CS 489, Winter 2021: Assignment 3

Question 1: Gradient Descent and Error Backpropagation

For the following questions, let us define the cost function:

$$L(x,t) = \frac{t}{x} - \ln\frac{t}{x} - 1 \tag{1}$$

and the activation function:

$$\rho(z) = z^2 \tag{2}$$

a) We want to find $\frac{\partial L}{\partial z_6}$. Let us first find the $\frac{\partial L}{\partial x}$:

$$\begin{aligned}\frac{\partial L}{\partial x} &= \frac{\partial}{\partial x}\left(\frac{t}{x} - \ln\frac{t}{x} - 1\right) \qquad \text{from (1)}\\ &= -\frac{t}{x^2} + \frac{1}{x}\\ &= \frac{x-t}{x^2}\end{aligned}$$

So, for x_6, we have:

$$\frac{\partial L}{\partial x_6} = \frac{x_6 - t}{x_6^2} \tag{3}$$

Next, if $x_6 = \rho(z_6)$, then by (2), $x_6 = z_6^2$. This means that:

$$\frac{\partial x_6}{\partial z_6} = \frac{\partial}{\partial z_6} z_6^2 = 2z_6 \tag{4}$$

Now, combining (3) and (4) together via chain rule:

$$\begin{aligned}\frac{\partial L}{\partial z_6} &= \frac{\partial L}{\partial x_6}\frac{\partial x_6}{\partial z_6}\\ &= \frac{x_6 - t}{x_6^2} 2z_6\\ &= \frac{z_6^2 - t}{(z_6^2)^2} 2z_6\\ &= \frac{2(z_6^2 - t)}{z_6^3}\end{aligned}$$

So, we conclude that:

$$\frac{\partial L}{\partial z_6} = \frac{2(z_6^2 - t)}{z_6^3} \tag{5}$$

b) We want to determine the update for w_{46}. We first need to obtain $\frac{\partial L}{\partial w_{46}}$.

$$\begin{aligned}
\frac{\partial L}{\partial w_{46}} &= \frac{\partial L}{\partial z_6} \times \frac{\partial z_6}{\partial w_{46}} && \text{by chain rule} \\
&= \frac{\partial L}{\partial z_6} \times \frac{\partial}{\partial w_{46}} \left(w_{36}x_3 + w_{46}x_4 w_{56}x_5 + b_6 \right) && \text{by the given neural network} \\
&= \frac{\partial L}{\partial z_6} \times x_4 \\
&= \frac{\partial L}{\partial z_6} \times \sigma(z_4) \\
&= \frac{2(z_6^2 - t)}{z_6^3} \sigma(z_4) && \text{using (5) from q1a)}
\end{aligned}$$

Using this, we can define an update for w_{46}:

$$\begin{aligned}
w_{46} &= w_{46} - k \left(\frac{\partial L}{\partial w_{46}} \right) \\
&= w_{46} - k \left(\frac{2(z_6^2 - t)}{z_6^3} \sigma(z_4) \right)
\end{aligned}$$

for some positive constant k.

c) We want to define an update fomr b_6. We first need to obtain $\frac{\partial L}{\partial b_6}$:

$$\begin{aligned}
\frac{\partial L}{\partial b_6} &= \frac{\partial L}{\partial z_6} \times \frac{\partial z_6}{\partial b_6} && \text{by chain rule} \\
&= \frac{\partial L}{\partial z_6} \times \frac{\partial}{\partial b_6} \left(w_{36}x_3 + w_{46}x_4 w_{56}x_5 + b_6 \right) && \text{by the given neural network} \\
&= \frac{\partial L}{\partial z_6} \times 1 \\
&= \frac{2(z_6^2 - t)}{z_6^3} && \text{using (5) from q1a)}
\end{aligned}$$

Using this, we can define an update for b_6:

$$\begin{aligned}
b_6 &= b_6 - k \left(\frac{\partial L}{\partial b_6} \right) \\
&= b_6 - k \frac{2(z_6^2 - t)}{z_6^3}
\end{aligned}$$

for some positive constant k.

d) We want to obtain $\frac{\partial L}{\partial z_3}$. Using chain rule:

$$\begin{aligned}
\frac{\partial L}{\partial z_3} &= \frac{\partial L}{\partial x_3} \times \frac{\partial x_3}{\partial z_3} && \text{by chain rule} \\
&= \left[\frac{\partial z_6}{\partial x_3}\frac{\partial L}{\partial z_6} + \frac{\partial z_5}{\partial x_3}\frac{\partial L}{\partial z_5} + \frac{\partial z_4}{\partial x_3}\frac{\partial L}{\partial z_4}\right] \times \frac{\partial x_3}{\partial z_3} \\
&= \left[\frac{\partial z_6}{\partial x_3}\frac{\partial L}{\partial z_6} + \frac{\partial z_5}{\partial x_3}\frac{\partial L}{\partial x_5}\frac{\partial x_5}{\partial z_5} + \frac{\partial z_4}{\partial x_3}\frac{\partial L}{\partial x_4}\frac{\partial x_4}{\partial z_4}\right] \times \frac{\partial x_3}{\partial z_3} \\
&= \left[\frac{\partial z_6}{\partial x_3}\frac{\partial L}{\partial z_6} + \frac{\partial z_5}{\partial x_3}\left(\frac{\partial z_6}{\partial x_5}\frac{\partial L}{\partial z_6}\right)\frac{\partial x_5}{\partial z_5} + \frac{\partial z_4}{\partial x_3}\left(\frac{\partial z_6}{\partial x_4}\frac{\partial L}{\partial z_6}\right)\frac{\partial x_4}{\partial z_4}\right] \times \frac{\partial x_3}{\partial z_3} \\
&= \frac{\partial L}{\partial z_6} \times \left[\frac{\partial z_6}{\partial x_3} + \frac{\partial z_5}{\partial x_3}\frac{\partial z_6}{\partial x_5}\frac{\partial x_5}{\partial z_5} + \frac{\partial z_4}{\partial x_3}\frac{\partial z_6}{\partial x_4}\frac{\partial x_4}{\partial z_4}\right] \times \frac{\partial x_3}{\partial z_3} \\
&= \frac{\partial L}{\partial z_6} \times \left[w_{36} + w_{35}w_{56}\frac{\partial x_5}{\partial z_5} + w_{34}w_{46}\frac{\partial x_4}{\partial z_4}\right] \times \frac{\partial x_3}{\partial z_3} && \text{as } \frac{\partial z_i}{\partial x_j} = w_{ij} \\
&= \frac{\partial L}{\partial z_6} \times \left[w_{36} + w_{35}w_{56}\frac{\partial x_5}{\partial z_5} + w_{34}w_{46}\frac{\partial x_4}{\partial z_4}\right] \times \frac{\partial x_3}{\partial z_3} && \text{as } \frac{\partial z_i}{\partial x_j} = w_{ij} \\
&= \frac{\partial L}{\partial z_6} \times \left[w_{36} + w_{35}w_{56}\frac{\partial}{\partial z_5}\sigma(z_5) + w_{34}w_{46}\frac{\partial}{\partial z_4}\sigma(z_4)\right] \times \frac{\partial}{\partial z_3}\sigma(z_3) && \text{by question def.}
\end{aligned}$$

Now since we know that $\frac{\partial \sigma(x)}{x} = \sigma(x)(1-\sigma(x))$ from exercise 2:

$$\begin{aligned}
\frac{\partial L}{\partial z_3} &= \frac{\partial L}{\partial z_6} \times [w_{36} + w_{35}w_{56}\sigma(z_5)(1-\sigma(z_5)) + w_{34}w_{46}\sigma(z_4)(1-\sigma(z_4))] \times \sigma(z_3)(1-\sigma(z_3)) \\
&= \frac{\partial L}{\partial z_6} \times \left[w_{36} + \sum_{i=4}^{5} w_{3i}w_{i6}\sigma(z_i)(1-\sigma(z_i))\right] \times \sigma(z_3)(1-\sigma(z_3))
\end{aligned}$$

Finally, sub in from (5) our definition of $\frac{\partial L}{\partial z_3}$, and we get:

$$\frac{\partial L}{\partial z_3} = \frac{2(z_6^2 - t)}{z_6^3} \times \left[w_{36} + \sum_{i=4}^{5} w_{3i}w_{i6}\sigma(z_i)(1-\sigma(z_i))\right] \times \sigma(z_3)(1-\sigma(z_3))$$