Q1. Neural Nets

Consider the following computation graph for a simple neural network for binary classification. Here $x$ is a single real-valued input feature with an associated class $y^*$ (0 or 1). There are two weight parameters $w_1$ and $w_2$, and non-linearity functions $g_1$ and $g_2$ (to be defined later, below). The network will output a value $a_2$ between 0 and 1, representing the probability of being in class 1. We will be using a loss function $\text{Loss}$ (to be defined later, below), to compare the prediction $a_2$ with the true class $y^*$.

1. Perform the forward pass on this network, writing the output values for each node $z_1, a_1, z_2$ and $a_2$ in terms of the node’s input values:

2. Compute the loss $\text{Loss}(a_2, y^*)$ in terms of the input $x$, weights $w_i$, and activation functions $g_i$:

3. Now we will work through parts of the backward pass, incrementally. Use the chain rule to derive $\frac{\partial \text{Loss}}{\partial w_2}$. Write your expression as a product of partial derivatives at each node: i.e. the partial derivative of the node’s output with respect to its inputs. (Hint: the series of expressions you wrote in part 1 will be helpful; you may use any of those variables.)
4. Suppose the loss function is quadratic, \( \text{Loss}(a_2, y^*) = \frac{1}{2}(a_2 - y^*)^2 \), and \( g_1 \) and \( g_2 \) are both sigmoid functions \( g(z) = \frac{1}{1+e^{-z}} \) (note: it's typically better to use a different type of loss, cross-entropy, for classification problems, but we'll use this to make the math easier).

Using the chain rule from Part 3, and the fact that \( \frac{\partial g(z)}{\partial z} = g(z)(1 - g(z)) \) for the sigmoid function, write \( \frac{\partial \text{Loss}}{\partial \omega_2} \) in terms of the values from the forward pass, \( y^*, a_1, \) and \( a_2 \):

5. Now use the chain rule to derive \( \frac{\partial \text{Loss}}{\partial \omega_1} \) as a product of partial derivatives at each node used in the chain rule:

6. Finally, write \( \frac{\partial \text{Loss}}{\partial \omega_1} \) in terms of \( x, y^*, w_1, a_1, z^*_i \):

7. What is the gradient descent update for \( \omega_1 \) with step-size \( \alpha \) in terms of the values computed above?
Q2. Neural Network Data Sufficiency

The next few problems use the below neural network as a reference. Neurons $h_1, h_2, h_3$ and $j_1, j_2$ all use ReLU activation functions. Neuron $y$ uses the identity activation function: $f(x) = x$. In the questions below, let $w_{a,b}$ denote the weight that connects neurons $a$ and $b$. Also, let $a_x$ denote the value that neuron $a$ outputs to its next layer.

Given this network, in the following few problems, you have to decide whether the data given are sufficient for answering the question.

(a) Given the above neural network, what is the value of $a_y$?
   Data item 1: the values of all weights in the network and the values $a_{h_1}, a_{h_2}, a_{h_3}$
   Data item 2: the values of all weights in the network and the values $a_{j_1}, a_{j_2}$
   ○ Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
   ○ Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
   ○ Both statements taken together are sufficient, but neither data item alone is sufficient.
   ○ Each data item alone is sufficient to answer the question.
   ○ Statements (1) and (2) together are not sufficient, and additional data is needed to answer the question.

(b) Given the above neural network, what is the value of $a_{h_1}$?
   Data item 1: the neuron input values, i.e., $a_{x_1}$ through $a_{x_4}$
   Data item 2: the values $a_{j_1}, a_{j_2}$
   ○ Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
   ○ Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
   ○ Both statements taken together are sufficient, but neither data item alone is sufficient.
   ○ Each data item alone is sufficient to answer the question.
   ○ Statements (1) and (2) together are not sufficient, and additional data is needed to answer the question.

(c) Given the above neural network, what is the value of $a_{j_2}$?
   Data item 1: the values of all weights connecting neurons $h_1, h_2, h_3$ to $j_1, j_2$
   Data item 2: the values $a_{h_1}, a_{h_2}, a_{h_3}$
   ○ Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
   ○ Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
   ○ Both statements taken together are sufficient, but neither data item alone is sufficient.
   ○ Each data item alone is sufficient to answer the question.
   ○ Statements (1) and (2) together are not sufficient, and additional data is needed to answer the question.
(d) Given the above neural network, what is the value of $\partial o_y / \partial w_{j_2,j_3}$?

Data item 1: the value of $o_{j_2}$.

Data item 2: all weights in the network and the neuron input values, i.e., $o_{x_1}$ through $o_{x_4}$.

- Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
- Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
- Both statements taken together are sufficient, but neither data item alone is sufficient.
- Each data item alone is sufficient to answer the question.
- Statements (1) and (2) together are not sufficient, and additional data is needed to answer the question.

(e) Given the above neural network, what is the value of $\partial o_y / \partial w_{h_2,j_2}$?

Data item 1: the value of $w_{j_2,y}$

Data item 2: the value of $\partial o_{j_2} / \partial w_{h_2,j_2}$

- Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
- Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
- Both statements taken together are sufficient, but neither data item alone is sufficient.
- Each data item alone is sufficient to answer the question.
- Statements (1) and (2) together are not sufficient, and additional data is needed to answer the question.

(f) Given the above neural network, what is the value of $\partial o_y / \partial w_{x_1,h_3}$?

Data item 1: the value of all weights in the network and the neuron input values, i.e., $o_{x_1}$ through $o_{x_4}$.

Data item 2: the value of $w_{x_1,h_3}$

- Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
- Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
- Both statements taken together are sufficient, but neither data item alone is sufficient.
- Each data item alone is sufficient to answer the question.
- Statements (1) and (2) together are not sufficient, and additional data is needed to answer the question.