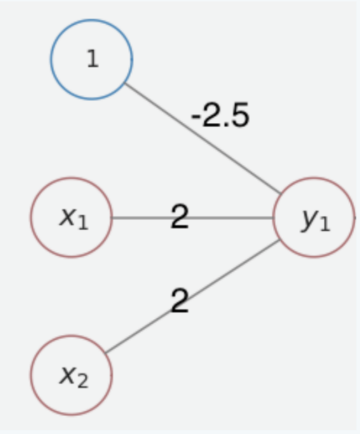
## Week 1a – Introduction to Deep Learning

1. Which of the following logical operations does the following perceptron represent? Note that ​ and ​ are binary.



* 1. ​ (Logical AND outputs 1 only when both inputs ​ and ​ are 1. For every other case, AND should output 0. The weights are the same for both inputs and ​. is negative except for when both ​ and ​ are 1.

1. The following training set can be classified exactly by a single perceptron.

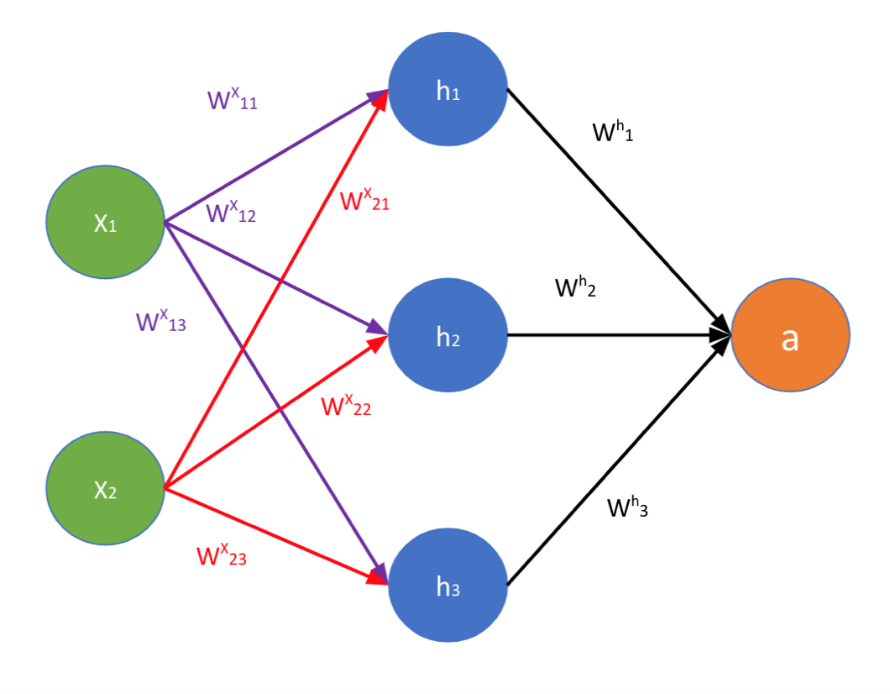
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 0 | 1 | 0 | 1 |
|  | 0 | 0 | 1 | 1 |
|  | 0 | 0 | 1 | 1 |

* 1. True. (A single perceptron is a linear classifier. A linear classifier can properly separate these data points.)

1. Which of the following can be guaranteed to behave as Activation functions and are not difficult to train?
   1. Sigmoid
   2. Step function
   3. Tanh
   4. ReLu
2. Consider your model is being trained using the Perceptron algorithm. Let W be the current Weight and x be a misclassified instance. Which of the following statements are valid?
   1. If x is a positive instance classified as negative, then W = W + x
   2. If x is a negative instance classified as positive, then W = W - x
3. What are the possible hyperparameters that can be tuned for a Multi-Layered Perceptron (MLP)?
   1. Number of hidden layers
   2. Activation functions
   3. Number of nodes in a layer

## Week 1b – Backpropagation

1. Which of the following statements is NOT true?
   1. Back propagation can always find the global optimum regardless of weights initialization.
2. Forward Propagation: Consider a neural network shown below.



It has an input , a hidden layer , and an output layer . The layers and has a sigmoid activation function.

The input is

The weight matrix for is

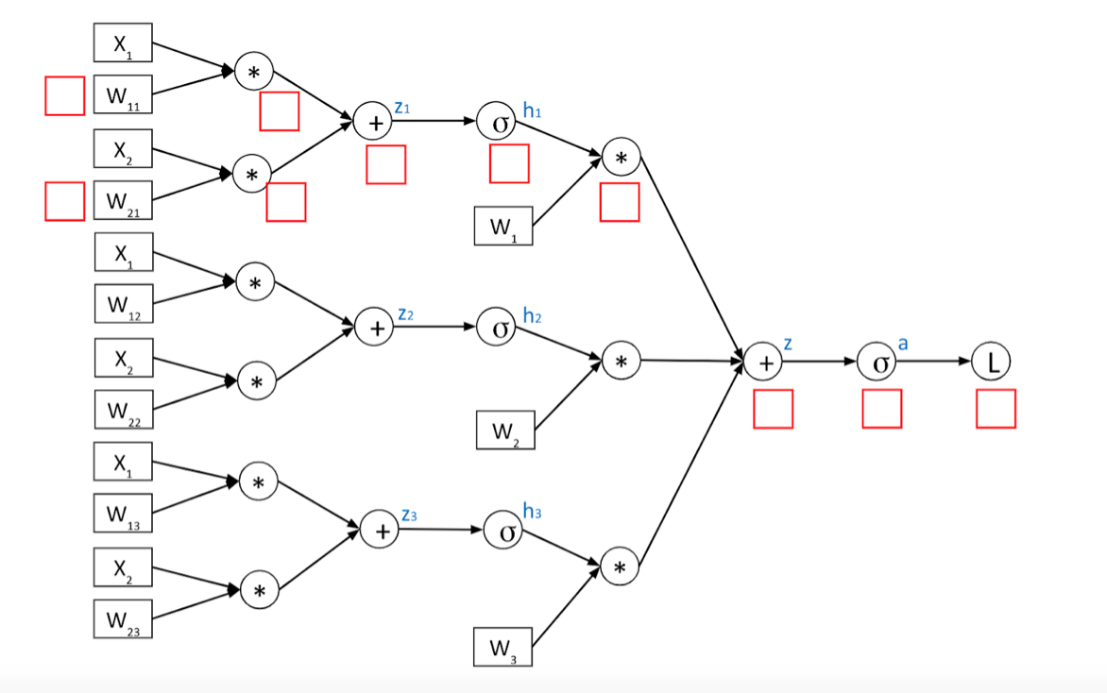
The weight matrix for is

What are the output values in the hidden layer output vector ?

* 1. What is the first element h1?
  2. What is the second element h2?
  3. What is the third element h3?

1. What is the output value at ?
2. Suppose we have a cross-entropy loss function for binary classification:

, where is the probability out from the output layer activation function. The computation graph of the network is presented below. The blue letters are intermediate variable labels to help you understand the connection between the network architecture graph above and the computation graph.



* 1. When , what is the gradient of the loss function with respect to ? (You may use the property of ).
  2. When and the learning rate , what is the updated weight ? (Calculate the updated weight using the old weight and the learning rate as follows: