CSPB 4622 - Truong - Machine Learning

Dashboard / My courses / 2247:CSPB 4622 / 28 October - 3 November / Quiz 10-2. backprop

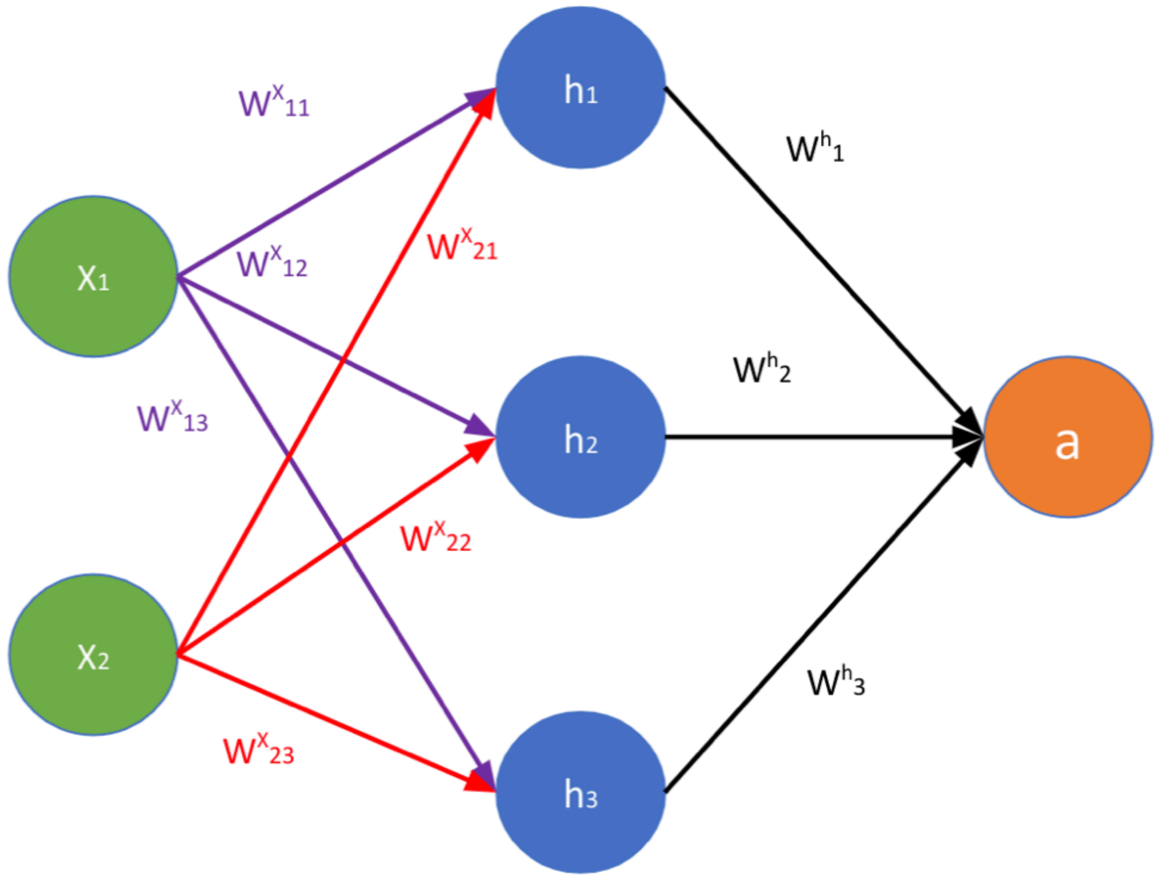
Started on	Friday, 8 November 2024, 12:16 AM
State	Finished
Completed on	Friday, 8 November 2024, 12:16 AM
Time taken	44 secs
Marks	7.00/9.00
Grade	**7.78** out of 10.00 (**78**%)

Question 1

Correct

Mark 4.00 out of 4.00

Consider a neural network shown below.



It has an input $X = (X_1, X_2)$,
a hidden layer $h = (h_1, h_2, h_3)$,
and an output $a = (a)$
The layers h and a has sigmoid activation function.
Input X is

$$X = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

The weight matrix for X-h is

$$W^X = \begin{pmatrix} 1/2 & 1/2 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$$

The weight matrix for h-a is

$$W^h = \begin{pmatrix} -1 \\ 1/2 \\ 1/2 \end{pmatrix}$$

Fill the values for the hidden layer and output layer neuron outputs.

h1 :

0.5

h2 :

0.2689

h3 :

0.7311

a :

0.5

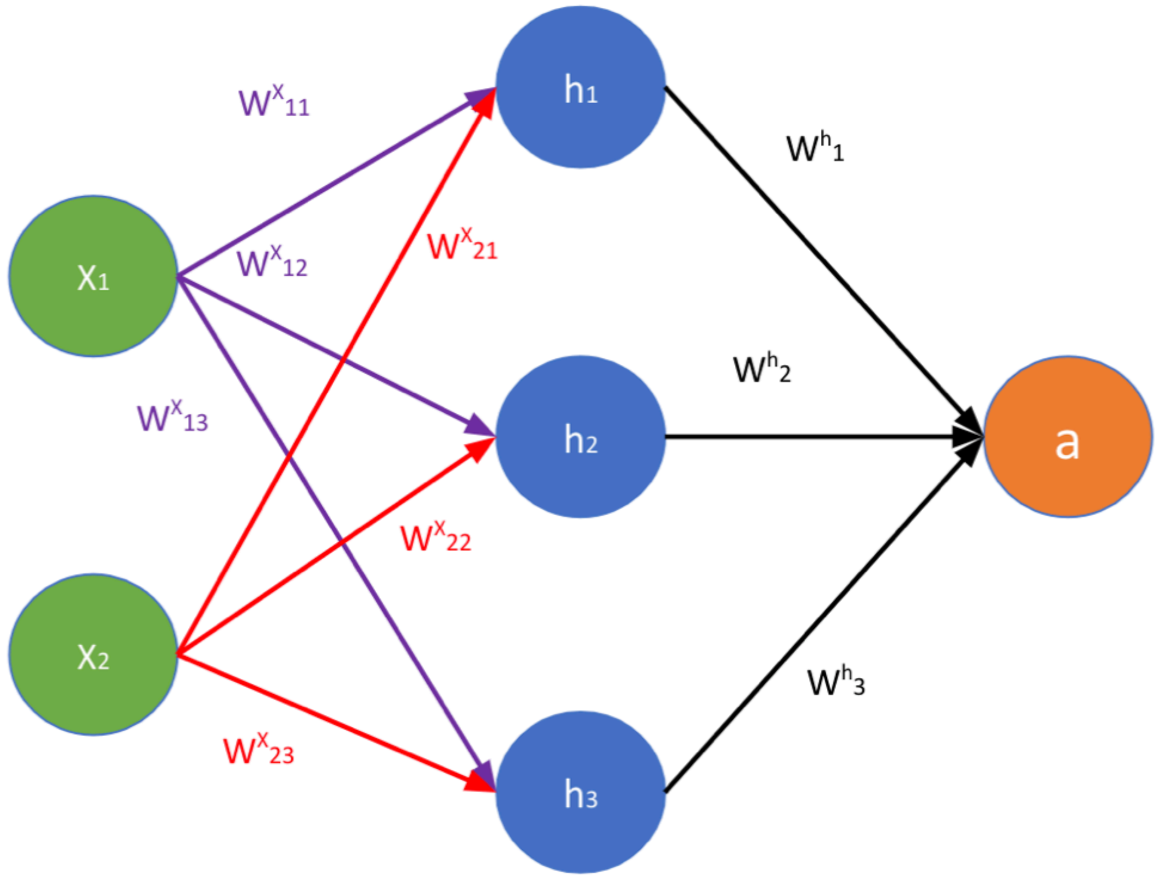
Your answer is correct.

Question **2**

Partially correct

Mark 2.00 out of 4.00

Consider a neural network shown below.



It has an input $X = (X_1, X_2)$,
a hidden layer $h = (h_1, h_2, h_3)$,
and an output $a = (a)$
The layers h and a has sigmoid activation function.
Input X is

$$X = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

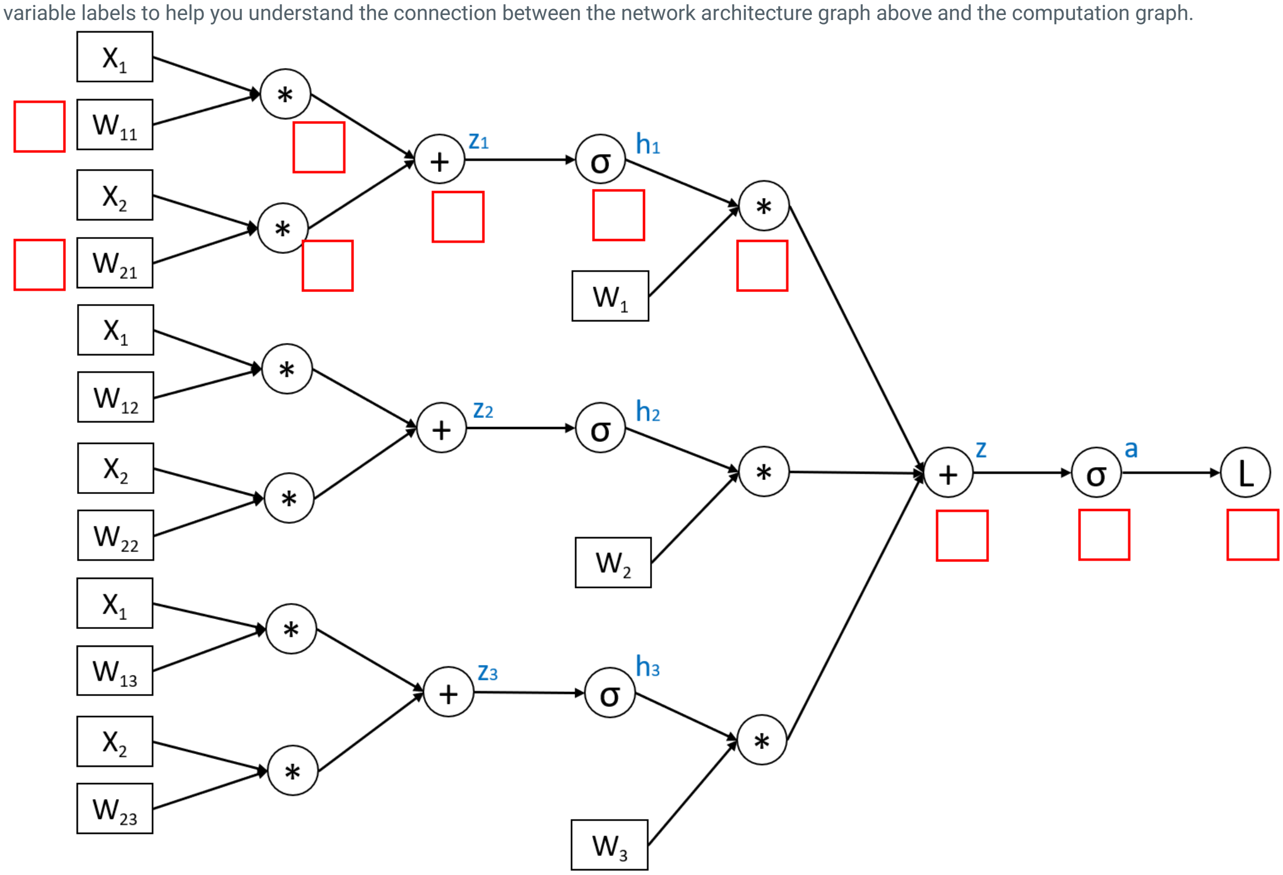
The weight matrix for X-h is

$$W^X = \begin{pmatrix} 1/2 & 1/2 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$$

The weight matrix for h-a is

$$W^h = \begin{pmatrix} -1 \\ 1/2 \\ 1/2 \end{pmatrix}$$

Consider we have a cross-entropy loss function for binary classification: L=−[y ln(a)+(1−y) ln(1−a)], where a is the probability out from the output layer activation function. We've built a computation graph of the network as shown below. The blue letters below are intermediate variable labels to help you understand the connection between the network architecture graph above and the computation graph.



(a) When $y = 1$, what is the gradient of the loss function w.r.t. W_{11} ?

0.125

✔

(b) With the same condition $y = 1$ and the learning rate 1/2, what is the updated weight W_{21} (new)?

0.125

✖

Write your answer to three decimal places. Note: Please use the computation graph method. One can calculate the gradient directly using chain rules, but if the computation graph is not used at all, it will not score properly. Try to fill the red boxes above. This question does not need coding and the answer can be easily obtained analytically.

Hint: You may use the property of $\frac{\partial\,\sigma(z)}{\partial\, z} = \sigma\,(1 - \sigma)$

Your answer is partially correct.

1 of your answers is correct.

$W_{21} \leftarrow W_{21} - \eta \frac{\partial □}{\partial W_{21}} = 1/2 - (1/2)(-1/8) = 9/16$

Question **3**

Correct

Mark 1.00 out of 1.00

Which of the following statements is NOT true?

- (•) a. Back propagation can always find the global optimum regardless of weights initialization. ✔
- ( ) b. Back propagation is more computationally complex than forward propagation.
- ( ) c. Forward propagation is necessary for computing the activations at each layer.
- ( ) d. Vanishing gradients can occur and lead to slow convergence.

Your answer is correct. Regardless of weight initialization, backpropagation cannot always find the global optimum.

The correct answer is: Back propagation can always find the global optimum regardless of weights initialization.