Assignment 2 Report
DD2424

Mehrdad Darraji

I. Analytic Gradient Computations

I computed the relative error between the numerically computed gradient value g_n and an analytically computed gradient value g_a with an eps value of 0.0001.

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

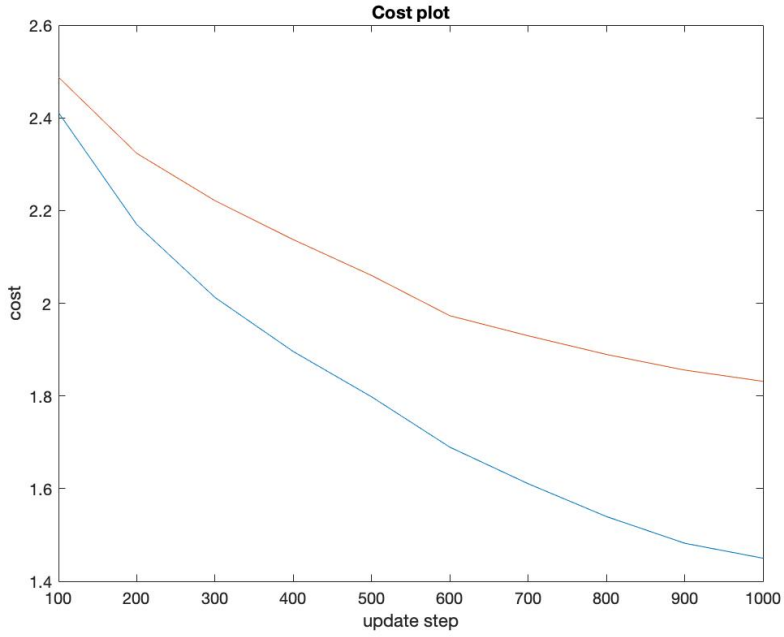
The result seemed to be very well as the error between my gradient function and the ComputeGradsNumSlow was smaller than 1e-6. There were relative error computations for each layer of the W and b model. Below are different tests I ran and results:

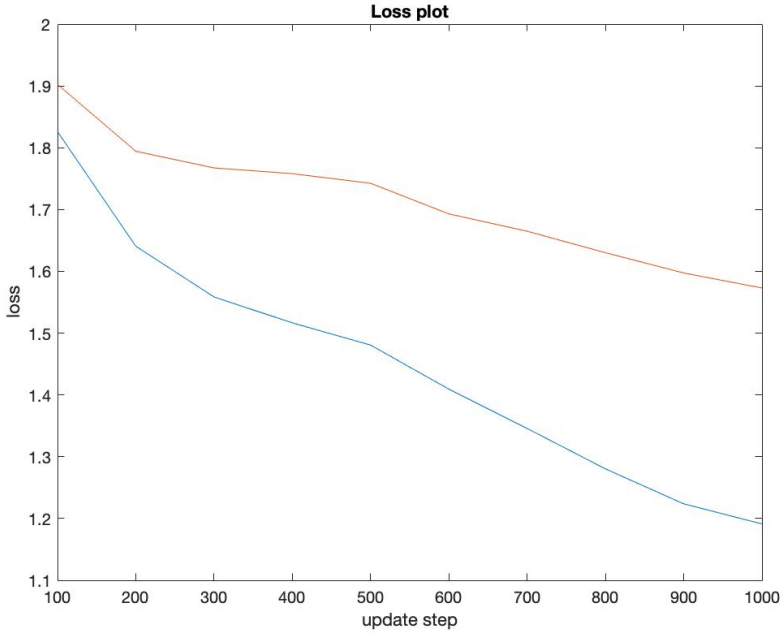
Batch Size	50	50	50	200	200	200
Lambda	0	0.1	1	0	0.1	1
W_1 error	2.35e-7	2.98e-7	6.81e-7	2.49e-7	3.63e-7	9.50e-7
W_2 error	5.19e-8	3.89e-8	9.26e-8	2.29e-8	4.46e-8	2.23e-8
b_1 error	1.69e-8	1.69e-8	5.86e-8	6.63e-9	1.29e-8	5.35e-8
b_2 error	1.72e-10	1.72e-10	4.37e-10	1.24e-10	2.12e-10	9.32e-10

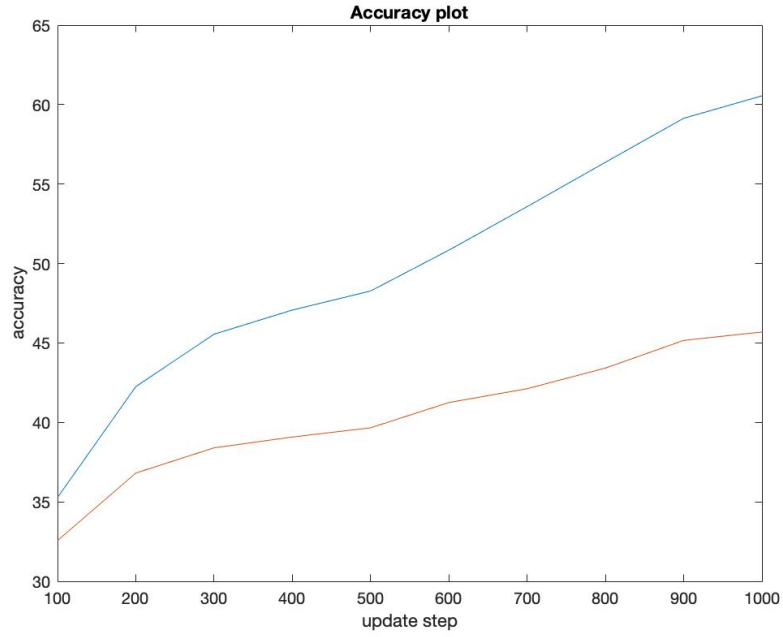
Note: in the plots below, blue represents the training, and red represent the validation dataset.

II. Cyclical Learning Rates with Default Values

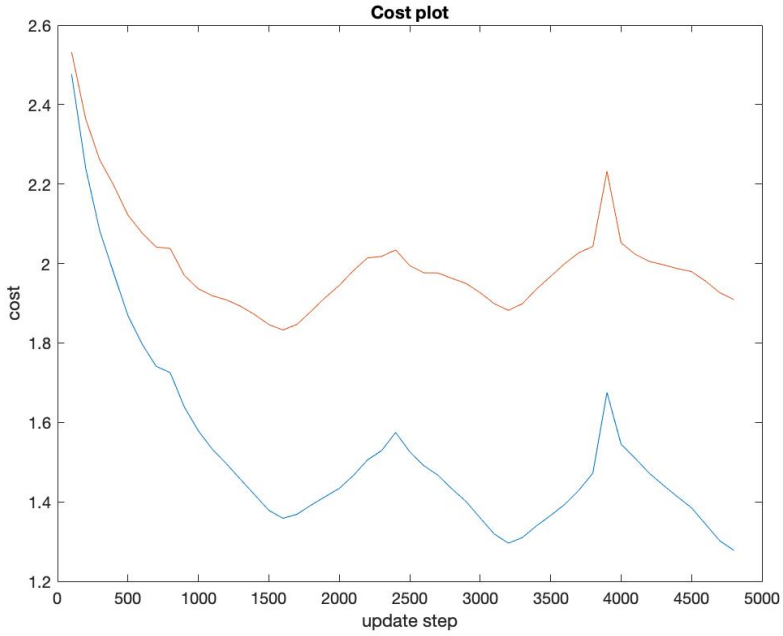
a. eta_min = 1e-5, eta_max = 1e-1, lambda = .01, number of steps = 500, batch size = 100, number of cycles = 1

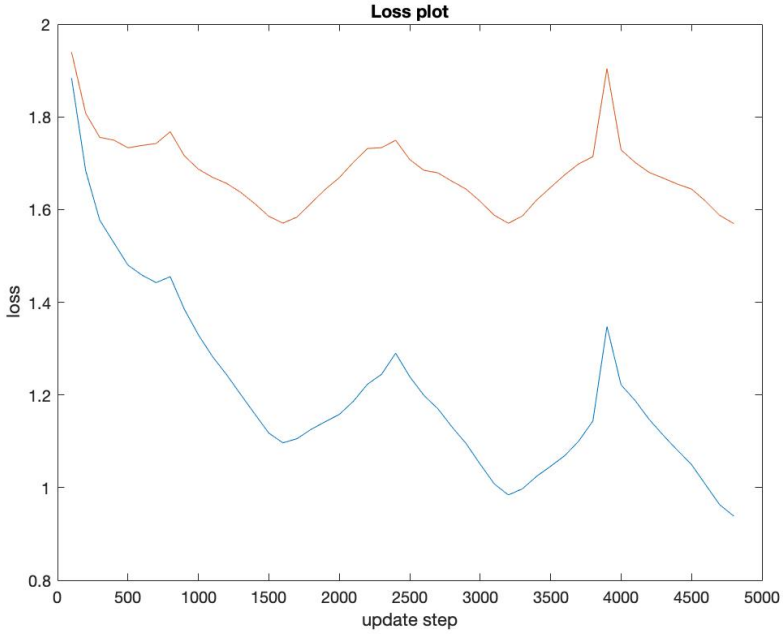


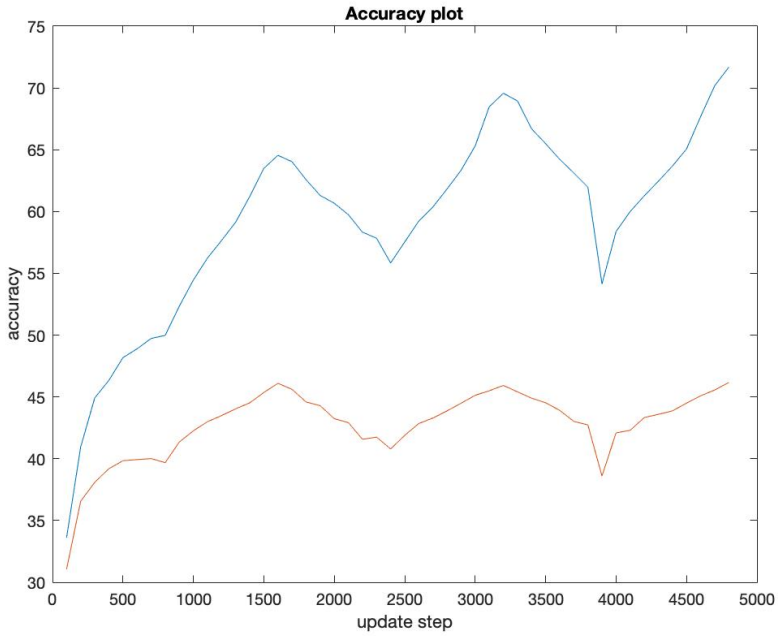

Loss plot
loss
update step
2
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
100
200
300
400
500
600
700
800
900
1000


Accuracy plot
accuracy
update step
65
60
55
50
45
40
35
30
100
200
300
400
500
600
700
800
900
1000

b. eta_min = 1e-5, eta_max = 1e-1, lambda = .01, number of steps = 800, batch size = 100, number of cycles = 3



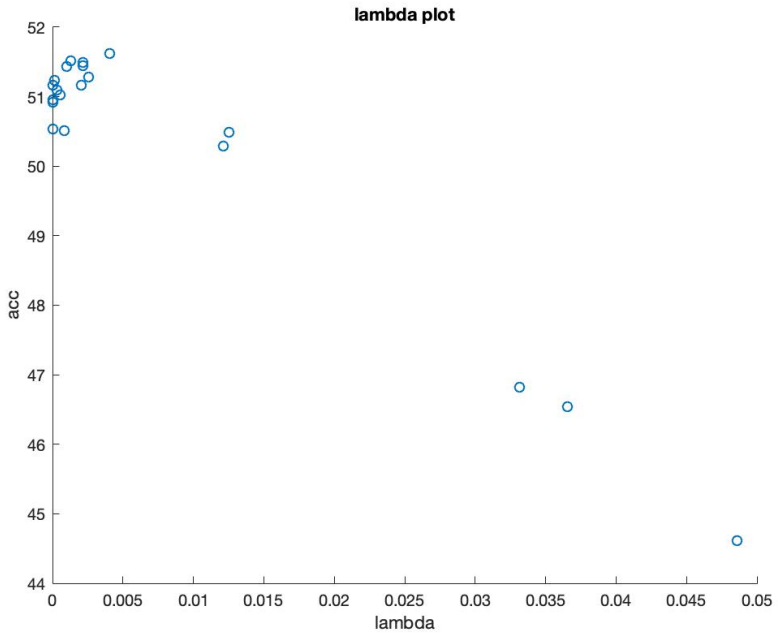




A test accuracy of 46.05% was achieved after performing one cycle of training, and test accuracy of 46.97% was achieved after performing three cycle of training. Even though increasing number of steps and cycles improved the accuracy, it appears that the increase made the plots very jagged and not smooth from each step forward.

III. Lambda Search

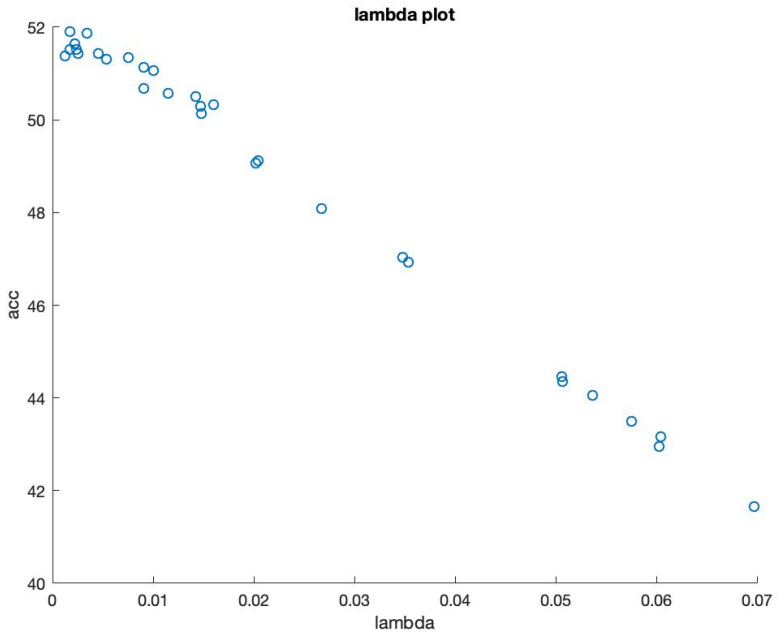
a. l_min = -5, l_max = -1, 20 values for lambda, n_s = 900, and 2 cycles



The table below shows the top three performing networks and the lambdas that were found:

Accuracy	51.62%	51.51%	51.49%
Lambda	4.073837e-03	1.324601e-03	2.179494e-03

b. l_min = -3, l_max = -1, 30 values for lambda, n_s = 900, and 3 cycles

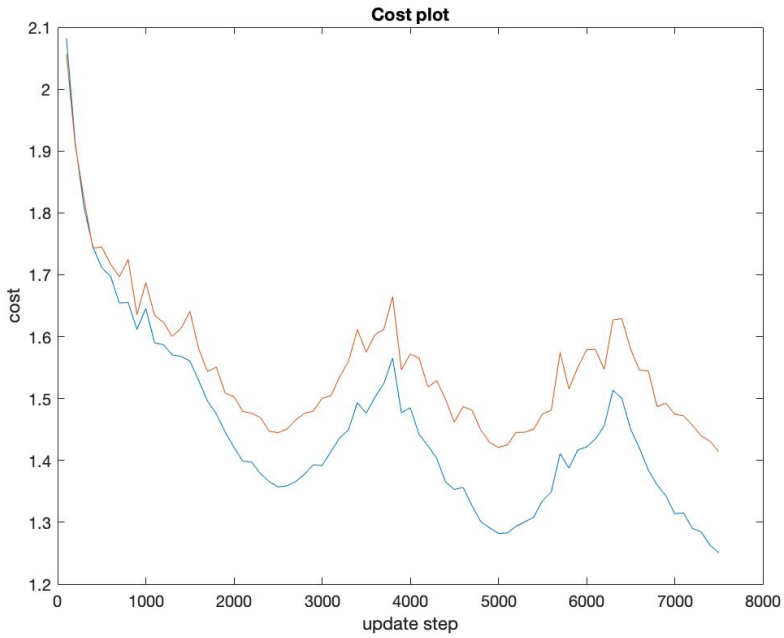


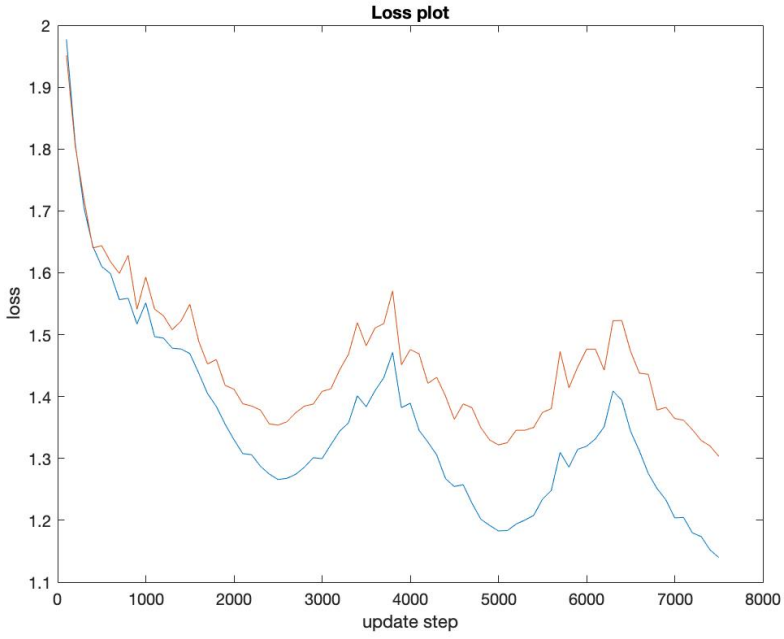
The table below shows the top three performing networks and the lambdas that were found:

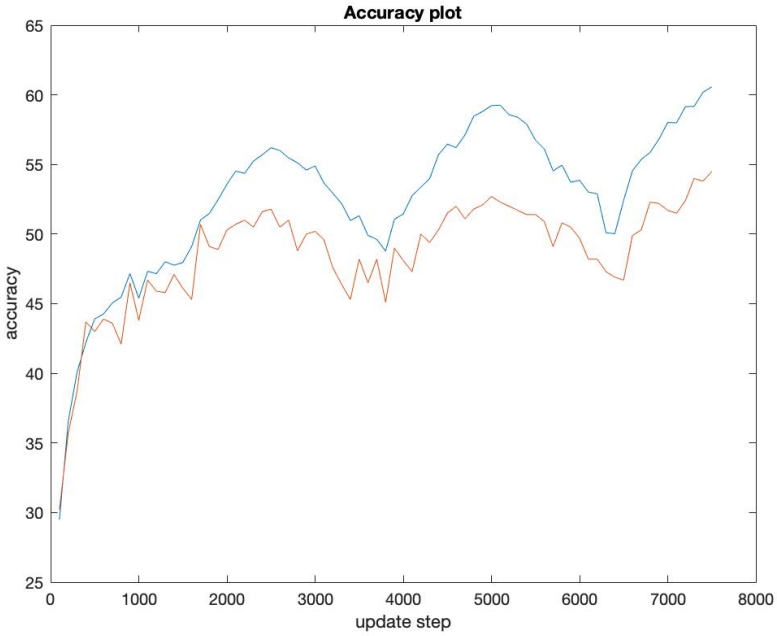
Accuracy	51.90%	51.87%	51.64%
Lambda	1.73825e-03	3.397505-03	2.210905e-03

IV. Training Using Best Lambda Setting

a. For the final test, I used the best lambda setting, 1.73825e-03, with a trained network on all the training data except 1000 examples for a validation set, n_s of 1250, and 3 cycles.







The final test accuracy achieved with the best lambda was 51.58%.