

KTH Royal Institute of Technology
DD2424-VT19-1 Deep Learning in Data Science
Assignment 2

Marina Herrera Sarrias

April 15, 2019

i) State how you checked your analytic gradient computations and whether you think that your gradient computations were bug free. Give evidence for these conclusions.

I verified the analytical computations of the gradients using numerical estimations. The results can be found in the table below.

	Maximum Relative Error	
	Finite Difference	Centered Difference
W_1	3.62e-03	1.96e-03
b_1	8.84e-06	8.54e-07
W_2	6.05e-05	8.90e-06
b_2	1.01e-04	1.15e-07

Table 1: Maximum Relative error between numerical and analytical gradient vectors computations.

We can verify that the centered difference formula provides more accuracy in approximating to the analytical computations of the gradients.

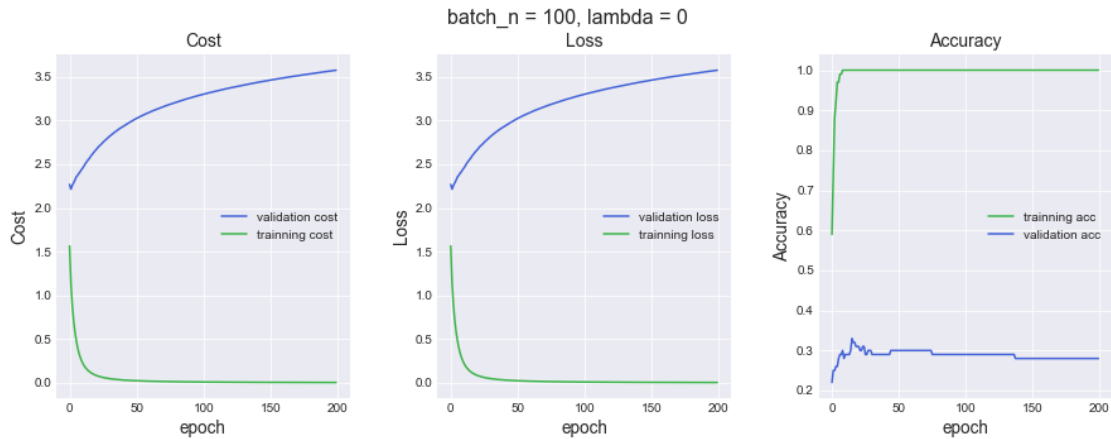


Figure 1: The graph of the training and validation cost, loss and accuracy computed after every epoch. The network was trained with the following parameter settings: `n_batch= 100 eta= .01, n_epochs= 40` and `lambda= 1`.

ii) The curves for the training and validation loss/cost when using the cyclical learning rates with the default values, that is replicate Figures 3 and 4. Also comment on the curves.

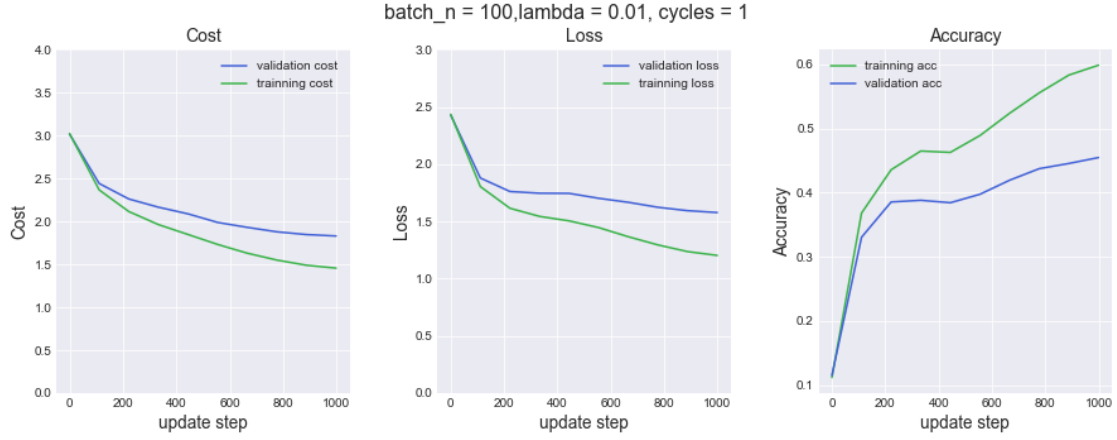


Figure 2: The graph of the training and validation cost, loss and accuracy computed after every epoch. The network was trained with the following parameter settings: `n_batch= 100 eta= .01, n_epochs= 40` and `lambda= 1`.

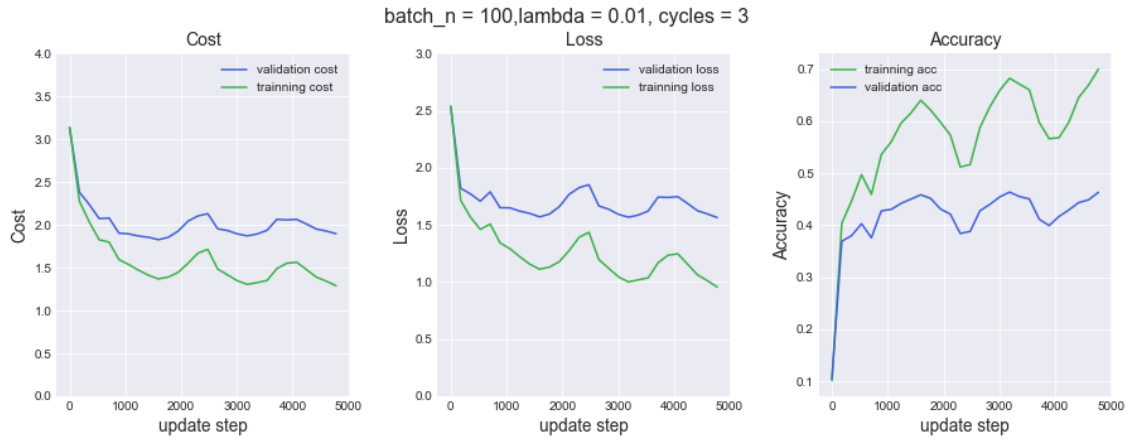


Figure 3: The graph of the training and validation cost, loss and accuracy computed after every epoch. The network was trained with the following parameter settings: `n_batch= 100 eta= .01, n_epochs= 40` and `lambda= 1`.

iii) State the range of the values you searched for `lambda`, the number of cycles used for training during the coarse search and the hyper-parameter settings for the 3 best performing networks you trained.

iv) State the range of the values you searched for `lambda`, the number of cycles used for training during the fine search, and the hyper-parameter settings for the 3 best performing networks you trained.

v) For your best found `lambda` setting (according to performance on the validation set), train the network on all the training data (all the batch data), except for 1000 examples in a validation set, for 3 cycles. Plot the training and validation loss plots and then report the learnt network's performance on the test data.