2022/07/08

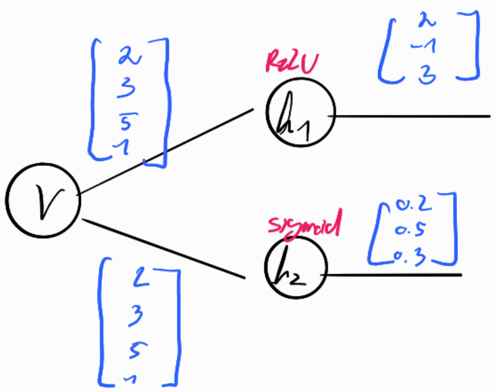
Exercise 3

Consider a neural network in which a vectored node v feeds into two distinct vectored nodes h_1 and h_2 computing different functions. The functions computed at the nodes are $h_1 = ReLU(W_1 v)$ and $h_2 = sigmoid(W_2 v)$. We do not know anything about the values of the variables in other parts of the network, but we know that $h_1 = [2, -1, 3]^T$ and $h_2 = [0.2, 0.5, 0.3]^T$, that are connected to the node $v = [2, 3, 5, 1]^T$. Furthermore the loss gradients are $\frac{\partial L}{\partial h_1} = [-2, 1, 4]^T$ and $\frac{\partial L}{\partial h_2} = [1, 3, 2]^T$, respectively. Show that the backpropagated loss gradient $\frac{\partial L}{\partial v}$ can be computed in terms of W_1 and W_2 as follows:

$$\frac{\partial L}{\partial v} = W_1^T \begin{bmatrix} -2 \\ 0 \\ 4 \end{bmatrix} + W_2^T \begin{bmatrix} 0.16 \\ 0.75 \\ -0.42 \end{bmatrix} \tag{5}$$

What are the sizes of W_1, W_2 and $\frac{\partial L}{\partial v}$?

Remember that $ReLU(x) = \max(0, x)$ and $sigmoid(x) = \frac{\exp(x)}{(\exp(x)+1)}$.



$$\frac{\partial L}{\partial h_1} = \begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix}$$

$$\frac{\partial L}{\partial h_2} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

- With backpropagation rule

$$\frac{\partial L}{\partial v} = \frac{\partial L}{\partial h_1} \cdot \frac{\partial h_1}{\partial v} + \frac{\partial L}{\partial h_2} \cdot \frac{\partial h_2}{\partial v}$$

$$= W_1^T \left(\frac{\partial L}{\partial h_1} \odot \frac{\partial h_1}{\partial z_1} \right) + W_2^T \left(\frac{\partial L}{\partial h_2} \odot \frac{\partial h_2}{\partial z_2} \right)$$

with $z_1 = W_1 v$, $h_1 = ReLU(z_1) \Longrightarrow \frac{\partial h_1}{\partial z_1} = ReLU'(z_1)$

$z_2 = W_2 v$, $h_2 = sigmoid(z_2) \to \frac{\partial h_2}{\partial z_2} = \sigma(z_2)(1 - \sigma(z_2))$

$\sigma(z_2)$

$\Longrightarrow \frac{\partial h_1}{\partial z_1}$ $\quad ReLU' = \begin{cases} 1 & z_i > 0 \\ 0 & z_i \leq 0 \end{cases}$

From $h_1 = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} \longrightarrow z_1 = W_1 v = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} \longrightarrow ReLU'(z_1) = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

$\to \frac{\partial L}{\partial h_1} \odot ReLU'(z_1) = \begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix} \odot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 4 \end{bmatrix}$

$$\Rightarrow \frac{\partial h_2}{\partial z_2} \qquad \sigma'(z) = \sigma(z)(1-\sigma(z))$$

From $h_2 = \begin{bmatrix} 0.2 \\ 0.5 \\ 0.3 \end{bmatrix} \longrightarrow z_2 = W_2 V \begin{bmatrix} 0.2 \\ 0.5 \\ 0.3 \end{bmatrix} \to \sigma'(z_2) = \begin{bmatrix} 0.2 \cdot 0.8 \\ 0.5 \cdot 0.5 \\ 0.3 \cdot 0.7 \end{bmatrix} = \begin{bmatrix} 0.16 \\ 0.25 \\ 0.21 \end{bmatrix}$

$$\Rightarrow \frac{\partial L}{\partial h_2} \odot \sigma'(z_2) = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \odot \begin{bmatrix} 0.16 \\ 0.25 \\ 0.21 \end{bmatrix} = \begin{bmatrix} 0,16 \\ 0.75 \\ 0,42 \end{bmatrix}$$

Sizes: $V \in \mathbb{R}^4$ $\quad h_1, h_2 \in \mathbb{R}^3$ $\quad W_1, W_2 \in \mathbb{R}^{3\times 4}$ $\quad \frac{\partial L}{\partial V} \in \mathbb{R}^4$