TTIC 31230 Fundamentals of Deep Learning
Problems For Fundamental Equations.

Problem 0: Backprogation through a ReLU linear threshold unit.
Consider the computation

$$
\begin{aligned}
y &= \sigma(w^\top x) \\
\ell &= \mathcal{L}(y)
\end{aligned}
$$

for $w, x \in R^d$ with $\sigma(z) = \max(z, 0)$ (the ReLU activation) and for $\mathcal{L}(y)$ an arbitrary function (a loss function). Let w_i denote the ith component of the weight vector w. Give an expression for $\frac{\partial \ell}{\partial w_i}$ as a function of $\frac{d\mathcal{L}(y)}{dy}$.

Solution: There are various correct ways of writing the answer. The following corresponds to a backpropagation computation.

$$
\begin{aligned}
\frac{d\ell}{dy} &= \frac{d\mathcal{L}(y)}{dy} \\
\frac{d\ell}{dw_i} &= \frac{d\ell}{dy}\frac{dy}{dw_i} \;=\; \frac{d\ell}{dy}\, x_i \mathbf{1}[w_i x_i \geq 0]
\end{aligned}
$$