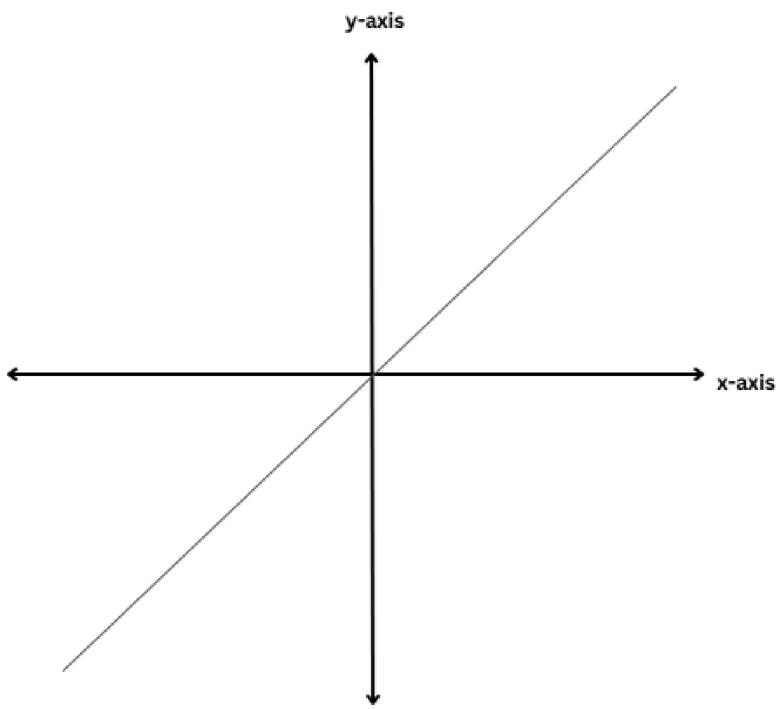
CSE 571 Midterm Review - Week 2 Module

1. Which of the following options will **always decrease** the training time of a Neural Network:
    a. Increasing the number of epochs
    b. Increasing the size of the training dataset
    c. Decreasing the number of neurons in each hidden layer
    d. Early-Stopping

early-stopping does not always decrease training time since training time depends on the convergence rate
early-stopping may require more iterations since it involves monitoring the validation loss and stopping training when it starts to increase - this means that the training process may require more iterations to find the optimal solution

2. Consider a function with the following graph:



Why would this function **NOT** qualify to be an activation function for Neural Networks?
    a. The function is linear
    b. The function is non-differentiable, y = x
    c. The function is non-monotonic

3. Which of the following is a desired property of an Activation Function for Neural Networks:
    a. Non-Differentiable
    b. Non-Monotonic
    c. Prone to Vanishing Gradients
    d. Computationally inexpensive

4. You have a network with current parameters (weights) [6.2, 3.8, 4.7], and a learning rate of 0.25. After one update step of gradient descent the network's parameters (weights) were updated to [6.7, 2.8, 3.45]. The gradient values used for this update step are:
    a. [2, 4, -5]
    b. [3, 7, -1]
    c. [-2, 4, 5]
    d. [2, -4, -5]

```
new_weight = old_weight - learning_rate * gradient_values
[6.7, 2.8, 3.45] = [6.2, 3.8, 4.7] - 0.25 * [a, b, c]
```

5. Which of the following best describes the differences between Softmax and Sigmoid Activation Functions?
    a. Softmax is used for binary classification, while Sigmoid is used for multi-class classification.
    b. Softmax can handle multiple class probabilities, while Sigmoid is limited to binary classification.
    c. Softmax outputs a probability distribution over multiple classes, while Sigmoid provides a single probability for binary classification.
    d. Softmax is suitable for regression tasks, whereas Sigmoid is designed for image classification.

6. Amy is trying to implement the Back-propagation Algorithm for a fully connected feed-forward Neural Network. She completes one forward pass for a particular input and retrieves the output from her Neural Network. Which of the following best describes the next step she has to take?

   `Option (a) suggests dE / dW (CORRECT). Option (c) suggests dW / dE (WRONG).`

   a. Calculate the loss with respect to the output and target values (based on the loss function) and use this to calculate the gradient of the loss function with respect to the network's weights and biases.
   b. Perform another forward pass with a different input to validate the network's performance.
   c. Calculate the loss with respect to the output and target values (based on the loss function) and use this to calculate the gradient of the network's weights with respect to the loss function.
   d. Save the current state of the network if it has already learned the desired output.

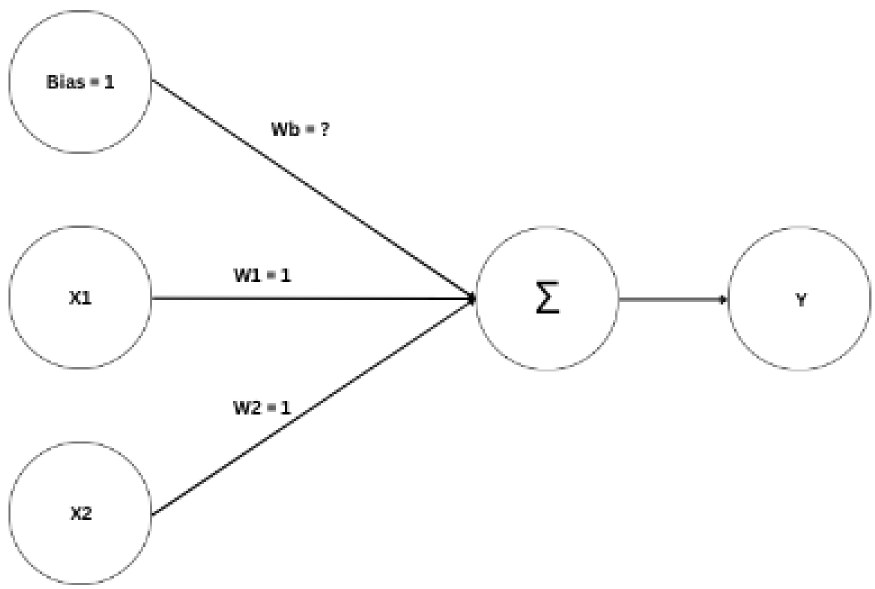
7. When would normalization of the dataset be most applicable prior to training a Neural Network?
   a. When the dataset contains images.
   b. When the dataset consists of categorical data.
   c. When the dataset has features with significantly different scales.
   d. When the dataset is balanced.

8. You have a neural network with an input layer of 4 neurons and an output layer with 1 neuron. Suppose the inputs to the networks are [8.3, -4.7, -2.5, 1.7] and the final output is 16.1. The network's parameters (weights) have missing values [2.0, x, 6.5, x]. What is the value of x?
   a. 5.25
   b. 6.38
   c. -6.38
   d. -5.25

9. Which of the following best describes the differences between Cross-Entropy Loss Function and Mean-Squared Error Loss Function?
    a. Cross-Entropy Loss is used for regression tasks, while Mean-Squared Error is used for classification tasks.
    b. Cross-Entropy Loss measures the difference between predicted probabilities and true labels, while Mean-Squared Error measures the squared difference between predicted and true values.
    c. Cross-Entropy Loss is suitable for problems with continuous output, while Mean-Squared Error is used when dealing with discrete class labels.
    d. Cross-Entropy Loss is less sensitive to outliers compared to Mean-Squared Error.

10. What is the objective function we are trying to solve using the Back Propagation Algorithm?
    a. Minimizing the number of hidden layers in the neural network.
    b. Maximizing the accuracy of the model on the training data.
    c. Minimizing the cost or loss function to make predictions closer to the true target values.
    d. Maximizing the complexity of the neural network to fit the training data perfectly.

11. 2-Input Logic OR Gate Construction with a Perceptron:



Consider the perceptron displayed above and the following pointers:

i. Truth Label TRUE (1) will have an Input Value **+1** to the network.

ii. Truth Label FALSE (0) will have an Input Value **-1** to the network.

iii. Y = 1 if the Sum > 0   Y = 0 if the Sum <= 0

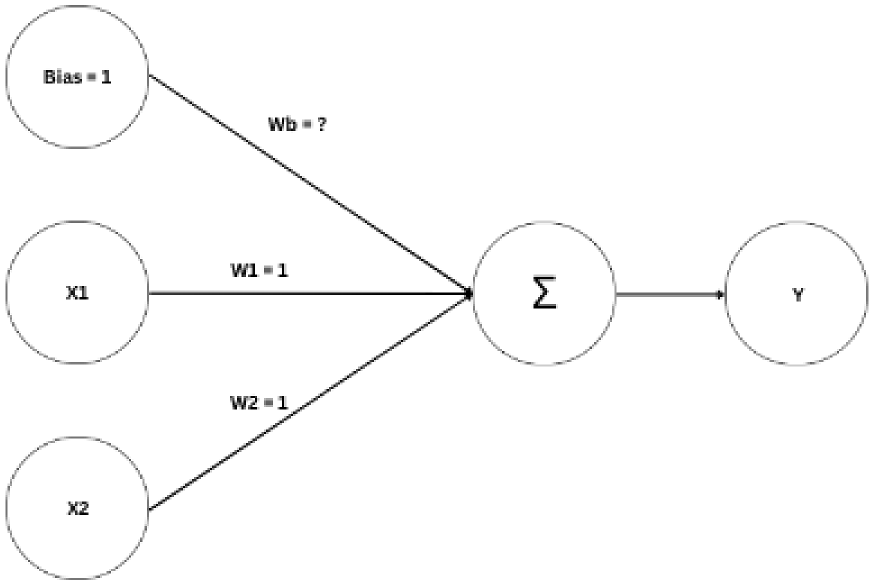
Among the following options, pick the **minimum possible value** for **Wb (Bias Weight)** which will convert the above 2-Input Perceptron to an **2-Input Logic OR Gate**
(NOTE: Make sure you consider all possible input combinations)

a. Wb = 1
b. Wb = -0.1
c. $\boxed{\text{Wb = 0.1}}$
d. Wb = 0

```
logic OR gate

(t OR t) -> t
(t OR f) -> t
(f OR f) -> f
```

12. 2-Input Logic AND Gate Construction with a Perceptron:



Consider the perceptron displayed above and the following pointers:

i. Truth Label TRUE (1) will have an Input Value **+1** to the network.

ii. Truth Label FALSE (0) will have an Input Value **-1** to the network.

iii. Y = 1 if the Sum > 0 Y = 0 if the Sum <= 0

Among the following options, pick the **minimum possible value** for **Wb (Bias Weight)** which will convert the above 2-Input Perceptron to an **2-Input Logic AND Gate**
(NOTE: Make sure you consider all possible input combinations)

a. Wb = 0.1
b. Wb = -1.9
c. Wb = -0.1
d. Wb = -2.0

```
logic AND gate

(t AND t) -> t
(t AND f) -> f
(f AND f) -> f
```