Contributors:
adrian.buchwald@uni-potsdam.de
jana.schulz@uni-potsdam.de

Exercise 1

EdYou: Adrian Buchwald, Jana Schulz

December 18, 2021

Determine $\frac{\delta E}{\delta w_{ji}^O}$ and $\frac{\delta E}{\delta w_{ji}^H}$ of loss function $E(w,b) = \frac{1}{2}\sum_{k=1}^{N^O}(O_k^O - t_k)^2$ for a network with one input layer (with N_I neurons), output layer (with N_O neurons)and hidden layer (with N_H neurons). Note that every neuron is assumed to be connected to every neuron of the next layer, i.e., a Multi Layer Perceptron is considered. Further the sigmoid function is the considered action function for every neuron in the hidden and output layer.

Variables The following variables are defined for each $L \in \{I, H, O\}$, where I stands for the input layer, H stands the hidden layer and O stands the output layer. For the index $k \in \mathbb{N}$ the following is assumed $0 \leq k \leq N_L$.

N_L ... the number of neurons in layer L

w_{ji}^L ... the weight in layer L for neuron j with the incoming neuron i

$O_k^{L'}$ ... the output of the k-th neuron in Layer $L' \in \{H, O\}$
which is produced by the perceptron with the activation function:
$O_k^{L'} = sig(x_k^{L'})$

$x_k^{L'}$ ... the input of the k-th neuron in layer $L' \in \{H, O\}$

$$x_k^{L'} = \sum_{n=1}^{N_L} w_{nk}^L O_n^{(L-1)} + b_k^O \text{ with } (L-1) \text{ representing the previous layer}$$

$sig(x)$ ... the sigmoid function of x. It is the activation function for every neuron in the hidden and output layer

t_k ... the k-th target value

b_k^L ... the bias of the k-th neuron in layer L

Part 1 Determine $\frac{\delta E}{\delta w_{ji}^O}$

$$\begin{aligned}
\frac{\delta E}{\delta w_{ji}^O} &= \frac{\delta}{\delta w_{ji}^O}\frac{1}{2}\sum_{k=1}^{N_O}(O_k^O - t_k)^2 \\
&\overset{(4)}{=} (O_i^O - t_i) * \frac{\delta}{\delta w_{ji}^O} O_i^O \\
&\overset{(1)}{=} (O_i^O - t_i)\frac{\delta}{\delta w_{ji}^O} sig(x_i^O) \\
&\overset{(2),(3)}{=} (O_i^O - t_i) sig(x_i^O)(1 - sig(x_i^O)) * \frac{\delta}{\delta w_{ji}^O}\sum_{k=1}^{N_H} w_{ki}^O O_k^H + b_i^O \\
&\overset{(5)}{=} (O_i^O - t_i) sig(x_i^O)(1 - sig(x_i^O)) * O_j^H
\end{aligned}$$

That equation holds because of the following properties:

$$O_k^O = sig(x_k^O) \text{ with} \tag{1}$$

$$x_k^O = \sum_{l=1}^{N_O} w_{lk}^O \cdot O_l^H + b_k \text{ and} \tag{2}$$

$$\frac{\delta}{\delta w_{ji}^O} sig(x_k^O) = sig(x_k^O)(1 - sig(x_k^O))\frac{\delta}{\delta w_{ji}^O}(x_k^O) \tag{3}$$

Because of (1) only O_i^O is influenced by w_{ji}, which means

$$\frac{\delta}{\delta w_{ji}^O}\sum_{k=1}^{N_O}(O_k^O - t_k)^2 = 2(O_i^O - t_i)\frac{\delta}{\delta w_{ji}^O}(O_i^O - t_i) = 2(O_i^O - t_i)\frac{\delta}{\delta w_{ji}^O} O_i^O \tag{4}$$

Furthermore, for every $l \neq j$: $\frac{\delta}{\delta w_{ji}^O} w_{lk}^O \cdot O_l^H + b_k^O = 0$
and because of that, the following holds:

$$\frac{\delta x_k^O}{\delta w_{ji}^O} = \frac{\delta}{\delta w_{ji}^O}\sum_{l=1}^{N_O} w_{lk}^O \cdot O_l^H + b_k^O = O_j^H \tag{5}$$

Part 2 Determine $\frac{\delta E}{\delta w_{ji}^H}$

$$\begin{aligned}
\frac{\delta E}{\delta w_{ji}^H} &= \frac{\delta}{\delta w_{ji}^H} \frac{1}{2} \sum_{k=1}^{N_O} (O_k^O - t_k)^2 \\
&\overset{(6)}{=} \sum_{k=1}^{N_O} (O_k^O - t_k) * \frac{\delta}{\delta w_{ji}} O_k^O \overset{(1)}{=} \sum_{k=1}^{N_O} (O_k^O - t_k) * \frac{\delta}{\delta w_{ji}} sig(x_k^O) \\
&\overset{(2),(3)}{=} \sum_{k=1}^{N_O} (O_k^O - t_k) * sig(x_k^O) * (1 - sig(x_k^O)) * \frac{\delta}{\delta w_{ji}} \sum_{l=1}^{N_H} w_{lk}^O O_l^H + b_k^O \\
&\overset{(7)}{=} \sum_{k=1}^{N_O} (O_k^O - t_k) * sig(x_k^O) * (1 - sig(x_k^O)) * w_{ik}^O \frac{\delta}{\delta w_{ji}} O_i^H \\
&\overset{(8)}{=} \sum_{k=1}^{N_O} (O_k^O - t_k) * sig(x_k^O) * (1 - sig(x_k^O)) * w_{ik} * sig(x_i^H) * (1 - sig(x_i^H)) * O_j^I
\end{aligned}$$

The following properties explain why the equation holds:

Because we are looking at a weight w_{ji}^H of the hidden layer, every O_k^O is influenced by w_{ji}^H, therefore (see also 4):

$$\frac{\delta}{\delta w_{ji}^O} \sum_{k=1}^{N_O} (O_k^O - t_k)^2 = 2 \sum_{l=1}^{N_O} (O_k^O - t_k) \frac{\delta}{\delta w_{ji}^O} (O_k^O) \tag{6}$$

Only O_i^H is influenced by w_{ji}^H because:

$$O_i^H = sig(x_i^H) \text{ with } x_i^H = \sum_{k=1}^{N_H} w_{li}^O O_l^H + b_i^O \tag{7}$$

Equivalent to (5), the following holds:

$$\frac{\delta O_i^H}{\delta w_{ji}^H} = sig(x_i^H)(1 - sig(x_i^H)) \frac{\delta}{\delta w_{ji}^H} \sum_{j=1}^{N_H} (O_j^I w_{ji}^H + b_j) = sig(x_i^H)(1 - sig(x_i^H)) O_j^I \tag{8}$$