DEEP LEARNING IN COMPUTER VISION - EXERCISE 3

KARSTEN HERTH, FELIX HUMMEL, FELIX KAMMERLANDER, AND DAVID PALOSCH

Exercise 3.1

Let $N, M \in \mathbb{N}$. Further, let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ and $g\colon \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ both be differentiable. Let $h\colon \mathbb{R}^n \to \mathbb{R}$ be defined by $h(x) := g(f(x), f(x))$.
Then, h is differentiable with

$$(Dh)(x) = (Dg)(f(x), f(x)) \cdot \begin{pmatrix} (Df)(x) \\ (Df)(x) \end{pmatrix} \in \mathbb{R}^{1\times N}$$

for all $x \in \mathbb{R}^N$.

Proof.
For any $x \in \mathbb{R}^N$ the matrix $(Df)(x)$ is of size $M\times N$ and for all $x \in \mathbb{R}^{2M} \cong \mathbb{R}^M \times \mathbb{R}^M$ the matrix $(Dg)(x)$ is of size $1 \times 2M$.
We define the functions

$$\begin{aligned} h_1(x) &:= (f(x), 0) \in \mathbb{R}^{1\times 2M} \\ h_2(x) &:= (0, f(x)) \in \mathbb{R}^{1\times 2M} \\ h_3(x) &:= h_1(x) + h_2(x) \in \mathbb{R}^{1\times 2M} \end{aligned}$$

for $x \in \mathbb{R}^N$, such that $h(x) = g(h_3(x))$. Obviously, h_1 and h_2 (and therefore, h_3 as well) are differentiable with derivatives

$$(Dh_1)(x) = \begin{pmatrix} \nabla f_1(x)^\top \\ \vdots \\ \nabla f_M^\top(x) \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} (Df)(x) \\ 0 \end{pmatrix} \in \mathbb{R}^{2M\times N}$$

$$(Dh_2)(x) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \nabla f_1(x)^\top \\ \vdots \\ \nabla f_M(x)^\top \end{pmatrix} = \begin{pmatrix} 0 \\ (Df)(x) \end{pmatrix} \in \mathbb{R}^{2M\times N}$$

for $x \in \mathbb{R}^N$.
In particular, $h = g \circ h_3$ is differentiable by the chain rule and we obtain, using the

Date: January 7, 2018.

linearity of the derivative,

$$\begin{aligned}
(Dh)(x) &= (Dg)(h_3(x)) \cdot (Dh_3)(x) \\
&= (Dg)(f(x), f(x)) \cdot [(Dh_1)(x) + (Dh_2)(x)] \\
&= (Dg)(f(x), f(x)) \cdot \begin{pmatrix} \nabla f_1(x)^\top \\ \vdots \\ \nabla f_M(x)^\top \\ \nabla f_1(x)^\top \\ \vdots \\ \nabla f_M(x)^\top \end{pmatrix} \\
&= (Dg)(f(x), f(x)) \cdot \begin{pmatrix} (Df)(x) \\ (Df)(x) \end{pmatrix} \in \mathbb{R}^{1\times N}
\end{aligned}$$

for $x \in \mathbb{R}^N$. □

Exercise 3.2

Let

$$f(x, w) := \sum_{i,j=0}^{2} w_{1,i}^{2,0} w_{0,j}^{1,i} x_j$$

and

$$E(w) = \frac{1}{L} \sum_{l=1}^{L} \left(d^l - f(x^l, w)\right)^2,$$

where x^l denotes the l-th sample and d^l denotes the l-th label.
We calculate according to the chain rule

$$\begin{aligned}
\frac{\partial}{\partial w_{1,1}^{2,0}} E(w) &= \frac{2}{L} \sum_{l=1}^{L} \left(d^l - f(x^l, w)\right) \frac{\partial}{\partial w_{1,1}^{2,0}} f(x^l, w) \\
&= \frac{2}{L} \sum_{l=1}^{L} \left(d^l - f(x^l, w)\right) \sum_{j=0}^{2} w_{0,j}^{1,1} x_j^l,
\end{aligned}$$

as well as

$$\begin{aligned}
\frac{\partial}{\partial w_{0,1}^{1,0}} E(w) &= \frac{2}{L} \sum_{l=1}^{L} \left(d^l - f(x^l, w)\right) \frac{\partial}{\partial w_{0,1}^{1,0}} f(x^l, w) \\
&= \frac{2}{L} \sum_{l=1}^{L} \left(d^l - f(x^l, w)\right) w_{1,0}^{2,0} x_1^l.
\end{aligned}$$