CS-E4850 Computer Vision
Exercise Round #9
Submitted by Chen Xu, ID 000000
2024-10-17

Exercise 1. Neural networks and backpropagation.

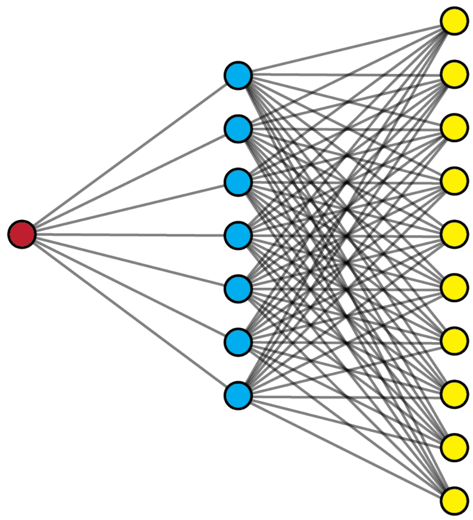

Figure 1: Schematics of a neural network

1)

$$
\begin{aligned}
E &= \frac{1}{m}\sum_{j=1}^{m} -\mathbf{t}_j \cdot log(\mathbf{y}_j) \\
&\overset{\text{m=1}}{=} -\mathbf{t}_1 \cdot log(\mathbf{y}_1) \qquad (1)
\end{aligned}
$$

2)

Proof.

$$y_i^{(2)} = \frac{e^{z_i^{(2)}}}{\sum_k e^{z_k^{(2)}}} \tag{2}$$

$$\begin{aligned} \frac{\partial y_i^{(2)}}{\partial z_i^{(2)}} &= \frac{e^{z_i^{(2)}}}{\sum_k e^{z_k^{(2)}}} - e^{z_i^{(2)}} \frac{e^{z_i^{(2)}}}{(\sum_k e^{z_k^{(2)}})^2} \\ &= y_i^{(2)} - (y_i^{(2)})^2 \end{aligned} \tag{3}$$

From equation (**??**) one can get:

$$\frac{\partial E}{\partial y_i^{(2)}} = \begin{cases} -\frac{1}{y_i^{(2)}} & \text{when } t_i = 1 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

From equations (**??**) and (**??**) one can get:

$$\begin{aligned} \frac{\partial E}{\partial z_i^{(2)}} &= \frac{\partial E}{\partial y_i^{(2)}} \frac{\partial y_i^{(2)}}{\partial z_i^{(2)}} \\ &= \begin{cases} (y_i^{(2)} - 1) & \text{when } t_i = 1 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{5}$$

$$\begin{aligned} \Rightarrow \frac{\partial E}{\partial \mathbf{z}^{(2)}} &= \frac{\partial E}{\partial \mathbf{y}^{(2)}} \frac{\partial \mathbf{y}^{(2)}}{\partial \mathbf{z}^{(2)}} \\ &= (\mathbf{y}^{(2)} - \mathbf{t})^\top \end{aligned} \tag{6}$$

□

3)

Proof.

From the chain rule and equation (**??**) one can get:

$$\begin{aligned} \frac{\partial E}{\partial \mathbf{y}^{(1)}} &= \frac{\partial E}{\partial \mathbf{z}^{(2)}} \frac{\partial \mathbf{z}^{(2)}}{\partial \mathbf{y}^{(1)}} \\ &= \frac{\partial E}{\partial \mathbf{z}^{(2)}} \mathbf{W}^{(2)} \\ &= (\mathbf{y}^{(2)} - \mathbf{t})^\top \mathbf{W}^{(2)} \end{aligned} \tag{7}$$

□

4)

Proof.

From the chain rule and equation (**??**) one can get:

$$\begin{aligned}
\frac{\partial E}{\partial w_{uv}^{(2)}} &= \frac{\partial E}{\partial \mathbf{z}^{(2)}} \frac{\partial \mathbf{z}^{(2)}}{\partial w_{uv}^{(2)}} \\
&= \frac{\partial E}{\partial z_u^{(2)}} y_v^{(1)} \\
&= (y_u^{(2)} - t_u) y_v^{(1)} \\
\Rightarrow \frac{\partial E}{\partial \mathbf{W}^{(2)}} &= (\mathbf{y}^{(2)} - \mathbf{t}) \mathbf{y}^{(1)\top}
\end{aligned}$$

□

5)

Proof.

$$\begin{aligned}
\because \mathbf{y}^{(1)} &= \sigma(\mathbf{z}^{(1)}) \qquad (8) \\
\therefore y_i^{(1)} &= \sigma(z_i^{(1)}) \\
&= \frac{1}{1 + e^{z_i^{(1)}}} \qquad (9) \\
\Rightarrow \frac{\partial y_i^{(1)}}{\partial z_i^{(1)}} &= \frac{e^{z_i^{(1)}}}{(1 + e^{z_i^{(1)}})^2} \\
&= \frac{1}{1 + e^{z_i^{(1)}}} - \frac{1}{(1 + e^{z_i^{(1)}})^2} \\
&= y_i^{(1)} - (y_i^{(1)})^2 \\
&= y_i^{(1)}(1 - y_i^{(1)}) \qquad (10)
\end{aligned}$$

Meanwhile

$$\frac{\partial y_i^{(1)}}{\partial z_j^{(1)}} = 0 \text{ When } i \neq j \qquad (11)$$

After some re-arrangement

$$\frac{\partial \mathbf{y}^{(1)}}{\partial \mathbf{z}^{(1)}} = diag(\mathbf{y}^{(1)} . * (\mathbf{1} - \mathbf{y}^{(1)})) \qquad (12)$$

□

6)

Proof.

From equations (**??**) and (**??**), one can get:

$$\begin{aligned}\frac{\partial E}{\partial \mathbf{z}^{(1)}} &= \frac{\partial E}{\partial \mathbf{y}^{(1)}}\frac{\partial \mathbf{y}^{(1)}}{\partial \mathbf{z}^{(1)}}\\ &= (\mathbf{y}^{(2)} - \mathbf{t})^\top \mathbf{W}^{(2)} diag(\mathbf{y}^{(1)} .* (\mathbf{1} - \mathbf{y}^{(1)}))\end{aligned}$$

□

7)

Proof.

From the chain rule, one can get:

$$\begin{aligned}\frac{\partial E}{\partial w_{uv}^{(1)}} &= \frac{\partial E}{\partial \mathbf{z}^{(1)}}\frac{\partial \mathbf{z}^{(1)}}{\partial w_{uv}^{(1)}}\\ &= \frac{\partial E}{\partial z_u^{(1)}} y_v^{(0)}\\ &= \frac{\partial E}{\partial z_u^{(1)}} x_v\end{aligned}$$

Hence:

$$\frac{\partial E}{\partial \mathbf{W}^{(1)}} = (\frac{\partial E}{\partial \mathbf{z}^{(1)}})^\top \mathbf{x}^\top$$

□