Problem 7.1

May 12, 2024

1 Problem

A two-layer network with two hidden units in each layer can be defined as:

$$
\begin{aligned}
y = \phi_0 + \phi_1 a\left[\nu_{01} + \nu_{11} a[\theta_{01} + \theta_{11} x] + \nu_{21} a[\theta_{02} + \theta_{12} x]\right] \\
+ \phi_2 a\left[\nu_{02} + \nu_{12} a[\theta_{01} + \theta_{11} x] + \nu_{22} a[\theta_{02} + \theta_{12} x]\right]
\end{aligned}
$$

Compute the derivatives of the output y with respect to each of the 13 parameters directly (not using back propagation). Use $\mathbb{I}$ for $\frac{\partial a[z]}{\partial z}$ where a represents the ReLU activation function.

2 Answer

$$
\begin{aligned}
\frac{\partial y}{\partial \phi_0} &= 1 \\
\frac{\partial y}{\partial \phi_1} &= a(\nu_{01} + \nu_{01} + \nu_{11} a[\theta_{01} + \theta_{11} x] + \nu_{21} a[\theta_{02} + \theta_{12} x]) \\
\frac{\partial y}{\partial \phi_2} &= a(\nu_{02} + \nu_{12} a[\theta_{01} + \theta_{11} x] + \nu_{22} a[\theta_{02} + \theta_{12} x])
\end{aligned}
$$

For the θ_{01} partial derivative I used chain rule twice, with u_1 as $\phi_1 \nu_{11} a[\theta_{01}]$ for the first symmetric part of the equation ϕ_1 (and similarly u_2 as $\phi_2 \nu_{12} a[\theta_{01}]$ for the ϕ_2)

$$
\begin{aligned}
y &= a[u_1] + a[u_2] \\
\frac{\partial y}{\partial \theta_{01}} &= \frac{\partial y}{\partial u_1} \cdot \frac{\partial u_1}{\partial \theta_{01}} + \frac{\partial y}{\partial u_2} \cdot \frac{\partial u_2}{\partial \theta_{01}} \\
\frac{\partial y}{\partial \theta_{01}} &= \mathbb{I}(\phi_1 \nu_{11} a[\theta_{01}]) \cdot \phi_1 \nu_{11} + \mathbb{I}(\phi_2 \nu_{12} a[\theta_{01}]) \cdot \phi_2 \nu_{12}
\end{aligned}
$$

I dropped the final a terms around θ_{01} to leave $\cdot \phi_1 \nu_{11}$ and $\cdot \phi_2 \nu_{12}$ because of the multiplication with the $\mathbb{I}$ covers the negative case as zero already.

I haven't yet worked out the other terms but I think similar logic applies.

I have asked for a sanity check on `math.stackexchange.com/questions/4914949`.