CS-E4950 Computer Vision
Exercise Round 9

Rui Qu
rui.qu@aalto.fi

November 14, 2019

Exercise 1 Solution

Schematic Picture of the Network[1]

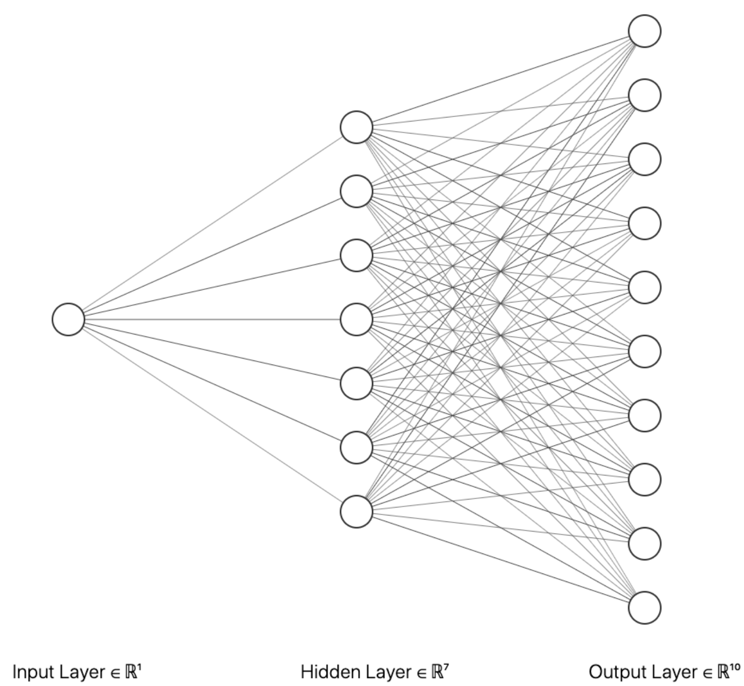

Figure 1: Fully Connected Neural Network with 1 input

[1] Picture generated via NN-architecture schematics (alexlenail.me/NN-SVG/)

1.

Given m=1:

$$E = \frac{1}{m}\sum_{j=1}^{m} -t_j \cdot log(y_j)$$
$$E = -t \cdot log(\mathbf{y}) = -t \cdot log(\sigma(\mathbf{W}x)) = -t \cdot log\frac{1}{1+e^{-\mathbf{W}x}} \tag{1}$$

Partial derivates of E with respect to t and x:

$$\begin{aligned}
\frac{\partial E}{\partial t} &= -log\frac{1}{1+e^{-\mathbf{W}x}} \\
\frac{\partial E}{\partial x} &= -\frac{t}{\sigma(\mathbf{W}x)} \cdot \frac{\partial \sigma(\mathbf{W}x)}{\partial x} \\
&= -t(1+e^{-\mathbf{W}x}) \cdot \mathbf{W}\frac{1}{1+e^{-\mathbf{W}x}}(1-\frac{1}{1+e^{-\mathbf{W}x}}) \\
&= -t\mathbf{W}\frac{e^{-\mathbf{W}x}}{1+e^{-\mathbf{W}x}}
\end{aligned} \tag{2}$$

2.

Utilise the chain rule:

$$\begin{aligned}
\frac{\partial E}{\partial z_i^{(2)}} &= \sum_{j=1}^{n} \frac{\partial E_j}{\partial y_j^{(2)}} \frac{\partial y_j^{(2)}}{\partial z_i^{(2)}} \\
\frac{\partial E_j}{\partial y_j^{(2)}} &= \frac{\partial(-t_j \cdot log y_j)}{\partial z_i^{(2)}} = -\frac{t_j}{y_i}
\end{aligned} \tag{3}$$

If $i \neq j$:

$$\frac{\partial y_j^{(2)}}{\partial z_i^{(2)}} = \frac{\partial \sigma(z_j^{(2)})}{\partial z_i^{(2)}} = -y_i y_j \tag{4}$$

If $i = j$:

$$\frac{\partial y_j^{(2)}}{\partial z_i^{(2)}} = \frac{\partial \sigma(z_i^{(2)})}{\partial z_i^{(2)}} = y_i(1-y_i) \tag{5}$$

Derived from Equation (3) (4) (5) and $\sum t_j = 1$:

$$
\begin{aligned}
\frac{\partial E}{\partial z_i^{(2)}} &= \sum_{j=1}^{n} \frac{\partial E_j}{\partial y_j^{(2)}} \frac{\partial y_j^{(2)}}{\partial z_i^{(2)}} \\
&= \sum_{i \neq j} \frac{\partial E_j}{\partial y_j^{(2)}} \frac{\partial y_j^{(2)}}{\partial z_i^{(2)}} + \sum_{i=j} \frac{\partial E_j}{\partial y_j^{(2)}} \frac{\partial y_j^{(2)}}{\partial z_i^{(2)}} \\
&= \sum_{i \neq j} [-\frac{t_j}{y_i}(-y_i y_j)] - \frac{t_j}{y_i} y_i (1 - y_i) \\
&= \sum_{i \neq j} [(t_j y_i)] + t_i y_i - t_i \\
&= y_i - t_i
\end{aligned} \tag{6}
$$

Hence:

$$
\frac{\partial E}{\partial \mathbf{z}^{(2)}} = (\mathbf{y}^{(2)} - \mathbf{t})^\top \tag{7}
$$

3.

$$
\begin{aligned}
\frac{\partial E}{\partial \mathbf{z}^{(2)}} &= (\mathbf{y}^{(2)} - \mathbf{t})^\top \\
\frac{\partial E}{\partial \mathbf{y}^{(1)}} &= \frac{\partial E}{\partial z(2)} \cdot \frac{\partial z(2)}{\partial y(1)} = (\mathbf{y}^{(2)} - \mathbf{t})^\top \cdot \mathbf{W}^{(2)}
\end{aligned} \tag{8}
$$

4.

$$
\begin{aligned}
\frac{\partial E}{\partial W_{uv}^{(2)}} &= \frac{\partial E}{\partial \mathbf{z}^{(2)}} \frac{\partial \mathbf{z}^{(2)}}{\partial W_{uv}^{(2)}} \\
&= (\frac{\partial E}{\partial \mathbf{z}^{(2)}})_u y_v^{(1)} \\
&= (y_u^{(2)} - t_u) y_v^{(1)} \\
i.e. \frac{\partial E}{\partial \mathbf{W}^{(2)}} &= [(\mathbf{y}^{(2)} - \mathbf{t}) \mathbf{y}^{(1)}]^\top
\end{aligned} \tag{9}
$$

5.

$$
\begin{aligned}
\frac{\partial \mathbf{y}^{(1)}}{\partial \mathbf{z}^{(1)}} &= \mathbf{y}^{(1)}(1 - \mathbf{y}^{(1)}) = diag(\mathbf{y}^{(1)} * (1 - \mathbf{y}^{(1)})) \\
\frac{\partial \sigma(z)}{\partial z} &= \frac{-(-1)e^{-z}}{(1 + e^{-z})^2} \\
&= \frac{1}{1 + e^{-z}} \cdot \frac{e^{-z}}{1 + e^{-z}} \\
&= \sigma(z)(1 - \sigma(z))
\end{aligned} \tag{10}
$$

6.

$$
\begin{aligned}
\frac{\partial E}{\partial \mathbf{z}^{(1)}} &= \frac{\partial E}{\partial \mathbf{y}^{(1)}}\frac{\partial \mathbf{y}^{(1)}}{\partial \mathbf{z}^{(1)}} \\
&= (\mathbf{y}^{(2)} - t)^\top \mathbf{W}^{(2)} diag(\mathbf{y}^{(1)}) * (1 - \mathbf{y}^{(1)})
\end{aligned}
\tag{11}
$$

7.

$$
\begin{aligned}
\frac{\partial E}{\partial W_{uv}^{(1)}} &= \frac{\partial E}{\partial \mathbf{z}^{(1)}}\frac{\partial \mathbf{z}^{(1)}}{\partial W_{uv}^{(1)}} \\
&= \frac{\partial E}{\partial z_u^{(1)}}\frac{\partial \mathbf{z}_u}{\partial w_{uv}^{(1)}} \\
&= \frac{\partial E}{\partial z_u^{(1)}}\frac{\partial (\sum_i w_{ui} x_i)}{\partial w_{uv}} \\
&= \frac{\partial E}{\partial z_u^{(1)}} x_v \\
i.e. \frac{\partial E}{\partial \mathbf{W}^{(1)}} &= (\frac{\partial E}{\partial \mathbf{z}^{(1)}})^\top \mathbf{x}^\top
\end{aligned}
\tag{12}
$$