LAB 2: Backpropagation

Andrii Zakharchenko

November 2019

1 Assignment 1

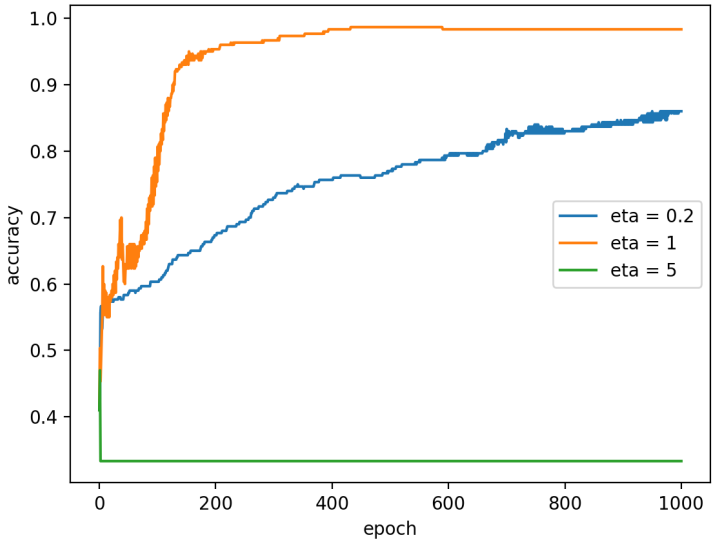

Figure 1: Comparison of test accuracy convergence for different learning rate

As we can see from the figure above, with learning rate $\eta = 1$ we achieved higher accuracy faster than for other learning rates. With every epoch we move towards some local minimum and learning rate specifies how fast we are moving there. So it is natural that we reached some local minimum faster with $\eta = 1$ rather than $\eta = 0.2$. But we can also see that for $\eta = 5$ we are stuck at very low accuracy. Since we update parameters on each minibatch it doesn't necessarily mean that with every update we will move in the right direction, and with big learning rate we change weights significantly and we basically oscillate around some minimum but always overshooting it.

2 Assignment 2

Show that $p(\mathbf{s}') = p(\mathbf{s})$, where $p'_k = p_k + c$, $c \in \mathbb{R}$ is some constant.

$$
\begin{aligned}
p(\mathbf{s}') &= \frac{e^{s_k+c}}{\sum_i^K e^{s_i+c}} = \frac{e^{s_k}e^c}{\sum_i^K e^{s_i}e^c} = \frac{e^{s_k}e^c}{(e^{s_1}e^c + e^{s_2}e^c + ... + e^{s_K}e^c)} \\
&= \frac{e^{s_k}e^c}{e^c(e^{s_1} + e^{s_2} + ... + e^{s_K})} = \frac{e^{s_k}}{(e^{s_1} + e^{s_2} + ... + e^{s_K})} = p(\mathbf{s})
\end{aligned} \tag{1}
$$

3 Assignment 3

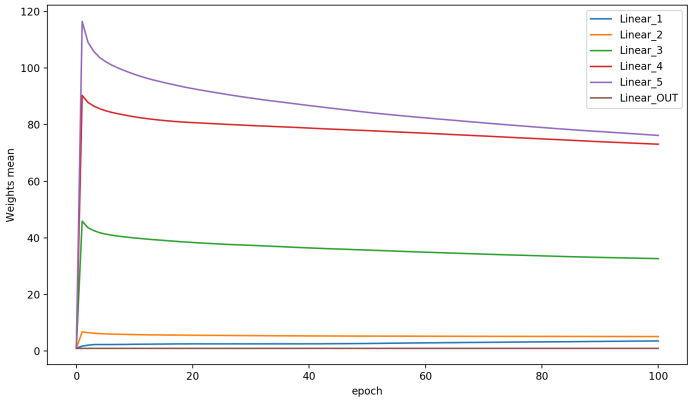

Figure 2: Mean absolute amplitude of weights for each layer (normalized)

We can see that in the first epoch weights in all layers increase their amplitude, this due to the fact that at the first epoch loss is high, meaning that local gradients are high, which in turn results in bigger changes in weights. In consecutive epochs we can see that amplitude of weights starts to decrease, since some weights may grow towards zero as these parameters are not important for the classification.