Assignment 2 - Report

Pietro Alovisi

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:

$$max(|W_1 - W_{1\ num}|)$$
$$max(|W_2 - W_{2\ num}|)$$
$$max(|b_1 - b_{1\ num}|)$$
$$max(|b_2 - b_{2\ num}|)$$

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

Network Performance

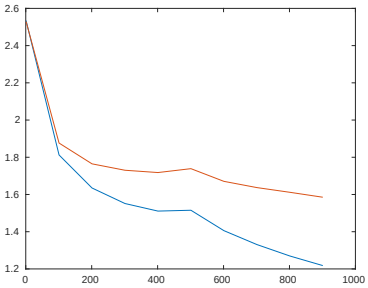
First I tested the network using the parameters in the Assignment instruction. My results to replicate Figure 3 and 4 in the instructions are respectively represented in Figure 1 and 2

λ search

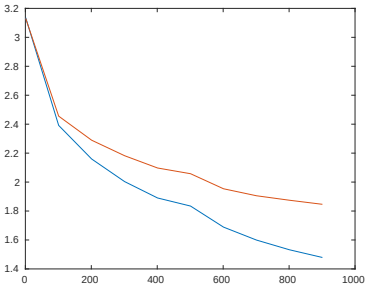
I have done a preliminary search for λ in the range $(10^{-5}, 10^{1})$ for a total of 30 samples. The result is shown in Figure 3.

Then with a second, more focused, search in the interval $(10^{-5}, 10^{1})$, I obtained the results shown in Figure 4.

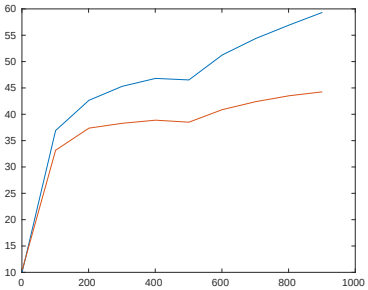
Full data with best λ



(a) Loss function.

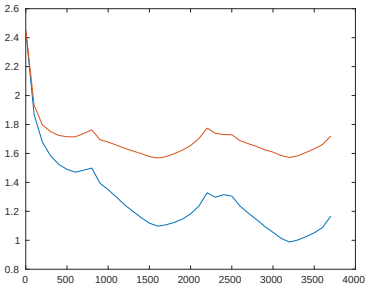


(b) Cost function.

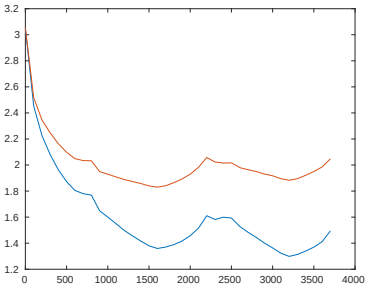


(c) Performance function.

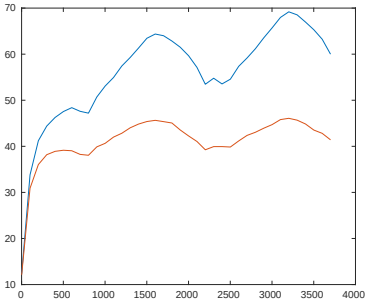
Figure 1: Representation of the cost, loss and performance of the network during training usign as parameters $n_s = 500$, $\lambda = 0.01$ for 1 cycle.



(a) Loss function.



(b) Cost function.



(c) Performance function.

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

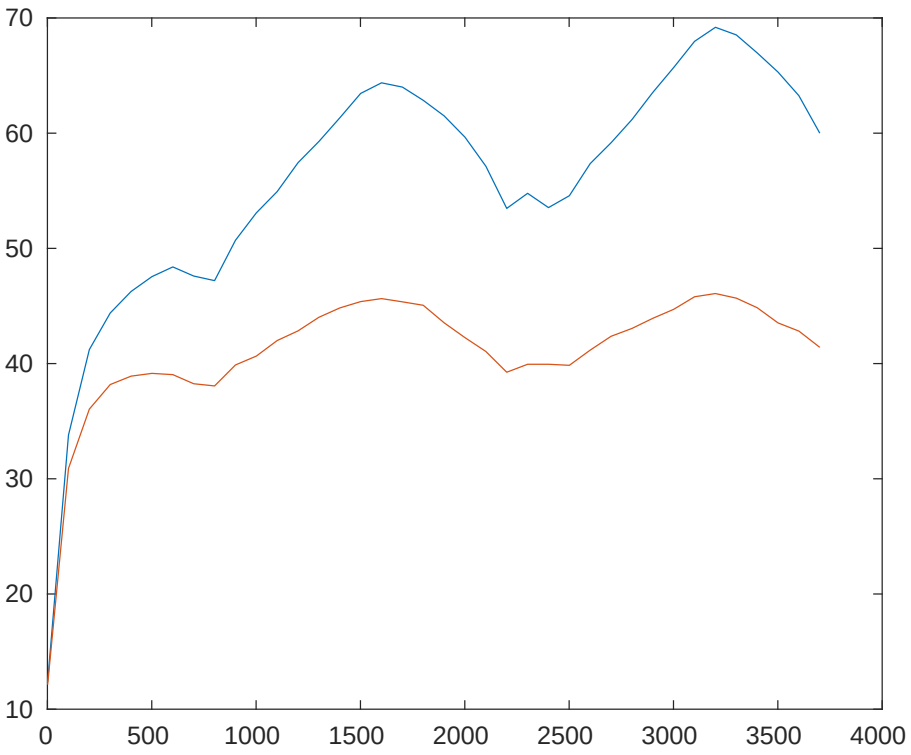


Figure 3: Results for different lambdas.

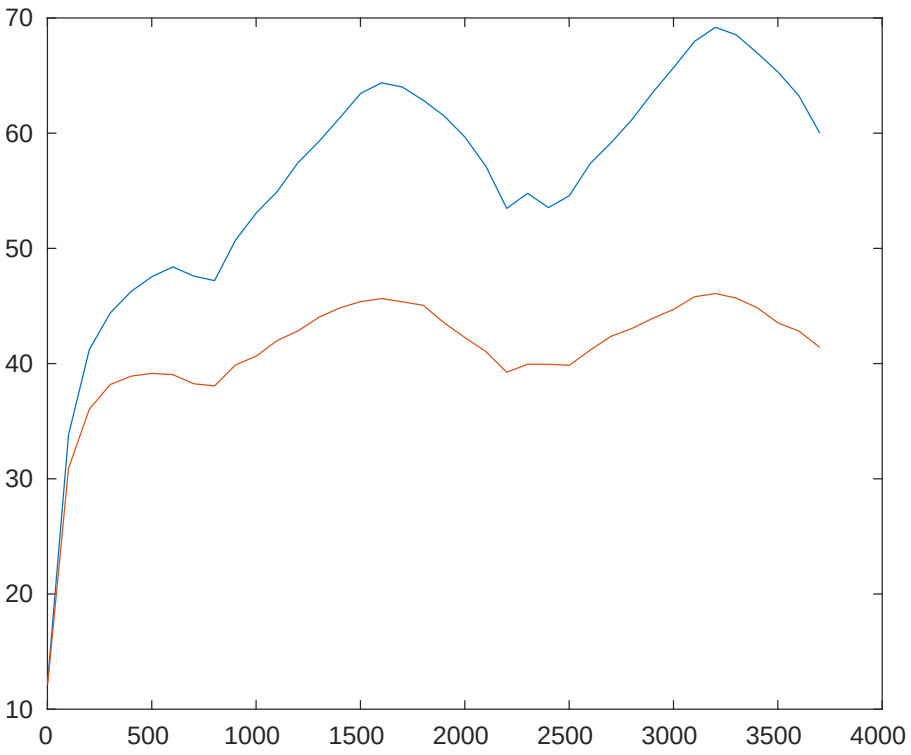


Figure 4: Results for different lambdas.