Report Assignment 1

Luc Weytingh

November 15, 2020

Abstract

insert abstracta

1 Evaluating the gradients

1.1a Linear Module

$$\begin{aligned}
\frac{\partial L}{\partial W} \Rightarrow \left[\frac{\partial L}{\partial W}\right]_{ij} &= \frac{\partial L}{\partial W_{ij}} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{\partial Y_{mn}}{\partial W_{ij}} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{\partial [XW^T + B]_{mn}}{\partial W_{ij}} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{\partial}{\partial W_{ij}} \left(\sum_p X_{mp} (W^T)_{pn} + B_{mn} \right) \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{\partial}{\partial W_{ij}} \left(\sum_p X_{mp} W_{np} + B_{mn} \right) \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \sum_p X_{mp} \frac{\partial W_{np}}{\partial W_{ij}} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \sum_p X_{mp} \delta_{ni} \delta_{pj} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} X_{mj} \delta_{ni} & &= \sum_{m} \frac{\partial L}{\partial Y_{mi}} X_{mj}
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial L}{\partial b} \Rightarrow \left[\frac{\partial L}{\partial b}\right]_i &= \frac{\partial L}{\partial b_i} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{\partial Y_{mn}}{\partial b_i} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{\partial [XW^T + B]_{mn}}{\partial b_i} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{\partial}{\partial b_i} \left(\sum_p X_{mp} W_{np} + B_{mn} \right) \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{\partial B_{mn}}{\partial b_i} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \delta_{ni} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mi}}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial X} \Rightarrow \left[\frac{\partial L}{\partial X}\right]_{ij} &= \frac{\partial L}{\partial X_{ij}} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{\partial Y_{mn}}{\partial X_{ij}} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{\partial [XW^T + B]_{mn}}{\partial X_{ij}} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{\partial}{\partial X_{ij}} \left(\sum_p X_{mp} W_{np} + B_{mn} \right) \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \sum_p \frac{\partial X_{mp}}{\partial X_{ij}} W_{np} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \sum_p \delta_{mi} \delta_{pj} W_{np} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \delta_{mi} W_{nj} \\
&= \sum_{n} \frac{\partial L}{\partial Y_{in}} W_{nj}
\end{aligned}
$$

1.1b Activation module

$$
\begin{aligned}
\frac{\partial L}{\partial X} \Rightarrow \left[\frac{\partial L}{\partial X}\right]_{ij} &= \frac{\partial L}{\partial X_{ij}} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{\partial Y_{mn}}{\partial X_{ij}} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{\partial h(X_{mn})}{\partial X_{ij}} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{\partial [h(X)]_{mn}}{\partial X_{ij}} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \delta_{mi} \delta_{nj} \frac{\partial h(X_{mn})}{\partial X_{ij}} \\
&= \sum_{m} \frac{\partial L}{\partial Y_{mj}} \delta_{mi} \frac{\partial h(X_{mj})}{\partial X_{ij}} \\
&= \frac{\partial L}{\partial Y_{ij}} \frac{\partial h(X_{ij})}{\partial X_{ij}}
\end{aligned}
$$

1.1c Softmax and Loss modules

$$
\begin{aligned}
\frac{\partial L}{\partial X} \Rightarrow \left[\frac{\partial L}{\partial X}\right]_{ij} &= \frac{\partial L}{\partial X_{ij}} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{\partial Y_{mn}}{\partial X_{ij}} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{\partial [Softmax(X)]_{mn}}{\partial X_{ij}} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{\partial}{\partial X_{ij}} \left(\frac{e^{X_{mn}}}{\sum_k e^{X_{mk}}} \right) \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{\delta_{im}\delta_{jn} e^{X_{mn}} \sum_k e^{X_{mk}} - \delta_{im} e^{X_{ij}} e^{X_{mn}}}{\left(\sum_k e^{X_{mk}}\right)^2} \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{e^{X_{mn}}}{\sum_k e^{X_{mk}}} \left(\frac{\delta_{im}\delta_{jn} e^{X_{mn}} \sum_k e^{X_{mk}}}{\sum_k e^{X_{mk}}} - \frac{\delta_{im} e^{X_{ij}} e^{X_{mn}}}{\sum_k e^{X_{mk}}} \right) \\
&= \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{e^{X_{mn}}}{\sum_k e^{X_{mk}}} \left(\delta_{im}\delta_{jn} - \frac{\delta_{im} e^{X_{ij}} e^{X_{mn}}}{\sum_k e^{X_{mk}}} \right) \\
&= \sum_{n} \frac{\partial L}{\partial Y_{in}} Y_{in} \left(\delta_{jn} - Y_{ij} \right)
\end{aligned}
$$

2 2 PyTorch MLP

2.1 2.1

The training en testing procedures yield a 0.48 accuracy with the default parameters. The loss and accuracy curves for this model can be found in Figure 1.

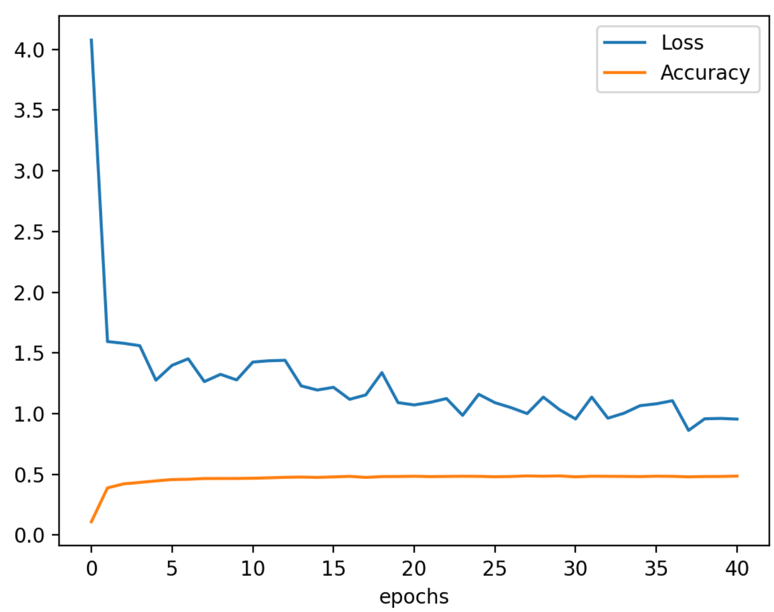

Figure 1: The loss and accuracy for a MLP with default parameters and 3 hidden layers of size 40, 50, and 40.