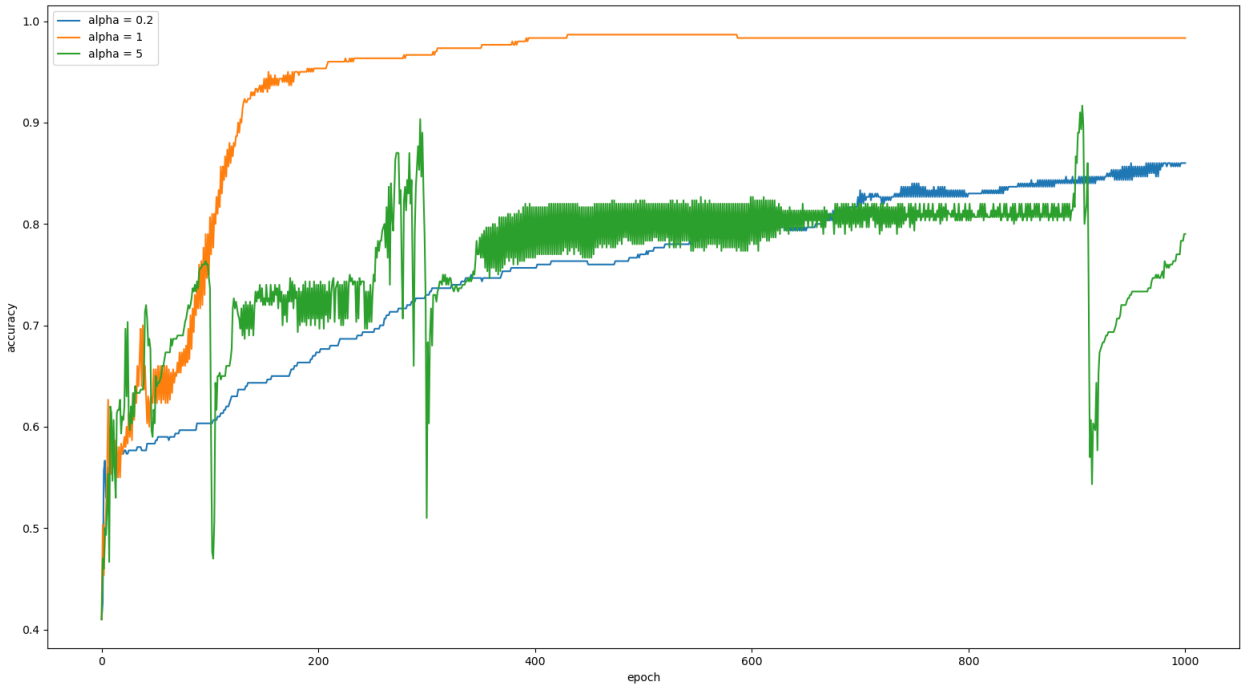
Report

Computer lab. Backpropagation

Morhunenko Mykola.

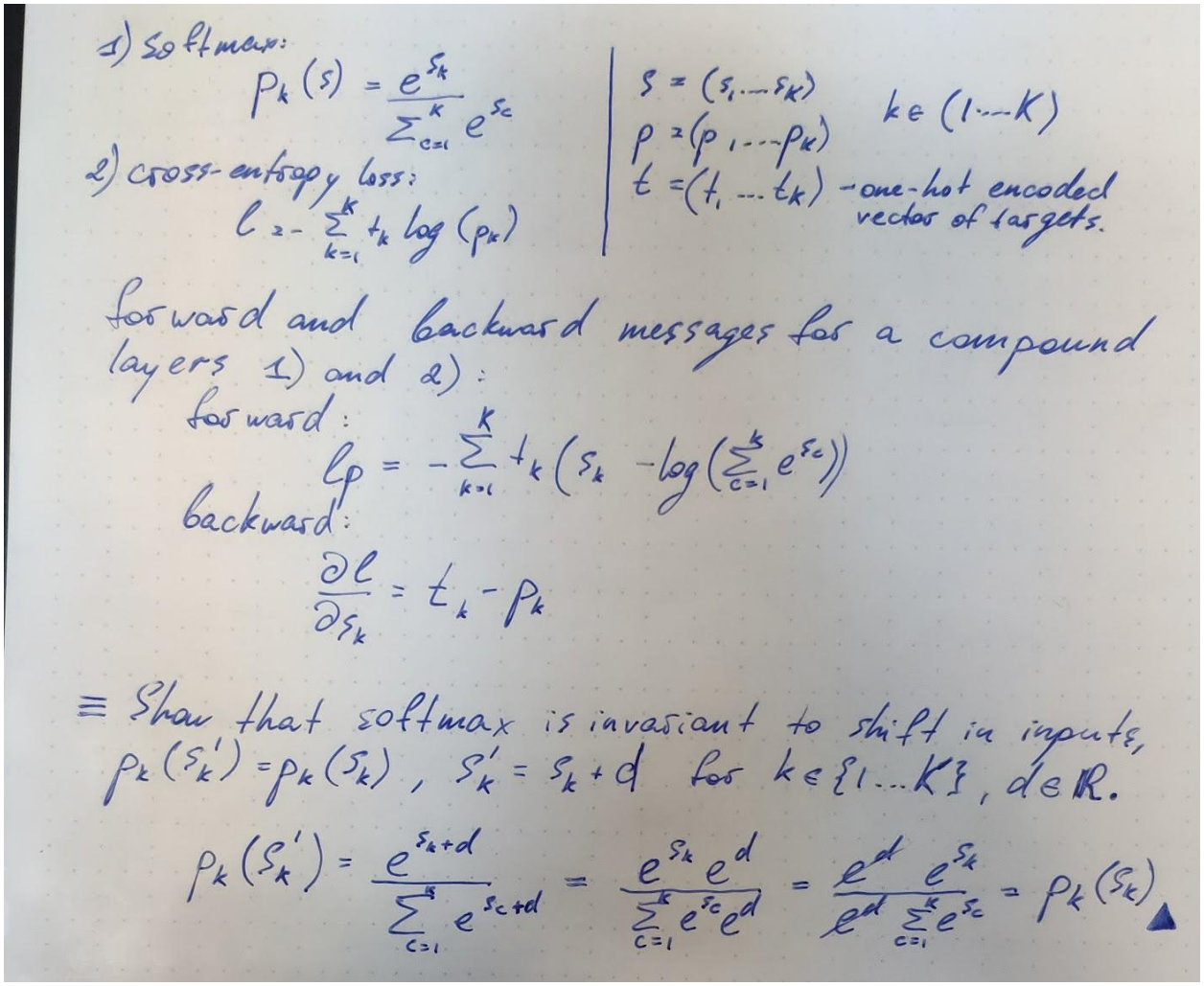
Assignment 1.

The convergence plots for three settings of the learning rate $\alpha \in \{0.2,\ 1,\ 5\}$ are:



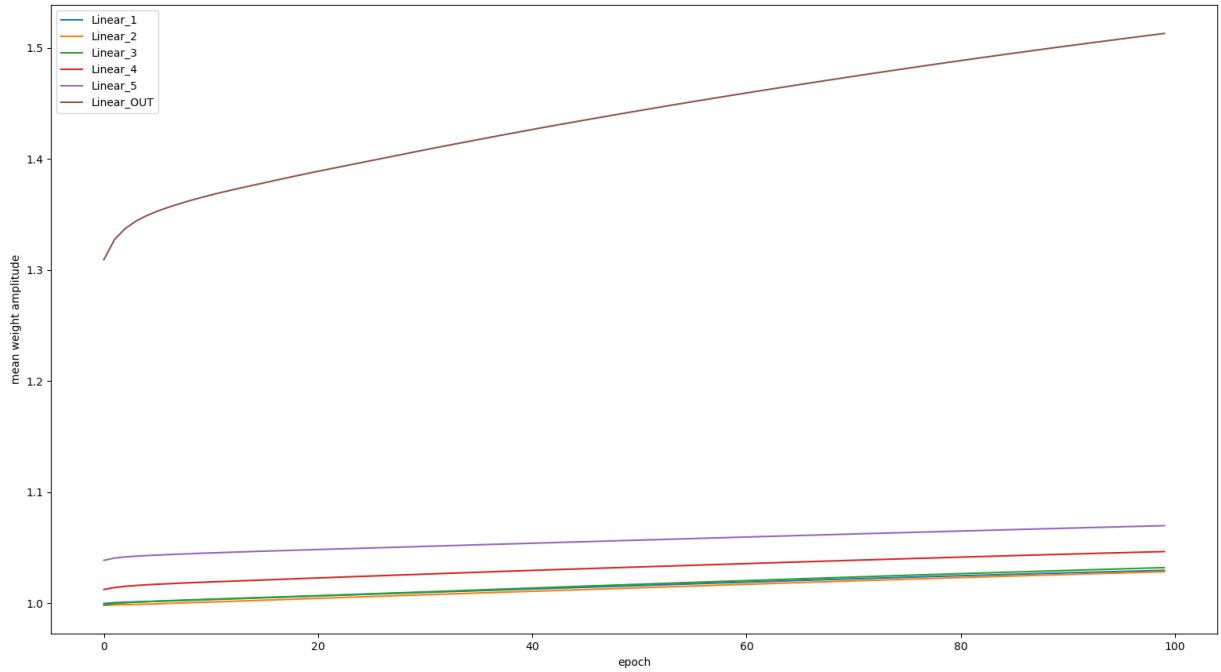
The result convergence improves with increasing the learning rate goes up to some point, and after that, it decreases and shows much worse results. It is because the learning rate is proportional to the chosen method parameter - Gradient Descent - and up to some point, decreasing of the parameter alpha improves the convergence, but after some point it will oscillate and show much worse results. The optimization function results are moving around the optimal value but can not reach it because of “overshooting”, which decreases the accuracy.

Assignment 2.



Assignment 3.

The plot shows the development of mean weight amplitude for each linear layer over the training epoch normalized w.r.t. The initial mean amplitude is shown below:



The weights are initialized randomly, so the training loss is higher for the first epochs. Because of the specifics of the Gradient Descent method, the weights amplitude firstly increases fast at the beginning and then slows down, because the training loss decreases.