Practical assignment 1 feedback

You

July 17, 2018

Abstract

Your abstract.

1 Possible Changes

1. lecture 3 slide 36

   $$\theta^{(t+1)} = \theta^{(t)} - \eta_t \dot{(}gradL + l\theta_l)$$

   i think it should be

   $$\theta^{(t+1)} = \theta^{(t)} - \eta_t \dot{(}gradL + l\theta^{(t)})$$

2. lecture 3 slide 59 typo save in the middle

3. paper assignment 1 pdf page 2

   $$\delta_{N-1} = \frac{\partial \mathcal{L}}{\partial s_{N-1}} = \delta_N \cdot W_N \cdot \frac{\partial f(s1)}{\partial s_1} \tag{1}$$

   shouldn't be more precise to be:

   $$\delta_{N-1} = \frac{\partial \mathcal{L}}{\partial s_{N-1}} = \delta_N \cdot W_N \cdot \frac{\partial f_{N-1}(s1)}{\partial s_1} \tag{2}$$

   and all the others similarly.

2 Feedback

In general i lost a lot of time trying to figure out what happens with the Loss functions. I think the slides do not cover the topic well enough. There is this suggested site http://ufldl.stanford.edu/wiki/index.php/Softmax in the practical-1.ipynb but I found these 2 slides very very helpful and way more clarifying than the recommended site. They can be found in http://www.psi.toronto.edu/~jimmy/ece521/Lec9-nn2.pdf (slides 4-5)

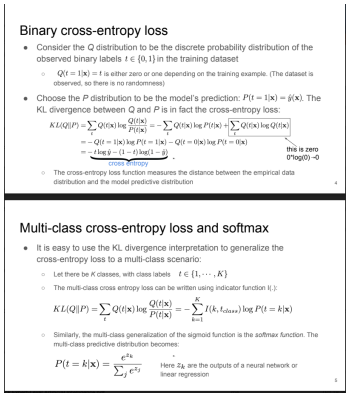

Figure 1:

In practical 1 i got really confused and lost quite some time to figure out what SoftMaxLoss is. I could not find this term on the internet. I ended up implementing a Softmax layer and a Multi class cross entropy loss, which i think is the expected answer. I do not understand why it is called SoftMaxLoss instead of Softmax Layer and Cross entropy Loss function. There should be at least some clarifying comments, in my opinion. My guess is that in the practical session some explanation was given.

3 Conclusion

Overall, i think the slides are very good and i have learned a lot from the assignment 1. I self studied with the given material and whatever i could find on the internet. I had no collaboration with any other student or guy that could help me and of course i did not attend any practical or theoretical lecture. Under these conditions it took me 3 weeks (i guess avg of 3-4 hours a day) to solve the whole practical 1 (programming and pen and paper), hopefully the time would reduce dramatically if i was attending the practical sessions or if i had collaboration with other students.