Assignment Set VIII

Amir Mehrpanah 02/11/2020

Assignment 8 of
Deep Learning Course Given by
Prof. H. Haji Abol Hassan
written in LaTeX

Observations After Using Different Optimizers

We have trained models on four different optimizers: SGD, Adam, RMSprop and Adadelta. Here are some of the interesting observations made:

- Using different optimizer may change the result of training dramatically from Nan or constant loss to a good fit with fairly small loss. See Figure 1 for example. The SDG is Nan and Adadelta is constant until epoch 40!

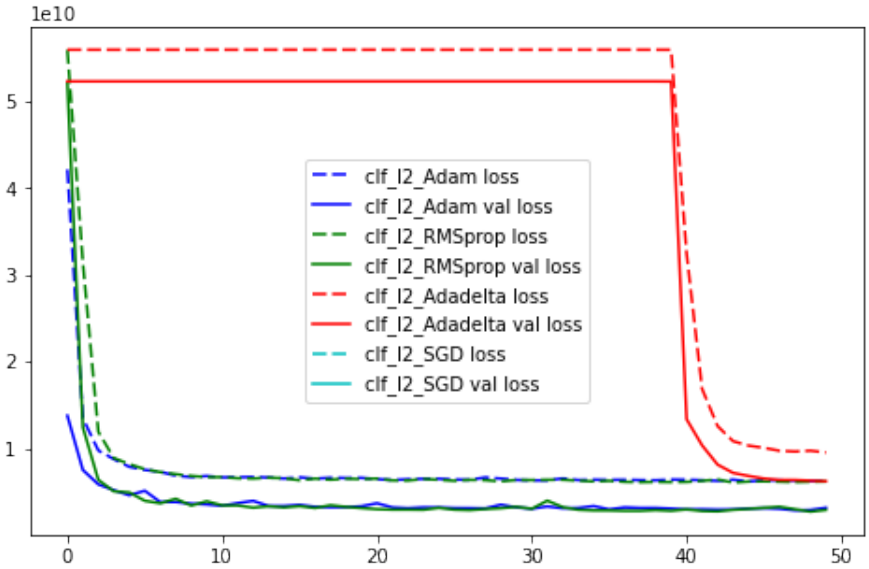

Figure 1: After the epoch 40!

- In some cases changing optimizer does not change the result. In Figure 3 and 2 we can see that all optimizers worked similarly and SGD has a slow rate of convergence which can be made faster by using a bigger learning rate.

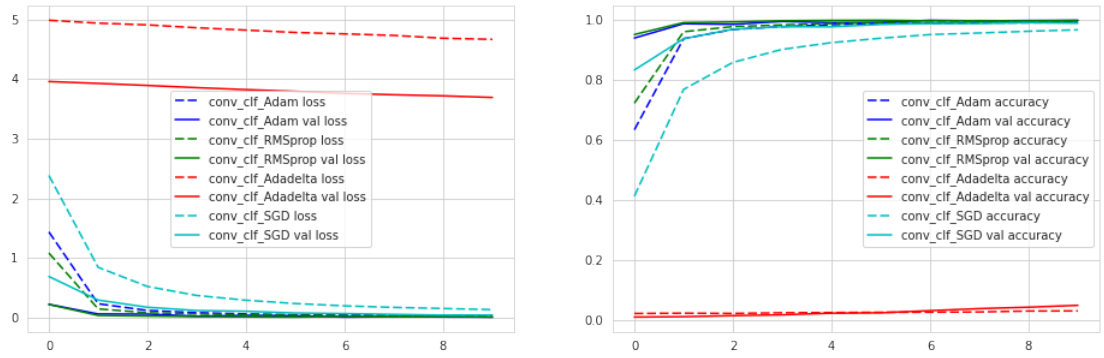

Figure 2: Different optimizer may give similar loss

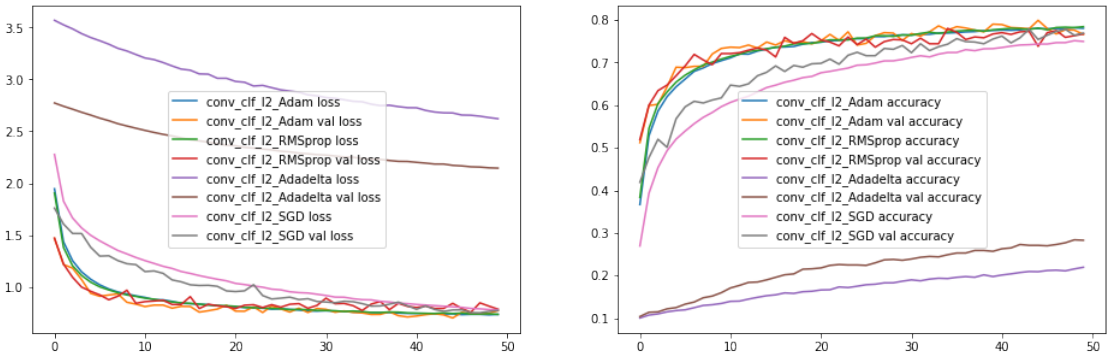

Figure 3: Different optimizer may give similar loss

- Optimization method can cause over fitting see Figure 4 and 5. SGD obtained the best validation loss at a higher risk of over fitting in both figures. In Figure 4 RMSprop and Adam obtained almost equal validation loss but Adam has smaller training loss which shows that the risk of over fitting is higher using Adam. Adadelta on the other hand is under fitting the data in Figure 4 and has not obtained as good results as RMSprop. Taking these risks into account, the best optimizer in these figures is RMSProp.

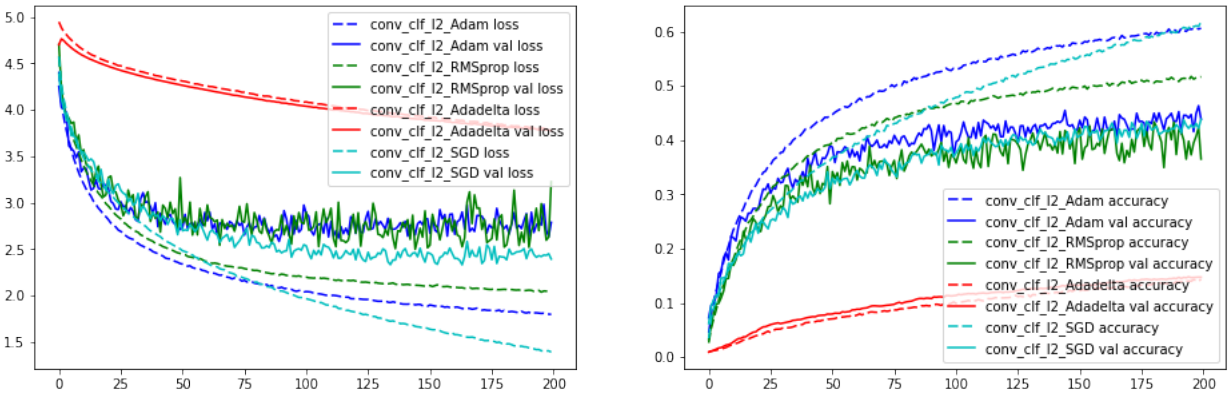

Figure 4: Different optimizer may cause over fitting or under fitting

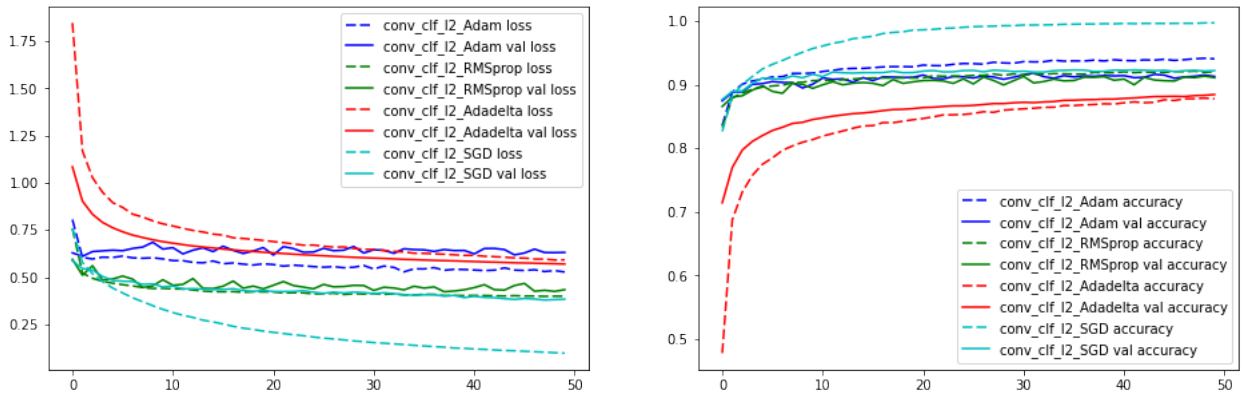

Figure 5: Different optimizer may cause over fitting or under fitting

- In Figure 6, RMSprop and SGD worked similar, however, There is something strange about SGD. It has decreased loss until epoch 10 then it started to go up later on! Adam and Adadelta worked very similar.

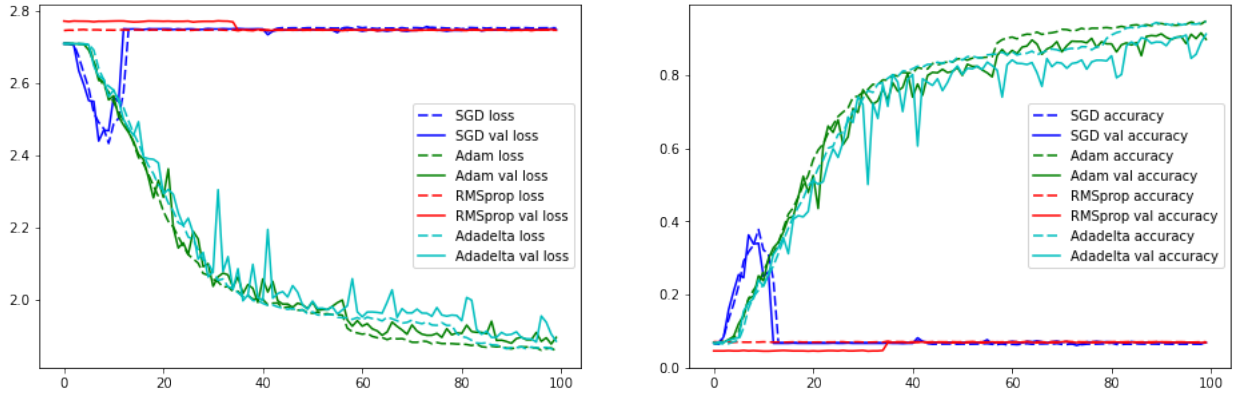

Figure 6: Spooky behaviour

- We have to omit SDG from Figure 7 because it has produced extremely large values, vanishing small changes in other losses, and has produced Nan values soon after the beginning epochs. Adadelta loss remained unchanged, Adam and RMSprop has produced very similar values.

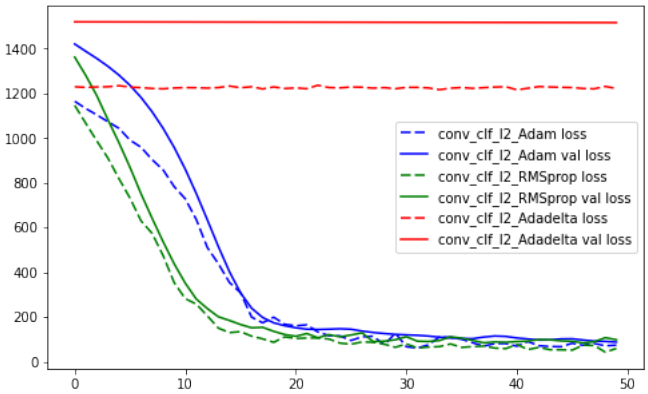

Figure 7: Extremely large loss values