

# Assignment 2 - Report

Pietro Alovise

## Gradient Check

To check the correctness of the algorithm I computed the maximum of the absolute value of the difference of the weight matrices and the bias vector:

$$\max(|W_1 - W_{1 \text{ num}}|)$$

$$\max(|W_2 - W_{2 \text{ num}}|)$$

$$\max(|b_1 - b_{1 \text{ num}}|)$$

$$\max(|b_2 - b_{2 \text{ num}}|)$$

Where  $\cdot_{\text{num}}$  represents the numerical computed value. All these values were smaller than  $1e-6$  for different initialization and for different choices of  $\lambda$ . The numerical step used was  $h = 1e-5$ .

## Network Performance

First I tested the network using the parameters in the Assignment instruction. My results to replicate Figure 3 and 4 in the instructions are respectively represented in Figure 1 and 2

## $\lambda$ search

To find the best fitting value for the hyperparameter  $\lambda$  I first fixed the other hyper parameters as follows:  $\eta_{\min} = 1e-5$ ,  $\eta_{\max} = 1e-1$ ,  $n_s = 900$  and run for each lambda for 2 cycles.

I have done a preliminary search for  $\lambda$  in the range  $(10^{-8}, 10^{-1})$  for a total of 100 samples. The result is shown in Figure 3a.

Then with a second, more focused, search in the interval  $(10^{-4}, 10^{-2})$  etching 60 samples. I obtained the results shown in Figure 3b.

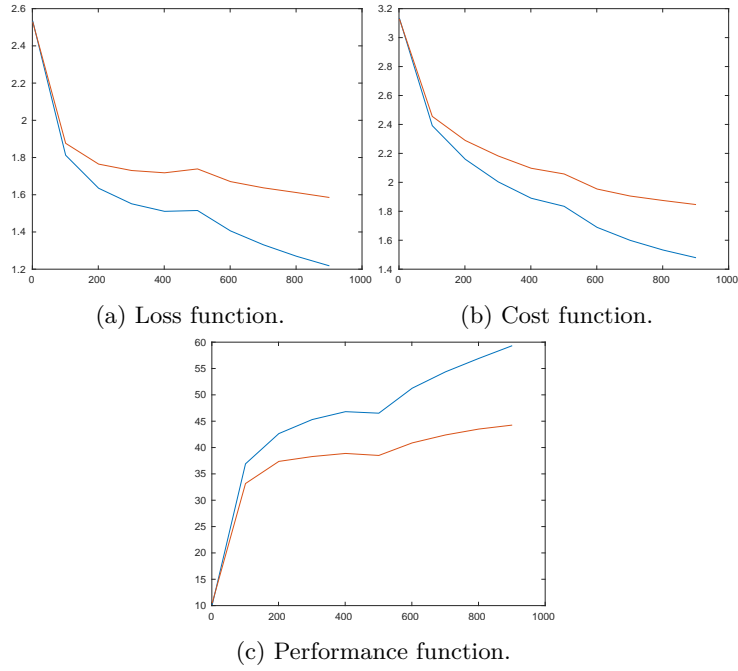


Figure 1: Representation of the cost, loss and performance of the network during training using as parameters  $n_s = 500$ ,  $\lambda = 0.01$  for 1 cycle.

In the table below are shown the results on the validation set for the best 3 values of  $\lambda$ .

$\lambda$	Performance on Validation
0.000305227	51.78%
0.000141724	51.70%
0.000928504	51.70%

## Full data with best $\lambda$

Now I can apply the best found  $\lambda$  to the whole dataset but 1000 samples for validation, and test the final result on the test data. The result is shown in the following table, and the loss and accuracy are shown in Figure 4.

$\lambda$	Performance on Test data
0.000305227	51.76%

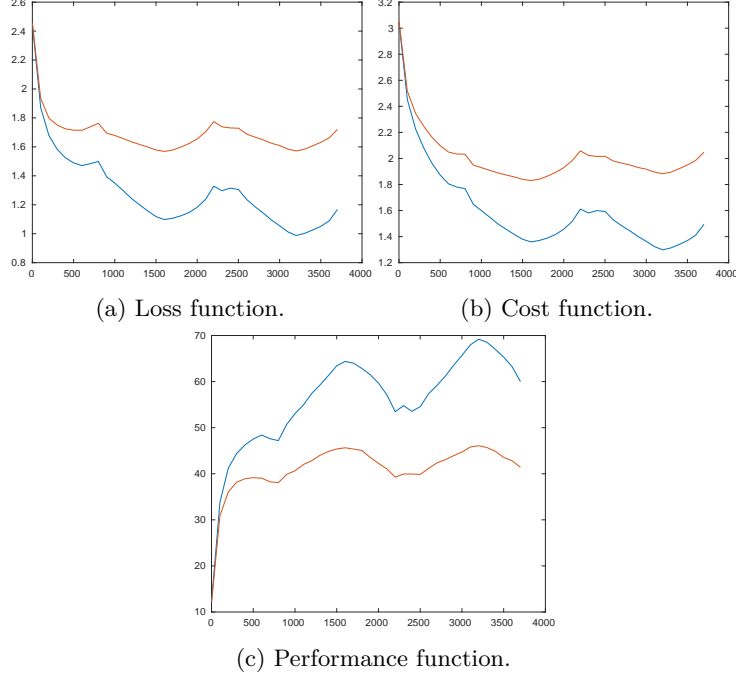


Figure 2: Representation of the cost, loss and performance of the network during training using as parameters  $n_s = 800$ ,  $\lambda = 0.01$  for 1 cycle.

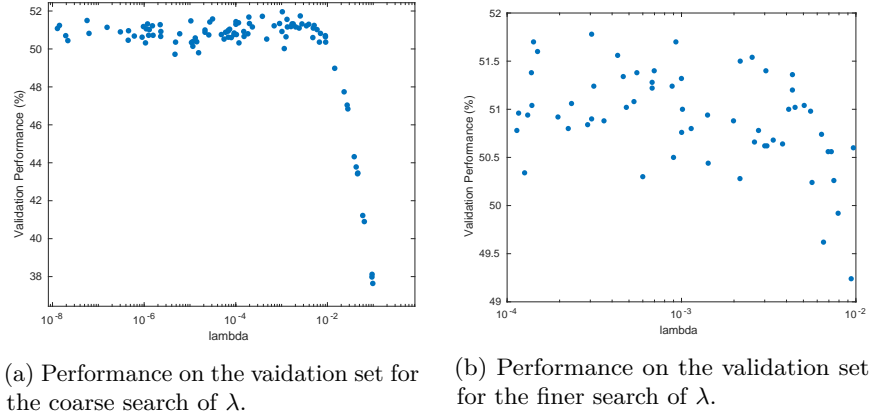


Figure 3: Result of search for the best hyperparameter  $\lambda$ .

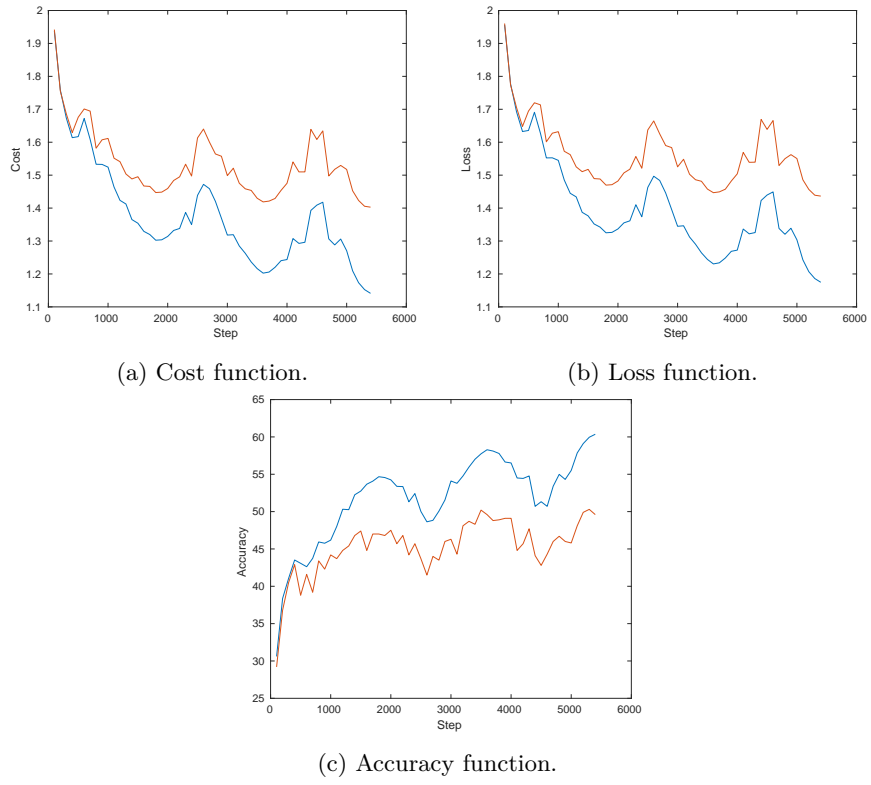


Figure 4: Tracked quantities during training with the best value of  $\lambda$ .