Introduction to Machine Learning
Problem Set: Neural Networks

Summer 2021

1. **This is not a programming question - compute the answers by hand, with a calculator, not by writing code. Show your work for each part.**

   Consider a neural network for classification with two features at the input, two hidden-layer nodes in a single hidden layer, and one output node. There is a sigmoid activation function at the hidden layer nodes *and* at the output layer node. The weights are indicated on the graph below. (The bias input is always +1.)

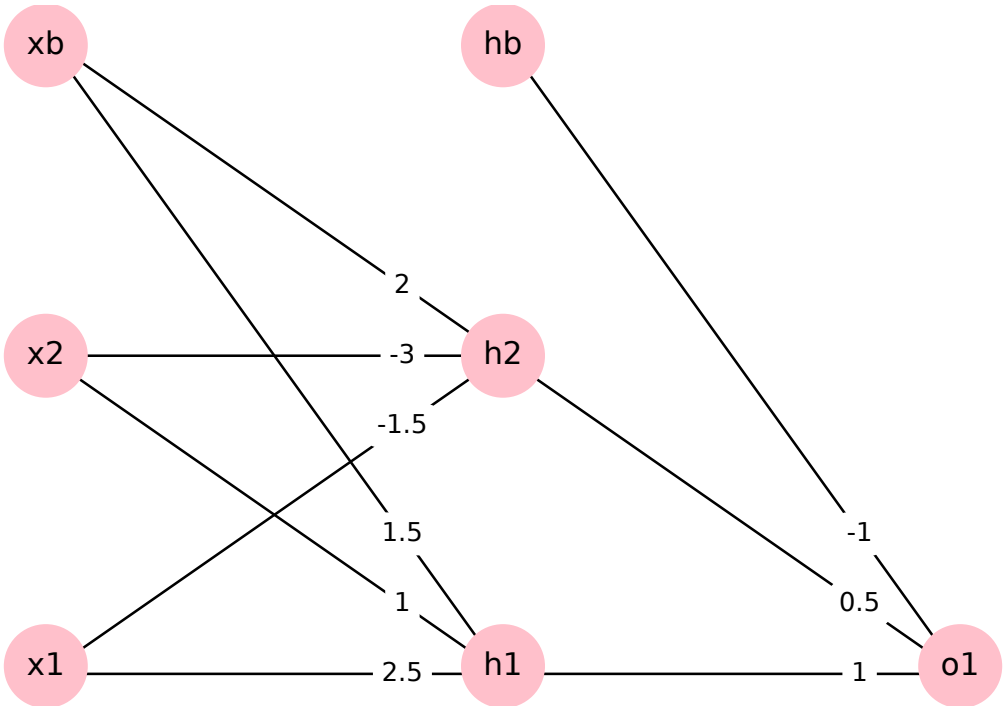

Figure 1: Network for neural network training example.

(a) For the input $x_1 = 0, x_2 = 1$, do a forward pass on the network and compute the output.

(b) Suppose the true value is $y = 1$ when the input is $x_1 = 0, x_2 = 1$,. Compute the gradients with respect to the weights, using the backpropagation algorithm and the squared loss function:

$$L(\theta) = \frac{1}{2} \sum_{N} (y - u_O)^2$$

Show your work.

Note that because of the sigmoid activation function at the output node, $u_O = \sigma(z_O)$, you will compute the backpropagation error at the output with

$$\delta_O = \frac{\partial L}{\partial z_O} = \frac{\partial L}{\partial u_O} \frac{\partial u_O}{\partial z_O}$$

(Note: we would usually use the binary cross entropy loss function for a classification problem, but the math is easier to work with for the squared error loss! So in this case, we'll use squared error.)

(c) Compute the updated weights for both the hidden layer and the output layer by performing one step of gradient descent. Use a learning rate of 0.1.