Peer Review of CA 7 – Assignment by Group 2, Review by Group 4 (Henrik)

Intro

Apologies for the tardiness in writing this peer review! We completely missed that the peer reviews were supposed to be written continuously as the course progressed until we got the email for CA7. Anyways, we are writing all our peer reviews now, so expect to receive a bunch.

Question a

Clear description of model dimensions and choice of loss function! However, some information is missing:

- Choice of regularizer.
- Regularization parameter.
- Learning rate.

In the text you say that: "The final accuracy which is obtained by the SGD optimizer is 0.68." but then in the table you have 0.95. I am not sure which one is accurate, but I will assume 95%, since 68% seems too low considering the size of your DNN.

I am also a bit surprised that the validation accuracy is so smooth in figure 1, in my experience the validation accuracy is always "choppier" which makes sense considering that the network is not directly trained on the validation dataset. The results in figure 2 looks much more like what I would expect! Do you have any clue as to why this is happening?

The results look good however! An accuracy of 97.5% is not shabby.

Question b

I believe you missed to report on the computational time and convergence speed here! That's a pretty important part of question b since the difference in computational time between pure SGD and mini-batch SGD is the main motivation for using mini-batches.

"As you can see in Figure 5, we have higher accuracy
at the end of the epochs in compared to previous methods. The final accuracy which
is obtained by this method is 0.975." – This is not true, you had exactly the same accuracy in the previous question.

Question c

"We note that the shallow model has many more trainable parameters at 318,010, while
the deeper model has less at 57,610." – Interesting, I like this comment! I guess that means it would be significantly cheaper to train the deep model over a network. (Lower comm cost / iteration)

Question d

Looks good to me! Just one comment surprised me:

“Then, regularization technique is recomended to solve this fact.”
This sentence makes it sound like you did not use a regularizer at all in the CA. I am surprised you can reach results such as ~98% validation accuracy without a regularizer! Never tested it myself but I definitely thought overfitting would be a huge problem if there is no regularization.

Question e

“Dropout can be preferable due to the decreased model size, e.g., for communications or storage or memory.”
I think dropout only temporarily removes nodes during training, they are still there for inference so the model size remains the same. At least in the original paper on dropout [1] they say “By dropping a unit out, we mean temporarily removing it from the network”

Question f

Interesting results! Nice to see that you at least managed to get some improvement both in convergence rate and final accuracy in Figure 11. First time I heard of label smoothing so it was interesting to read!

Summary

Nice report overall! All the exercises have been completed successfully. Some parts were a bit hard to understand as explained under each question, but the results look good.

[1] Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." *The journal of machine learning research* 15.1 (2014): 1929-1958.