CS 188 Introduction to
Summer 2020 Artificial Intelligence

Final Review

NN and Logistic Regression

Q1. Optimization

We would like to classify some data. We have N samples, where each sample consists of a feature vector $\mathbf{x} = \{x_1, \cdots, x_k\}$ and a label $y = \{0, 1\}$.

We introduce a new type of classifier called logistic regression, which produces predictions as follows:

$$P(Y = 1|X) = h(\mathbf{x}) = s\left(\sum_i w_i x_i\right) = \frac{1}{1 + \exp(-(\sum_i w_i x_i))}$$

$$s(\gamma) = \frac{1}{1 + \exp(-\gamma)}$$

where $s(\gamma)$ is the logistic function, $\exp x = e^x$, and $\mathbf{w} = \{w_1, \cdots, w_k\}$ are the learned weights.

Let's find the weights w_j for logistic regression using stochastic gradient descent. We would like to minimize the following loss function for each sample:

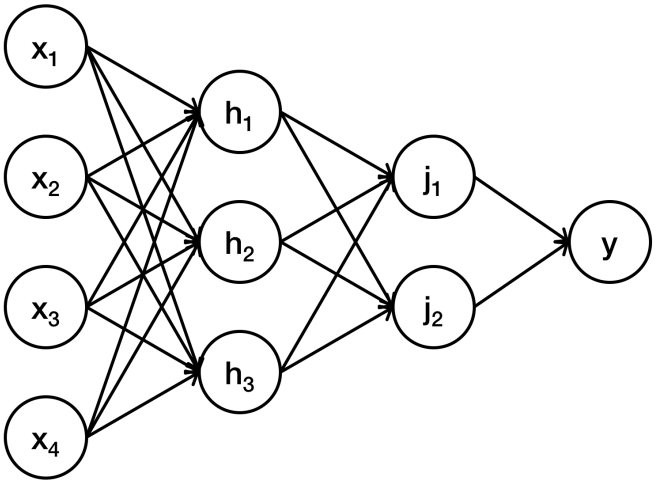
$$L = -[y \ln h(\mathbf{x}) + (1 - y) \ln(1 - h(\mathbf{x}))]$$

(a) Find dL/dw_i. Hint: $s'(\gamma) = s(\gamma)(1 - s(\gamma))$.

(b) Write the stochastic gradient descent update for w_i. Our step size is η.

Q2. 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 $\partial o_{j_2} / \partial w_{h_2,j_2}$

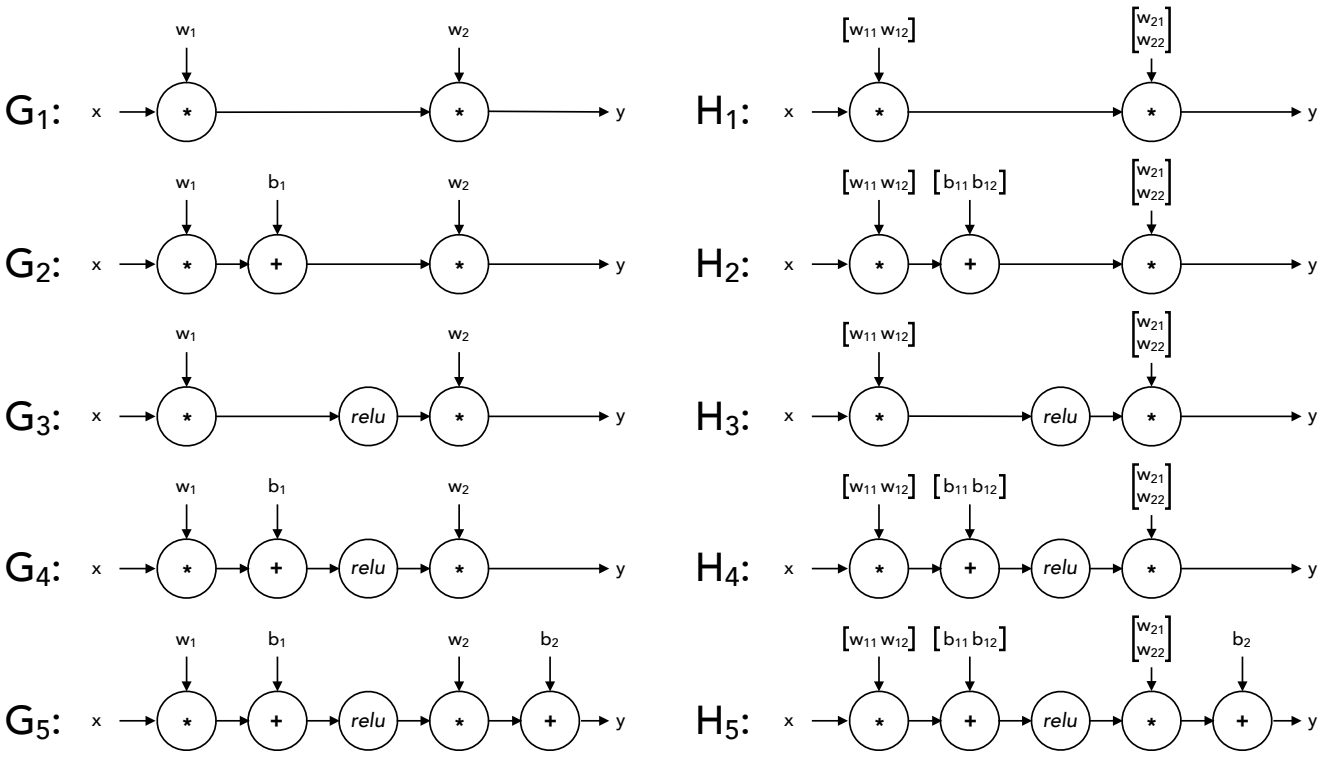
- ◯ Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
- ◯ Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
- ◯ Both statements taken together are sufficient, but neither data item alone is sufficient.
- ◯ 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}

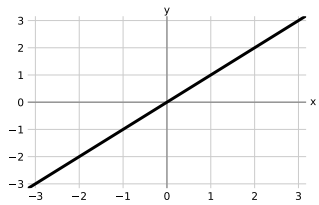
- ◯ Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
- ◯ Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
- ◯ Both statements taken together are sufficient, but neither data item alone is sufficient.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

Q3. Neural Networks: Representation

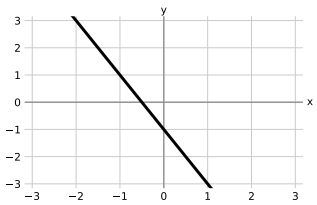


For each of the piecewise-linear functions below, mark all networks from the list above that can represent the function **exactly** on the range $x \in (-\infty, \infty)$. In the networks above, *relu* denotes the element-wise ReLU nonlinearity: $relu(z) = max(0, z)$. The networks G_i use 1-dimensional layers, while the networks H_i have some 2-dimensional intermediate layers.

(a)

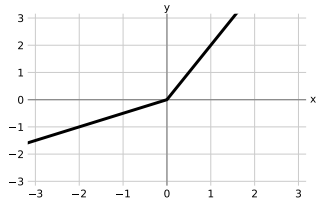


☐ G_1 ☐ G_2 ☐ G_3 ☐ G_4 ☐ G_5 ☐ H_1 ☐ H_2 ☐ H_3 ☐ H_4 ☐ H_5 ○ None of the above

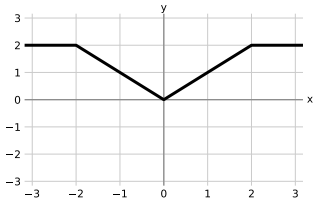


☐ G_1 ☐ G_2 ☐ G_3 ☐ G_4 ☐ G_5 ☐ H_1 ☐ H_2 ☐ H_3 ☐ H_4 ☐ H_5 ○ None of the above

(b)

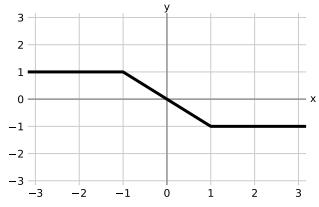


☐ G_1 ☐ G_2 ☐ G_3 ☐ G_4 ☐ G_5 ☐ H_1 ☐ H_2 ☐ H_3 ☐ H_4 ☐ H_5 ○ None of the above

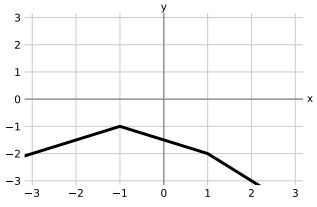


☐ G_1 ☐ G_2 ☐ G_3 ☐ G_4 ☐ G_5 ☐ H_1 ☐ H_2 ☐ H_3 ☐ H_4 ☐ H_5 ○ None of the above

(c)



☐ G_1 ☐ G_2 ☐ G_3 ☐ G_4 ☐ G_5 ☐ H_1 ☐ H_2 ☐ H_3 ☐ H_4 ☐ H_5 ○ None of the above



☐ G_1 ☐ G_2 ☐ G_3 ☐ G_4 ☐ G_5 ☐ H_1 ☐ H_2 ☐ H_3 ☐ H_4 ☐ H_5 ○ None of the above