4. Multilayer Perceptron

Problem 4.1.1.1. Show that $\frac{\partial J}{\partial W_{ij}} = \frac{\partial J}{\partial y_i} x_j$, where $x = (x_1, \ldots, x_d)^T$. [Hint: Although not necessary, you might find it helpful to use the notation $\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & \text{else} \end{cases}$.

Solution Note that

$$\frac{\partial y_r}{\partial W_{ij}} = \delta_{rj} \frac{\partial y_r}{\partial W_{ij}}$$

And

$$\frac{\partial y_j}{\partial W_{ij}} = x_j$$

Then

$$\begin{aligned}
\frac{\partial J}{\partial W_{ij}} &= \sum_{r=1}^{m} \frac{\partial J}{\partial y_r} \frac{\partial y_r}{\partial W_{ij}}. \\
&= \frac{\partial J}{\partial y_j} \frac{\partial y_j}{\partial W_{ij}} \\
&= \frac{\partial J}{\partial y_j} x_j
\end{aligned}$$

Problem 4.1.1.2. Now let's vectorize this. Let's write $\frac{\partial J}{\partial y} \in \mathbf{R}^{m\times 1}$ for the column vector whose ith entry is $\frac{\partial J}{\partial y_i}$. Let's also define the matrix $\frac{\partial J}{\partial W} \in \mathbf{R}^{m\times d}$, whose ij'th entry is $\frac{\partial J}{\partial W_{ij}}$. Generally speaking, we'll always take $\frac{\partial J}{\partial A}$ to be an array of the same size ("shape" in numpy) as A. Give a vectorized expression for $\frac{\partial J}{\partial W}$ in terms of the column vectors $\frac{\partial J}{\partial y}$ and x. [Hint: Outer product.]

Solution

$$\frac{\partial J}{\partial W} = \frac{\partial J}{\partial y} \bigotimes x$$

Problem 4.1.1.3. In the usual way, define $\frac{\partial J}{\partial x} \in \mathbf{R}^d$, whose i'th entry is $\frac{\partial J}{\partial x_i}$. Show that

$$\frac{\partial J}{\partial x} = W^T \left(\frac{\partial J}{\partial y} \right)$$

Solution et W_i denotes the ith column of W.
Then observe that

$$\begin{aligned} \frac{\partial J}{\partial x_i} &= \sum_{r=1}^{m} \frac{\partial J}{\partial y_r} \frac{\partial y_r}{\partial x_i} \\ &= \sum_{r=1}^{m} \frac{\partial J}{\partial y_r} W_{r,i} \\ &= (W_i)^T \frac{\partial J}{\partial y} \end{aligned}$$

Thus

$$\frac{\partial J}{\partial x} = W^T \frac{\partial J}{\partial y}$$

Problem 4.1.1.4. Show that $\frac{\partial J}{\partial b} = \frac{\partial J}{\partial y}$, where $\frac{\partial J}{\partial b}$ is defined in the usual way.

Solution Note that $\frac{\partial y}{\partial b} = 1$. Then

$$\frac{\partial J}{\partial b} = \frac{\partial J}{\partial y}\frac{\partial y}{\partial b} = \frac{\partial J}{\partial y}$$

Problem 4.1.2. Show that $\frac{\partial J}{\partial A} = \frac{\partial J}{\partial S} \odot \sigma'(A)$, where we're using $\odot$ to represent the **Hadamard product**. If A and B are arrays of the same shape, then their Hadamard product $A \odot B$ is an array with the same shape as A and B, and for which $(A \odot B)_i = A_i B_i$. That is, it's just the array formed by multiplying corresponding elements of A and B. Conveniently, in numpy if A and B are arrays of the same shape, then A*B is their Hadamard product.

Solution Because of how $\sigma()$ is defined, $\frac{\partial S_i}{\partial A_i} = \sigma'(A_i)$.
Thus $\frac{\partial J}{\partial A_i} = \frac{\partial J}{\partial S_i} \frac{\partial S_i}{\partial A_i} = \frac{\partial J}{\partial S_i} \sigma'(A_i)$.
So then $\frac{\partial J}{\partial A} = \frac{\partial J}{\partial S} \odot \sigma'(A)$.