Perceptron and Multi-layer Perceptron

Question 1. A perceptron has two input units, a unipolar step function, weights w_1 and w_2, and a threshold θ (thus, θ can be considered as a weight for an extra input which is always set to -1).

a) Let $w_1 = 0.2$, $w_2 = -0.5$, and $\theta = -0.2$. What is the actual output for the input pattern $\mathbf{x} = [1, 1]^T$?

b) The perceptron is now trained using the learning rule $\Delta\mathbf{w} = \eta(d - y)\mathbf{x}$, where $\mathbf{x}$ is the input vector, η is the learning rate, **w** is the weight vector, *d* is the desired output, and *y* is the actual output. What are the new values of the weights and threshold after one step of training with the input vector $\mathbf{x} = [0, 1]^T$ and desired output 1, using a learning rate $\eta = 0.2$?

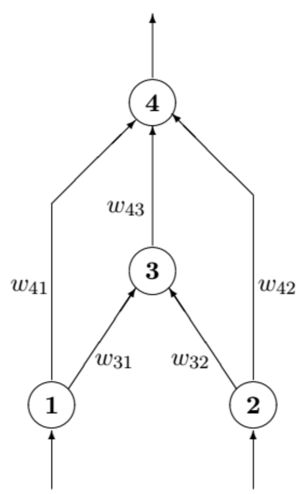
c) What is the equation of the decision line for the perceptron before training? Choose the correct answer from the following options.

A. $x_2 = 0$
B. $x_2 = 0.667x_1$
C. $x_2 = 0.400x_1 + 0.400$
D. $x_2 = 0.667x_1 + 1.333$
E. $x_2 = 0.800x_1 + 0.800$

d) What is the equation of the decision line for the perceptron before training? Choose the correct answer from the following options.

A. $x_2 = 0$
B. $x_2 = 0.667x_1$
C. $x_2 = 0.400x_1 + 0.400$
D. $x_2 = 0.667x_1 + 1.333$
E. $x_2 = 0.800x_1 + 0.800$

Question 2. In the network shown aside, all the units have binary inputs (0 or 1), unipolar step functions and binary outputs (0 or 1). The weights for this network are $w_{31} = 1$, $w_{32} = 1$, $w_{41} = 1$, $w_{42} = 1$ and $w_{43} = -2$. The threshold of the hidden unit (3) is 1.5 and the threshold of the output unit (4) is 0.5. The threshold of both input units (1 and 2) is 0.5, so the output of these units is the same as the input.



a) Which Boolean functions can be computed by this network?

b) Is the following statement true or false? "A multi-layer perceptron with separate layers of weights could perform exactly the same function as the neural shown aside."

Question 3. A single-layer perceptron has 3 input units and 3 output units. How many weights does this network have?

Question 4. Is the following statement true or false? "The Boolean function [(a AND b AND c) OR d] can be solved with a perceptron of binary output." If TRUE, draw the perceptron and note the specific values of weights, w_1, w_2,…, w_n and threshold w_0. If FALSE, explain why.

Question 5. Consider a single sigmoid unit with three inputs, x_1, x_2 and x_3, and output y. The output is given by $y = g(w_0 + w_1x_1 + w_2x_2 + w_3x_3)$ where $g(z) = 1 / (1 + e^{-z})$. We input values of either 0 or 1 for each of the inputs. Assign values to weights, w_0, w_1, w_2 and w_3 so that the output y is greater than 0.5 if an only if the following Boolean condition is satisfied: **[(x_1 AND x_2) OR x_3]**. Note: Your assignment of values to weights should be in multiples of 0.25. That is, try weight values from the set {-1, -0.75, -0.5, -0.25, 0, +0.25, ..., +1}.

Question 6. A multi-layer feedforward network has 5 input units, a first hidden layer with 4 units, a second hidden layer with 3 units, and 2 output units. How many weights does this network have?

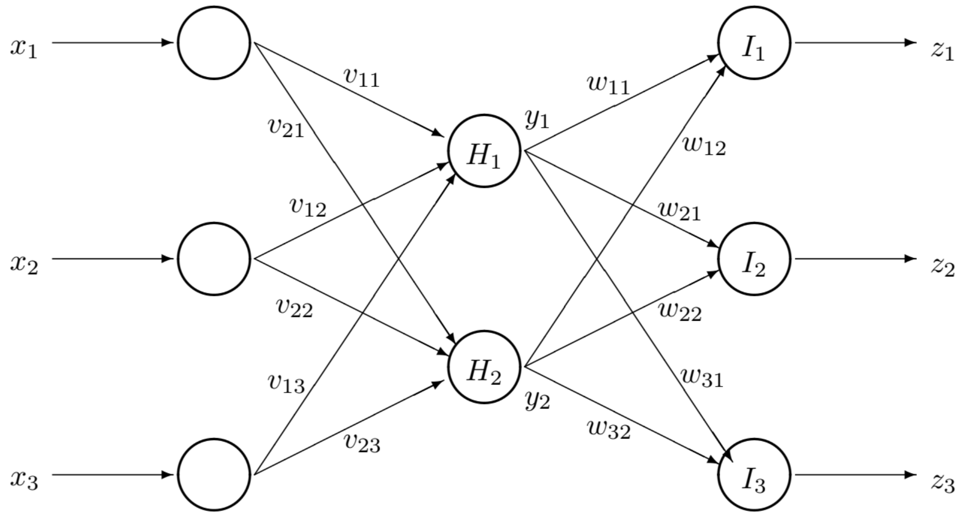
Question 7. Why is it not a good idea to have linear activation functions in the hidden units of a multi-layer feedforward network?

A. This type of function is smooth but non-decreasing.
B. It is not possible to calculate the derivative for this type of function.
C. A multi-layer network with this type of activation functions would be no more powerful than a single-layer network.
D. When the errors are backpropagated, this function gives no information about the amount of adjustment required to the weights of the hidden units.

Question 8. Which of the following techniques is NOT a strategy for dealing with local minima in the back-propagation algorithm?

A. Add random noise to the weights during training.
B. Add random noise to the training examples at each epoch of training.
C. Repeat the training process several times with new random weights, and pick the network that gives the best generalization performance.
D. Train and test an ensemble of networks, then use the average of the network outputs.
E. After each epoch of training, test the network with an independent data set and stop the training when the generalization error starts to increase.

Question 9. Consider the following feedforward network with one hidden layer of units.



The input vector to the network is $x = [x_1, x_2, x_3]^T$, the vector of hidden layer outputs is $y = [y_1, y_2]^T$, the vector of actual outputs is $z = [z_1, z_2, z_3]^T$, and the vector of desired outputs is $t = [t_1, t_2, t_3]^T$. The network has the following weight vectors:

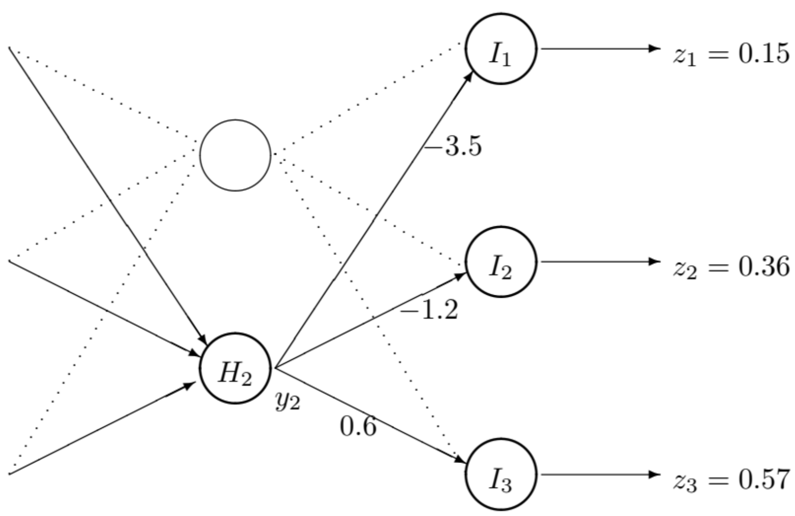
$$\mathbf{v}_1 = \begin{bmatrix} -2.0 \\ 2.0 \\ -2.0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1.0 \\ 1.0 \\ -1.0 \end{bmatrix}, \quad \mathbf{w}_1 = \begin{bmatrix} 1.0 \\ -3.5 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} 0.5 \\ -1.2 \end{bmatrix} \quad \text{and} \quad \mathbf{w}_3 = \begin{bmatrix} 0.3 \\ 0.6 \end{bmatrix}$$

Assume that all units have sigmoid activation functions given by $f(x) = 1 / 1 + \exp(-x)$ and that each unit has a bias $\theta = 0$ (zero).

a) If the network is tested with an input vector $x = [1.0, 2.0, 3.0]^T$ then what will the activation H_1 of the first hidden neuron be?

b) With the same input as above, what is the activation I_3 of the third output neuron?

Question 10. The following figure shows part of the neural network described in the above question (i.e., the same weights, activation functions and bias values). A new input pattern is presented to the network and training proceeds as follows. The actual outputs of the network are given by $z = [0.15, 0.36, 0.57]^T$ and the corresponding target outputs are given by $t = [1.00, 1.00, 1.00]^T$. The weights w_{12}, w_{22} and w_{32} are also shown below.



a) What is the error for each of the output units?

b) What is the error for hidden unit 2 given that its activation for the pattern being processed is currently $y_2 = 0.5$?