Hand-in 2

Michael Iversen
Student ID: 201505099

October 11, 2022

PART I: Derivative

When implementing the neural network, we apply the following loss function,

$$L(z) = -\sum_{i=1}^{k} y_i \ln(\mathrm{softmax}(z)_i).$$

Let $y_j = 1$ be the correct label and $y_i = 0$ for $i \neq j$, then the loss function can be rewritten as,

$$L(z) = -\ln(\mathrm{softmax}(z)_j).$$

In this section, we determine the derivative of the loss function with respect to all entries of z: $\frac{\partial L}{\partial z_i}$. We determine the derivative by direct calculation.

$$\frac{\partial L}{\partial z_i} = \frac{\partial}{\partial z_i}\left[-\ln\left(\frac{e^{z_j}}{\sum_{a=1}^{k} e^{z_a}}\right)\right]$$

Applying the chain rule and quotient rule, we find,

$$\begin{aligned}
&= -\frac{\sum_{a=1}^{k} e^{z_a}}{e^{z_j}} \cdot \frac{\delta_{ij} e^{z_j}\left(\sum_{a=1}^{k} e^{z_a}\right) - e^{z_j}\left(\sum_{a=1}^{k} \delta_{ia} e^{z_a}\right)}{\left(\sum_{a=1}^{k} e^{z_a}\right)^2} \\
&= -\frac{\sum_{a=1}^{k} e^{z_a}}{e^{z_j}} \cdot \frac{\delta_{ij} e^{z_j}\left(\sum_{a=1}^{k} e^{z_a}\right) - e^{z_j} e^{z_i}}{\left(\sum_{a=1}^{k} e^{z_a}\right)^2} \\
&= -\delta_{ij} + \frac{e^{z_i}}{\sum_{a=1}^{k} e^{z_a}} \\
&= -\delta_{ij} + \mathrm{softmax}(z_i).
\end{aligned}$$

In total, the derivative is given by the simple expression, $\frac{\partial L}{\partial z_i} = -\delta_{ij} + \mathrm{softmax}(z_i)$.