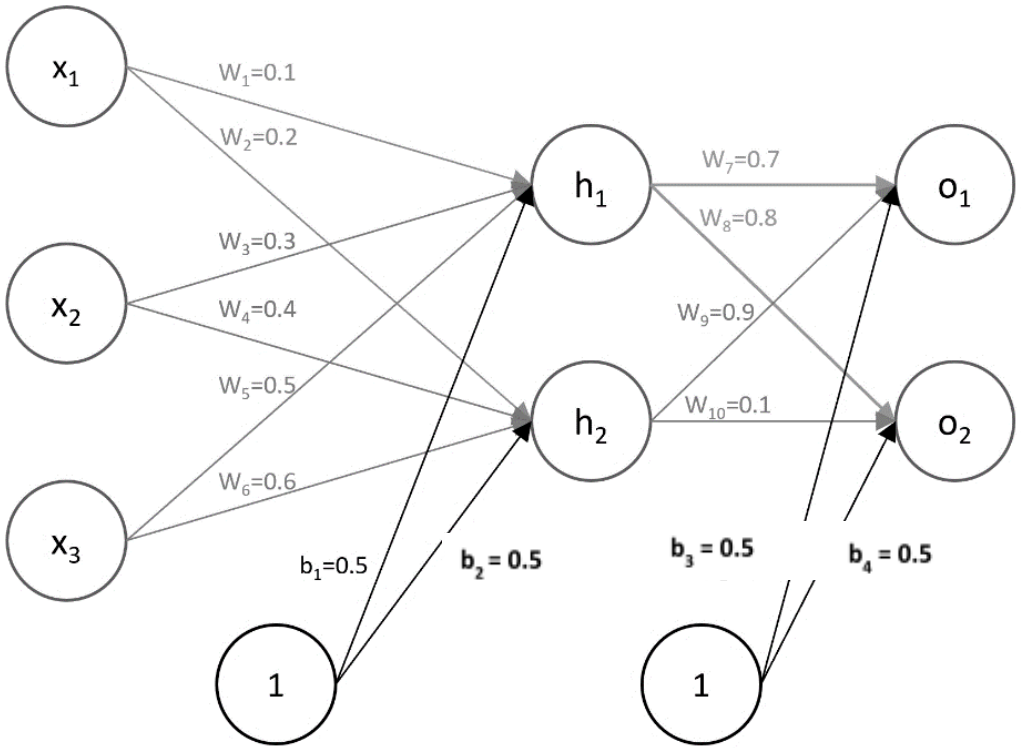
QUESTION: A Multi-layer feed-forward neural network with initialization of weights is given below.



a) Do a forward pass and compute the output at O_1 and O_2. Use linear activation function at hidden layer $\boldsymbol{h_1}$, and $\boldsymbol{h_2}$ and sigmoid activation function at $\boldsymbol{O_1}$ and $\boldsymbol{O_2}$. All biases are 0.5, the input values are x1 = 1, x2 = 4, x3 = 5 and target values are t1= 0.1, t2 = 0.05. Show all the working.

$$net\_h_1 = (0.1\times1) + (0.3\times4) + (0.5\times5) + (0.5\times1) = 4.3$$

$$O(net\_h_1) = 4.3$$

$$net\_h_2 = (0.2\times1) + (0.4\times4) + (0.6\times5) + (0.5\times1) = 5.3$$

$$O(net\_h_2) = 5.3$$

$$net\_O_1 = (0.7\times4.3) + (0.9\times5.3) + (0.5\times1) = 8.28$$

$$O(net\_O_1) = \frac{1}{1+e^{-8.28}} = 0.9997$$

$$net\_O_2 = (0.8\times4.3) + (0.1\times5.3) + (0.5\times1) = 4.47$$

$$O(net\_O_2) = \frac{1}{1+e^{-4.47}} = 0.9887$$

b) Do a backward pass (backpropagation) and compute update in weights $\boldsymbol{b_1}$, $\boldsymbol{w_4}$ and $\boldsymbol{w_{10}}$. Use learning rate η=0.01. Show all the working.

$$new\_w_{10} = 0.1 + 0.01 \times (0.05 - 0.9887) \times 0.9887 \times (1 - 0.9887) \times 5.3$$

$$new\_w_{10} = 0.1 + 0.01 \times (-0.9387) \times 0.9887 \times (0.0113) \times 5.3$$

$$new\_w_{10} = 0.1 - 0.000556 = 0.0994$$

$$new\_w_{10} = 0.09944$$

$$new\_w_4 = 0.4 + 0.01 \times [(0.1 - 0.9997) \times 0.9997 \times (1 - 0.9997) \times 0.9 + (0.05 - 0.9887) \times 0.988$$

$$new\_w_4 = 0.4 + 0.01 \times [-0.000243 - 0.00105] \times 4$$

$$new\_w_4 = 0.4 - 0.0000517 = 0.39995$$

$$new\_w_4 = 0.39995$$

$$new\_b_1 = 0.5 + 0.01 \times [(0.1 - 0.9997) \times 0.9997 \times (1 - 0.9997) \times 0.7 + (0.05 - 0.9887) \times 0.988$$

$$new\_b_1 = 0.5 + 0.01 \times [-0.000189 - 0.00839]$$

$$new\_b_1 = 0.5 - 0.0000858 = 0.49991$$

$$new\_b_1 = 0.49991$$

NN and GA Concepts:

a) Why it is not a good idea to initialize all the weights of a neural network with zeros?
Initializing weights of a neural network with the same value (irrespective of it being zero) makes it difficult for the neural network to converge. If all weights are the same, all hidden units are also the same. In the case of zeros, all hidden neuron values are also zero. Therefore, it is plausible not to have symmetry between any neurons, thus, we initialize all weights randomly.

b) Why activation functions are used in neural networks?
Activation functions introduce non-linearity in the neural networks. Without them, the network cannot learn a non-linear decision boundary.

c) When not to use multilayered neural network?
When problem is linearly separable and simple.

d) When is it a must to use softmax activation function?
When we are dealing with multiclass classification and need probabilities at the output layer.

e) What is the mutation rate?
mutation rate is the probability of how many chromosomes should be mutated in one generation. It is a number between 0 and 1
f). Mutation using a coin toss
In the case of [101010] individual
For each bit a random number between 0-1 is generated, if the number is >=0.5 the bit will be flipped else remains the same.