CSC2516: Homework #1

Due on January 24

Roger Grosse

Thomas Hollis

Problem 1

Please give a set of weights and biases for the network which correctly implements this function.

Solution

The weights and biases that correctly implement this function are as follows:

$$\mathbf{W}^{(1)} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \tag{1}$$

$$\mathbf{b}^{(1)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{2}$$

$$\mathbf{w}^{(2)} = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix} \tag{3}$$

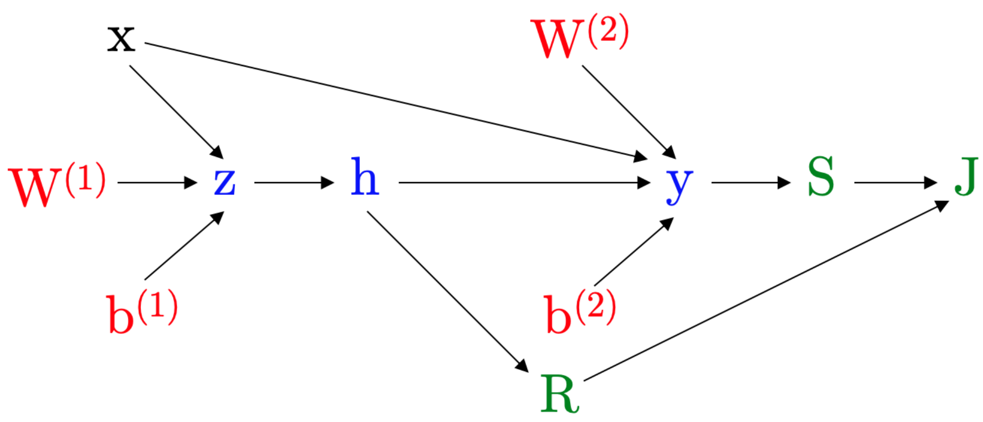
$$b^{(2)} = 0 \tag{4}$$

Problem 2

Draw the computation graph relating $\mathbf{x}, \mathbf{z}, \mathbf{h}, \mathbf{y}, R, S$, and J.

Solution

The computation graph is as follows:



Problem 3

Derive the backprop equations for computing $\bar{\mathbf{x}} = \partial J / \partial \mathbf{x}$.

Solution

The network equations are as follows:

$$J = R + S$$

$$R = \sum_i r_i h_i$$

$$S = \frac{1}{2} \sum_i (y_i - s_i)^2$$

$$y_i = x_i + \sum_j w_{ij}^{(2)} h_j + b_i^{(2)}$$

$$h_i = \sigma(z_i)$$

$$z_i = \sum_j w_{ij}^{(1)} x_j + b_i^{(1)}$$

From our computation graph, backprop will ultimately compute the following:

$$\frac{\partial J}{\partial \mathbf{x}} = \frac{\partial R}{\partial \mathbf{x}} + \frac{\partial S}{\partial \mathbf{x}}$$

Where:

$$\frac{\partial R}{\partial \mathbf{x}} = \frac{\partial R}{\partial \mathbf{h}} \frac{\partial \mathbf{h}}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial \mathbf{x}}$$

$$\frac{\partial S}{\partial \mathbf{x}} = \frac{\partial S}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \mathbf{h}} \frac{\partial \mathbf{h}}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial \mathbf{x}}$$

Now we can show the efficient derivative computations (using class notation, avoiding naive approach):

$$\frac{\partial J}{\partial R} = \bar{R} = 1$$

$$\frac{\partial J}{\partial S} = \bar{S} = 1$$

$$\frac{\partial J}{\partial y_i} = \bar{y_i} = \bar{S}(y_i - s_i)$$

$$\frac{\partial J}{\partial h_i} = \bar{h_i} = \bar{R} r_i + \sum_k \bar{y_k} w_{ik}^{(2)}$$

$$\frac{\partial J}{\partial z_i} = \bar{z_i} = \bar{h_i} \sigma'(z_i)$$

$$\frac{\partial J}{\partial x_i} = \bar{x_i} = \bar{y_i} + \sum_k \bar{z_k} w_{ik}^{(1)}$$

Which gives us the equivalent vectorised forms:

$$\bar{\mathbf{y}} = \bar{S}(\mathbf{y} - \mathbf{s})$$

$$\bar{\mathbf{h}} = \bar{R}\mathbf{r} + \mathbf{W}^{(2)T} \bar{\mathbf{y}}$$

$$\bar{\mathbf{z}} = \bar{\mathbf{h}} \circ \sigma'(\mathbf{z})$$

$$\bar{\mathbf{x}} = \bar{\mathbf{y}} + \mathbf{W}^{(1)T} \bar{\mathbf{z}}$$

Thus:

$$\bar{\mathbf{x}} = (\mathbf{y} - \mathbf{s}) + \mathbf{W}^{(1)T} \Bigg(\Big(\mathbf{r} + \mathbf{W}^{(2)T} (\mathbf{y} - \mathbf{s}) \Big) \circ \sigma'(\mathbf{z}) \Bigg)$$

Problem 4

Which of the weight derivatives are guaranteed to be 0 for this training case?

Solution

The weight derivatives that are guaranteed to be 0 are as follows:

$\frac{\partial L}{\partial w_1}$: YES

$\frac{\partial L}{\partial w_2}$: YES

$\frac{\partial L}{\partial w_3}$: NO

The justification behind this can be easily provided by backprop. First, we list out the equations that characterise the network:

$$
\begin{gathered}
L = \frac{1}{2}(y-2)^2 \\
y = ReLU(h_1 w_1 + h_2 w_5) \\
h_1 = ReLU(h_3 w_2 + \dots) \\
h_2 = ReLU(h_3 w_4 + \dots) \\
h_3 = ReLU(x_1 w_3 + \dots) \\
\bar{L} = (y-t)
\end{gathered}
$$

Now we can find the equations of the derivatives $\bar{w_1}$, $\bar{w_2}$ and $\bar{w_3}$.

$$
\begin{gathered}
\bar{w_1} = \bar{L}\frac{\partial y}{\partial w_1} = (y-t) \cdot ReLU'(h_1 w_1 + h_2 w_5) h_1 \\
\therefore \bar{w_1} = \bar{L}\frac{\partial y}{\partial w_1} = (y-t) \cdot ReLU'(h_1 w_1 + h_2 w_5) \cdot 0 = 0
\end{gathered}
$$

For $\bar{w_1}$, from the above equations, we can clearly see than when h_1 receives an input of -1 the ReLU evaluates to 0. Hence $\bar{w_1}$ will also necessarily be equal to 0 since $h_1 = 0$. Thus the answer is: YES.

As for $\bar{w_2}$, we know that $(h_3 w_2 + \cdots) = -1$ since it is the input to h_1 so we have:

$$
\begin{gathered}
\bar{w_2} = \bar{L}\frac{\partial y}{\partial h_1}\frac{\partial h_1}{\partial w_2} = (y-t) \cdot ReLU'(h_1 w_1 + h_2 w_5) w_1 \cdot ReLU'(h_3 w_2 + \cdots) h_3 \\
\therefore \bar{w_2} = (y-t) \cdot ReLU'(h_1 w_1 + h_2 w_5) w_1 \cdot ReLU'(-1) h_3 = 0
\end{gathered}
$$

Hence since $ReLU'(-1) = 0$ we know that $\bar{w_2} = 0$. Thus the answer is: YES.

Finally for $\bar{w_3}$, we also know that $(h_3 w_2 + \cdots) = -1$ since it is the input to h_1 so we have:

$$
\begin{gathered}
\bar{w_3} = \bar{L}\frac{\partial y}{\partial h_1}\frac{\partial h_1}{\partial h_3}\frac{\partial h_3}{\partial w_3} + \bar{L}\frac{\partial y}{\partial h_2}\frac{\partial h_2}{\partial h_3}\frac{\partial h_3}{\partial w_3} \\
\therefore \bar{w_3} = (y-t) \cdot ReLU'(h_1 w_1 + h_2 w_5) w_1 \cdot ReLU'(h_3 w_2 + \cdots) w_2 \cdot ReLU'(x_1 w_3 + \cdots) x_1 \\
+ (y-t) \cdot ReLU'(h_1 w_1 + h_2 w_5) w_5 \cdot ReLU'(h_3 w_4 + \cdots) w_4 \cdot ReLU'(x_1 w_3 + \cdots) x_1 \\
\therefore \bar{w_3} = (y-t) \cdot ReLU'(h_1 w_1 + h_2 w_5) w_1 \cdot ReLU'(-1) w_2 \cdot ReLU'(x_1 w_3 + \cdots) x_1 \\
+ (y-t) \cdot ReLU'(h_1 w_1 + h_2 w_5) w_5 \cdot ReLU'(h_3 w_4 + \cdots) w_4 \cdot ReLU'(x_1 w_3 + \cdots) x_1
\end{gathered}
$$

$$
\therefore \bar{w_3} = 0 + (y-t) \cdot ReLU'(h_1 w_1 + h_2 w_5) w_5 \cdot ReLU'(h_3 w_4 + \cdots) w_4 \cdot ReLU'(x_1 w_3 + \cdots) x_1
$$

Hence since $\bar{w_3}$ only reduces to 0 on one side of the sum we cannot guarantee that it will be equal to 0. Hence the answer is: NO.