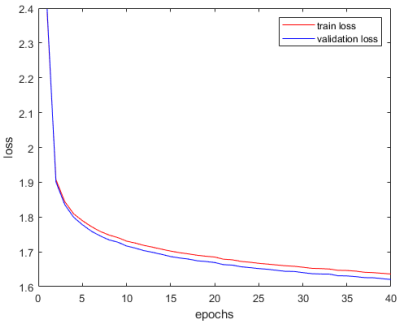

Figure 3: Loss function for 2 parameter setting.

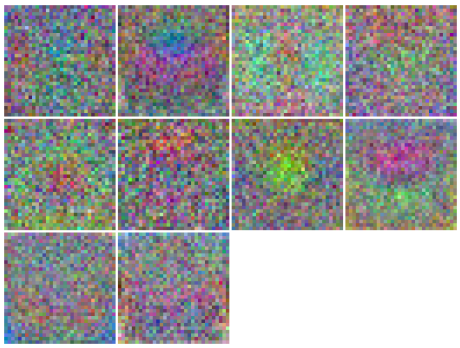

Figure 4: Template for 2 parameter setting.

0.1.3 Parameters:

lambda=0.1,n_epochs=40,_batches=100,eta=0.001:

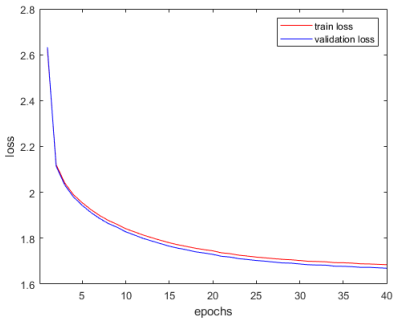

Figure 5: Loss function for 3 parameter setting.

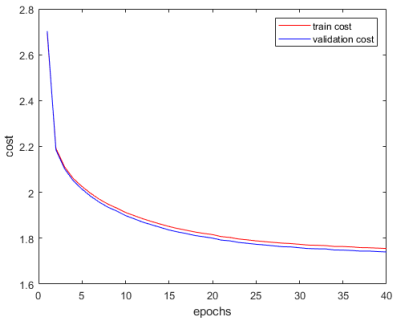

Figure 6: Cost function for 3 parameter setting.

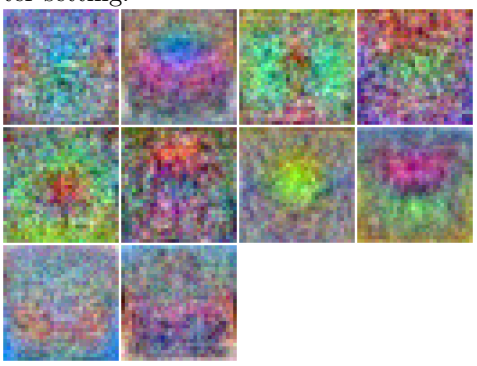

Figure 7: Template for 3 parameter setting.

0.1.4 Parameters:

lambda=1,n_epochs=40,_batches=100,eta=0.001:

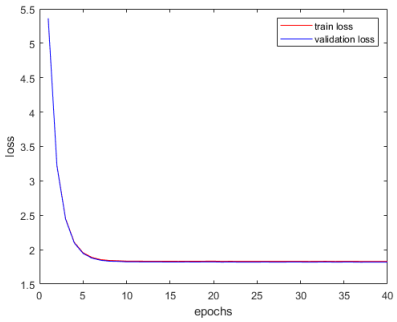

Figure 8: Loss function for 4 parameter setting.

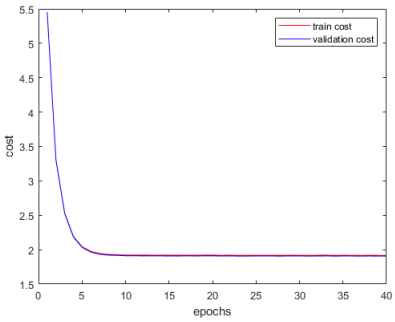

Figure 9: Cost function for 4 parameter setting.

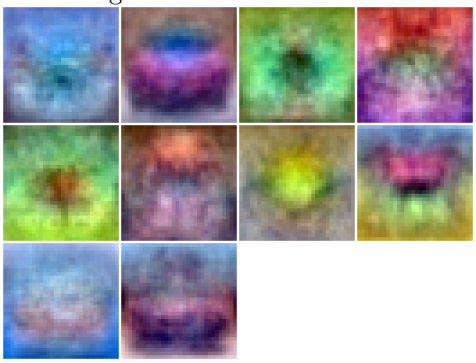

Figure 10: Template for 4 parameter setting.

0.2 Accuracy matrix

lambda	n_epochs	n_batches	eta	Test accuracy	Train accuracy
0	40	100	0.1	0.3840	0.3844
0	40	100	0.001	0.4587	0.4518
0.1	40	100	0.001	0.4428	0.4369
1	40	100	0.001	0.3904	0.3862

Comments

1.According to the test condition 1 and 2, we can tell that the larger learning rate is, the more unstable optimization process could be. Large learning rate may be benefit to the quick convergence, but it oscillate during the training epochs. The small learning rate made the loss function much smoother and it is intuitively that the loss function graph have the tendency to minimize itself.
2.From the condition 3 and 4, we can tell the small regularization parameter will

be benefit to the accuracy of prediction, but the large regularization parameter will take control of the loss function and the parameter adjustment procedure will be affected, so the accuracy is generally lower than the condition above. 3.Perhaps the 0.001 learning rate is really suitable for this specific task and the generated output is already optimal.