Assignment 2 - Report

Pietro Alovisi

Gradient Check

To check the gradient I computed for each layer the maximum of the absolute difference of the gradients, as in the equations below:

$$max(|W_i - W_{i\ num}|)$$
$$max(|b_i - b_{i\ num}|)$$

Where $\cdot_{num}$ represents the numerical computed value. The

Table 1: Maximum absolute difference between the computed gradients for the parameter of each layer (first column) for different network configuration. The lists like $[10, 10, 10]$ mean that there are 3 hidden nodes with 10 nodes each. For all the networks λ was set to 1.0.

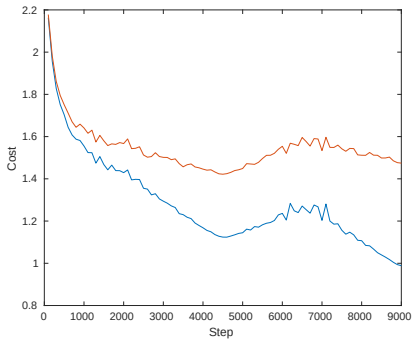
Inner nodes $\rightarrow$	$[50]$	$[20, 10]$	$[10, 10, 10]$
W_1	7.01e-10	8.33e-10	6.71e-10
b_1	3.49e-10	2.33e-10	3.18e-10
W_2	5.51e-10	5.48e-10	4.97e-10
b_2	3.12e-10	2.86e-10	3.18e-10
W_3	-	4.58e-10	5.77e-10
b_3	-	3.05e-10	1.80e-10
W_4	-	-	4.90e-10
b_4	-	-	3.36e-10

Test Multi-Layer

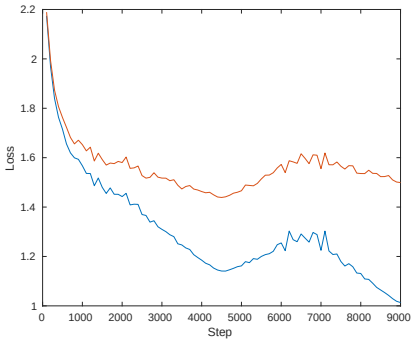
As suggested on the instructions I created a network with 2 hidden layers of 50 neurons each. Then I trained it on 45000 samples and validate it on 5000. The parameters I used were the one stated in the instructions : $eta_min = 1e-5$, $eta_max 1e-1$, $n_s = 5 \cdot 450$ $\lambda = 0.0001$ (but I guess thre was a typo, n_s

was supposed to be $n_s = 2 \cdot 450$).

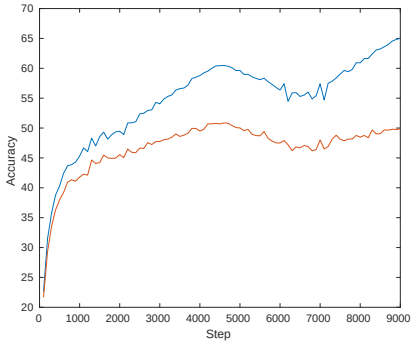
The loss and the accuracy throughout the 2 cycles of training are shown in Figure 1. The accuracy I got on the test data is 50.17%.



(a) Cost Fuction.



(b) Loss Fuction.

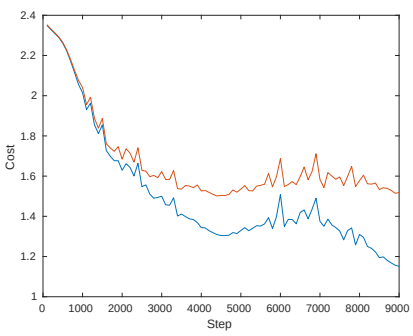


(c) Accuracy of the network (%).

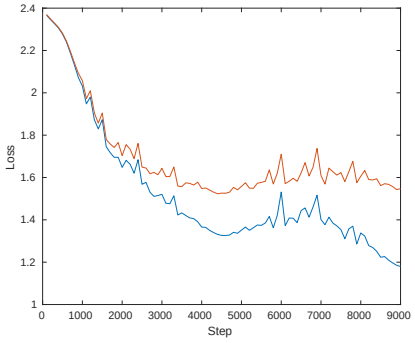
Figure 1: Training evolution of cost loss and accuracy on test and validation sets for the 2 layer network, with parameters $eta_min = 1e-5$, $eta_max 1e-1$, $n_s = 5 \cdot 450$ $\lambda = 0.0001$, run for 2 cycles.

Then I tested with the same hyperparameters the 9 layer network described by the sequence of hidden states $[50, 30, 20, 20, 10, 10, 10, 10]$. The performance I get is 48.75%. For completeness I also polot again in Figure 2 the cost, loss and accuracy.

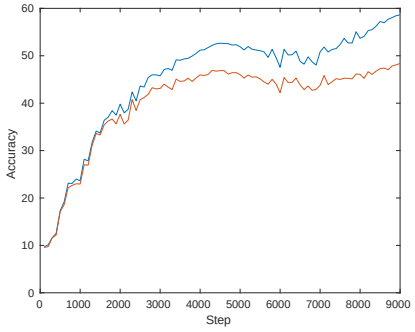
Test Multi-Layer



(a) Cost Fuction.



(b) Loss Fuction.



(c) Accuracy of the network (%).

Figure 2: Training evolution of cost loss and accuracy for the 9-layer network.