Exercise 3.1 Derivatives

$f : \mathbb{R}^N \to \mathbb{R}^M$
$g : \mathbb{R}^{M \times M} \to \mathbb{R}$
$h : \mathbb{R}^N \to \mathbb{R}, \qquad h(x) := g(f(x), f(x))$

using the information given on slide DL3.15/69:

$$f : \mathbb{R}^N \to \mathbb{R}^M, \quad g : \mathbb{R}^M \to \mathbb{R}^K, \quad h = g \circ f, \qquad \text{then}$$

$$Dh(x) = Dg(f(x))Df(x)$$

for $K = 1$:

$$\nabla h(x)^T = \nabla g(f(x))^T Df(x)$$

using the information given on slide DL3.22/69:

"the gradients being propagated back can just be added"

$$\Rightarrow \nabla h(x)^T = \nabla g(f_1(x), f_2(x))^T Df_1(x) + \nabla g(f_1(x), f_2(x))^T Df_2(x)$$

since $f_1 = f_2$

$$\nabla h(x)^T = 2\nabla g(f(x), f(x))^T Df(x)$$

Exercise 3.2 Backpropagation

$$f(x;w) = \sum_{i=0}^{2} w_{1,i}^{2,0} (\sum_{j=0}^{2} w_{0,j}^{1,i} x_j), \qquad f : \mathbb{R}^3 \to \mathbb{R}$$
$$\frac{\partial}{\partial w_{1,1}^{2,0}} f(x_l;w) = \sum_{j=0}^{2} w_{0,j}^{1,1} x_j$$
$$\frac{\partial}{\partial w_{0,1}^{1,1}} f(x_l;w) = w_{1,1}^{2,0} x_1$$

$$E(w) = \frac{1}{L} \sum_{l=1}^{L} \|d_l - f(x_l;w)\|_2^2$$

$$\begin{aligned}
\frac{\partial}{\partial w_{1,1}^{2,0}} E(w) &= \frac{1}{L} \frac{\partial}{\partial w_{1,1}^{2,0}} \sum_{l=1}^{L} (d_l - f(x_l;w))(d_l - f(x_l;w)) \\
&= \frac{1}{L} \frac{\partial}{\partial w_{1,1}^{2,0}} \sum_{l=1}^{L} (d_l^2 - 2d_l f(x_l;w) + f(x_l;w)^2) \\
&= \frac{1}{L} \sum_{l=1}^{L} (-2d_l \frac{\partial}{\partial w_{1,1}^{2,0}} f(x_l;w) + 2 \frac{\partial}{\partial w_{1,1}^{2,0}} f(x_l;w)) \\
&= \frac{2}{L} \sum_{l=1}^{L} (f(x_l;w) - d_l) \frac{\partial}{\partial w_{1,1}^{2,0}} f(x_l;w) \\
&= \frac{2}{L} \sum_{l=1}^{L} (f(x_l;w) - d_l) \sum_{j=0}^{2} w_{0,j}^{1,1} x_j
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial w_{0,1}^{1,1}} E(w) &= \frac{2}{L} \sum_{l=1}^{L} (f(x_l;w) - d_l) \frac{\partial}{\partial w_{0,1}^{1,1}} f(x_l;w) \\
&= \frac{2}{L} \sum_{l=1}^{L} (f(x_l;w) - d_l) w_{1,1}^{2,0} x_1
\end{aligned}$$