CS188: Exam Practice Session 11

Q1. Deep Learning

(a) Consider the neural network below. We refer to nodes that apply the activation function as neurons numbered from top-down, left-right. Draw the computational graph of the following neural network, labeling all the bolded variables listed below. (Please use the same conventions as from project 6)

- $\mathbf{x}$, a scalar, is the input. $\mathbf{a_4}$, also a scalar, is the output.
- There are three hidden layers of size 1, 2, and 1.
- All neurons $\mathbf{n_i}$ have the same activation function $g(x) = e^x$.
  (numbered from top-down, left-right)
- Each $\mathbf{a_i}$ value is the final output for neuron i in the network.
- Each $\mathbf{z_i}$ value is the pre-activation value for neuron i in the network (i.e. after the dot product, before the nonlinear activation).
- Let $\mathbf{w_{ij}}$ be the weight used between neuron n_i and n_j.
  (You can consider x to be n_0 for the purposes of labeling weights.)
- $\mathbf{L}(y, a_4)$ is the loss function. It is $(y - a_4)^2$, where $\mathbf{y}$ is the training label.

(b) Use the Chain Rule and the equations given to calculate $\delta L/\delta z_2$. You may use x and y, along with all of the w, a, and z values.

(c) Use the Chain Rule and the equations given to calculate $\delta L/\delta z_3$. You may use x and y, along with all of the w, a, and z values.

(d) Use the Chain Rule and the equations given to calculate $\delta L/\delta w_{01}$. You may use x and y, along with all of the w, a, and z values. You may symbolically use previous parts.

Q2. Gradients

Given that

$$p(y = 1 \mid f(x); w) = \frac{e^{w^\top f(x)}}{e^{w^\top f(x)} + e^{-w^\top f(x)}}$$

$$p(y = 0 \mid f(x); w) = \frac{e^{-w^\top f(x)}}{e^{w^\top f(x)} + e^{-w^\top f(x)}}$$

$$g(x; w) = \begin{cases} w^\top x & w^\top x > 0 \\ \alpha(e^{w^\top x} - 1) & w^\top x \leq 0 \end{cases}$$

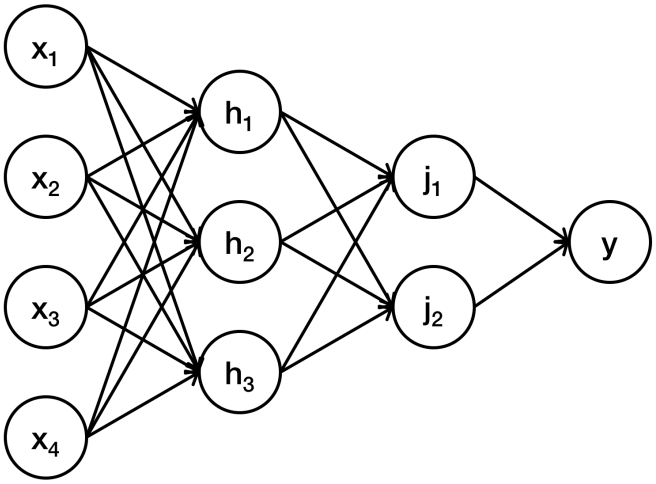
Take the following gradients:

(a) $\frac{\partial}{\partial w} g(x; w)$

(b) $\frac{\partial}{\partial w} \sum_{i=1}^{m} \log p(y = y^{(y)} \mid f(x^{(i)}); w)$

Q3. Neural Network Data Sufficiency

The next few problems use the below neural network as a reference. Neurons h_{1-3} and j_{1-2} 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 o_a 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 o_y?

Data item 1: the values of all weights in the network and the values o_{h_1}, o_{h_2}, o_{h_3}
Data item 2: the values of all weights in the network and the values o_{j_1}, o_{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 o_{h_1}?

Data item 1: the neuron input values, i.e., o_{x_1} through o_{x_4}
Data item 2: the values o_{j_1}, o_{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 o_{j_1}?

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 o_{h_1}, o_{h_2}, o_{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 $\partial o_y / \partial w_{j_2,y}$?

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.