

CA7 Review

from SG 1

Well done for the comparison between different hyperparameters and depth of the NN. Hope you could finish the rest parts (d)(e)(f)!

Some ideas:

- 1) LR=1 is too high for the task; and it's a bit weird while doing comparison with lower learning rate and higher learning rate. We think it might be the problem of setting "decay=0.99" in your SGD function, while we think $lr = lr * (1 - \text{decay})$, so you should set it as 0.01.
- 2) MSE may not be a good measurement for evaluating the classification performance when accuracy and loss are already used.
- 3) It will be better to show some successful and failed cases for the model.
- 4) When you increase the batch size to an extremely large value like 1000, you need to increase the number of training epochs, otherwise it is easy for the network to underfit.
- 5) When you do a comparison between the deep and shallow network for question c, if you track the loss and accuracy, you can easily notice that both two models are underfitting as well.