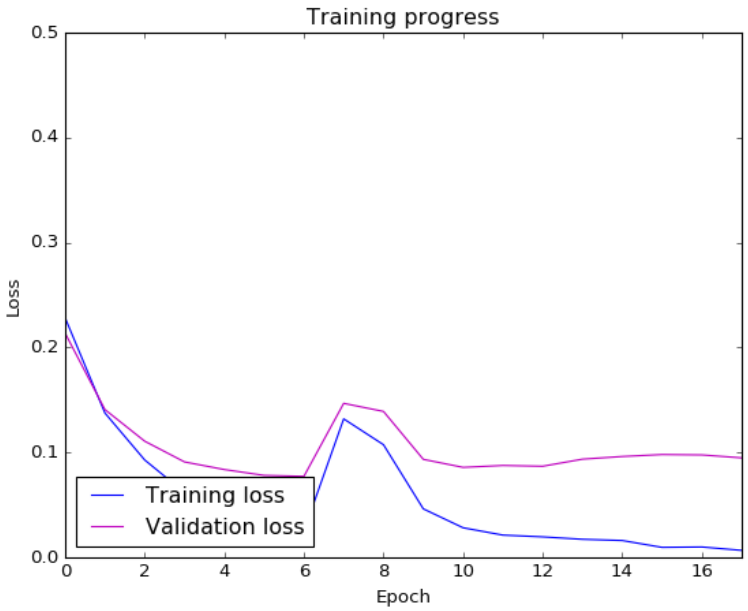
Deep Learning Lab

Assignment 1

Name: Aleksandar Petrov

Immatriculation number: 3743542

I have implemented the very basics so that the NN can be trained on the MNIST data set. I had trouble with my gradients but it was all just coding mistakes. I reached a validation error of ~ 1.8% but sadly my NN overfits really strongly. I am using 3 hidden layers with 300, 400, 10 neurons respectively (starting from input to output) the idea I took from some of the results on the MNIST web-site. Other settings: learning rate: 0.5, increased mini batches to 128. The first layer is a ReLU and the second one as well. After reading the DL book they suggest that ReLUs are sufficient to solve most problems. The output is of course a Softmax because our task is in the scope of multiclass classification. I've also implemented Adam but it throws an "overflow encountered in true_divide" error when I try to use it. I didn't have enough time to invest this problem, so I would be happy if you could answer that.



On the plot the performance of my NN can be seen. The final best achieved test error is 1.68%.

Questions:

1. Can you please provide me with as much feedback on my code as possible? I really want to get better ☺
2. Why is my Adam implementation throwing that error? How can I fix it?
3. Why does one comment in the code say to assume that X_train and Y_train are shuffled, when I couldn't find any explicit shuffling of the data?