Congratulations! You passed!

Grade received 100%

Latest Submission Grade 100%

To pass 83% or higher

Go to next item

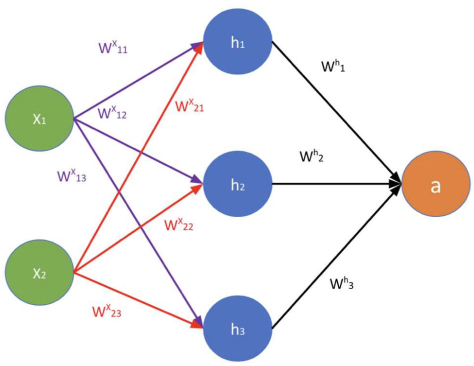
1. Which of the following statements is NOT true? — 7 / 7 points

- ○ Back propagation is more computationally complex than forward propagation.
- ○ Forward propagation is necessary for computing the activations at each layer.
- ○ Vanishing gradients can occur and lead to slow convergence.
- ◉ Back propagation can always find the global optimum regardless of weights initialization.

Correct

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

2. Forward Propagation: Consider a neural network shown below. — 2 / 2 points



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

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

What are the output values in the hidden layer output vector h? Give each answer to three decimal places.

What is the first element in the hidden layer output vector (h1) ?

0.500

Correct
Your answer is correct.

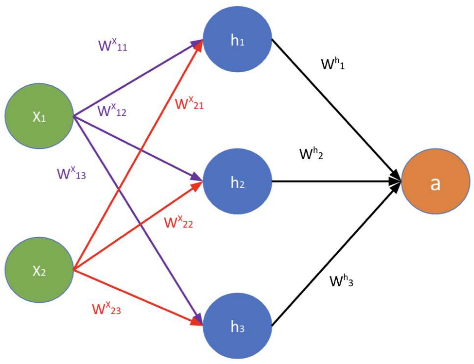
Good notation examples for written answers:
$W_{1,2}^{X}$ for the weight between X_1 and h_2
W_1^{h} for the weight between h_1 and a

Good notation examples for coding answers:
`Wx` for the W^X matrix
`Wh` for the W^h matrix
`x` for input vector
`h` for hidden layer's output vector
`a` for output layer's output

$z^X = W^X \cdot X = [0, -1, 1]$
$h = \sigma(z^X) = [1/2, 1/(1+e), e/(1+e)]$
$z^h = W^h \cdot h = -1/2 + 1/2 = 0$
$a = \sigma(0) = 1/2$

Note to graders:
h = $[\sigma(0), \sigma(-1), \sigma(1)]$ also valid.

3. Forward Propagation: Consider a neural network shown below. 2 / 2 points



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

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

What are the output values in the hidden layer output vector h? Give each answer to three decimal places.
What is the second element in the hidden layer output vector (h2)?

0.268

Correct
Your answer is correct.

Good notation examples for written answers:

$W_{1,2}^{X}$ for the weight between X_1 and h_2

W_1^{h} for the weight between h_1 and a

Good notation examples for coding answers:

`Wx` for the W^X matrix

`Wh` for the W^h matrix

`x` for input vector

`h` for hidden layer's output vector

`a` for output layer's output

$z^X = W^X \cdot X = [0, -1, 1]$

$h = \sigma(z^X) = [1/2, 1/(1+e), e/(1+e)]$

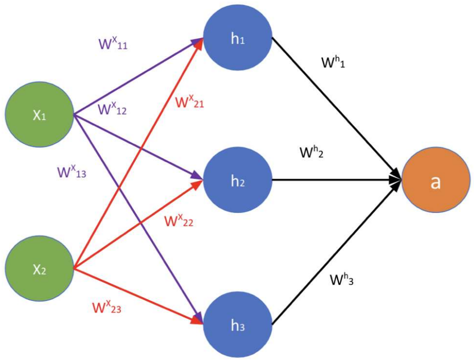
$z^h = W^h \cdot h = -1/2 + 1/2 = 0$

$a = \sigma(0) = 1/2$

Note to graders:

h = $[\sigma(0), \sigma(-1), \sigma(1)]$ also valid.

4. Forward Propagation: Consider a neural network shown below. 2 / 2 points



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

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

What are the output values in the hidden layer output vector h? Give each answer to three decimal places.

What is the third element in the hidden layer output vector (h3)?

0.731

Correct

Your answer is correct.

Good notation examples for written answers:

$W^X_{1,2}$ for the weight between X_1 and h_2

W^h_1 for the weight between h_1 and a

Good notation examples for coding answers:

`Wx` for the W^X matrix

`Wh` for the W^h matrix

`x` for input vector

`h` for hidden layer's output vector

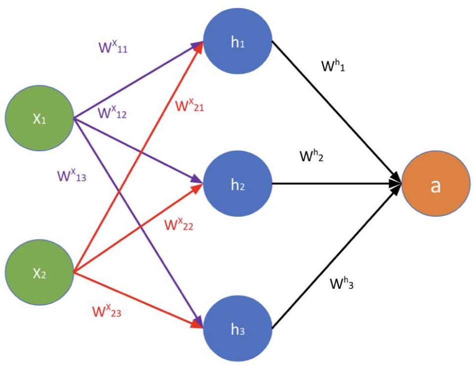
`a` for output layer's output

$z^X = W^X \cdot X = [0, -1, 1]$
$h = \sigma(z^X) = [1/2, 1/(1+e), e/(1+e)]$
$z^h = W^h \cdot h = -1/2 + 1/2 = 0$
$a = \sigma(0) = 1/2$

Note to graders:

h = $[\sigma(0), \sigma(-1), \sigma(1)]$ also valid.

5. Forward Propagation: Consider a neural network shown below. 2 / 2 points



What is the output value at a? Give your answer to three decimal places.

0.499

Correct

Your answer is correct.

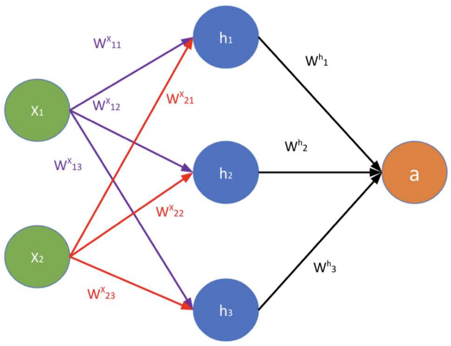
Good notation examples for written answers:
$W_{1,2}^X$ for the weight between X_1 and h_2
W_1^h for the weight between h_1 and a

Good notation examples for coding answers:
`Wx` for the W^X matrix
`Wh` for the W^h matrix
`x` for input vector
`h` for hidden layer's output vector
`a` for output layer's output

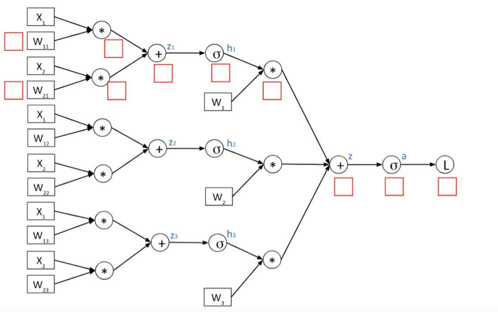
$$z^X = W^X \cdot X = [0, -1, 1]$$
$$h = \sigma(z^X) = [1/2, 1/(1+e), e/(1+e)]$$
$$z^h = W^h \cdot h = -1/2 + 1/2 = 0$$
$$a = \sigma(0) = 1/2$$

Note to graders:
h = $[\sigma(0), \sigma(-1), \sigma(1)]$ also valid.

6. Consider a neural network shown below. 6 / 6 points



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.



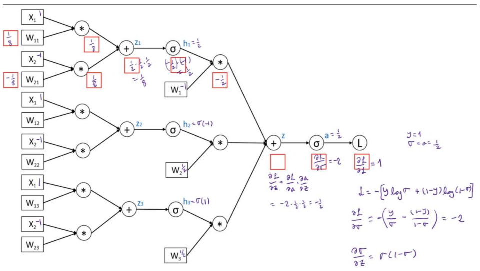
When y=1, what is the gradient of the loss function w.r.t. $W11$? Write your answer to three decimal places.

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 $\frac{\partial \sigma(z)}{\partial z} = \sigma\,(1-\sigma)$

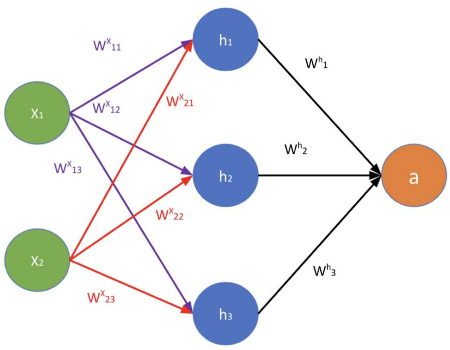
0.125

Correct
Your answer is correct.

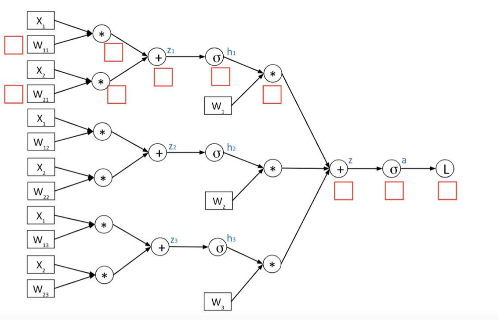


$$\frac{\partial L}{\partial W11} = 1/8$$

7. Consider a neural network shown below. 6 / 6 points



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.



With the same condition (y=1) and the learning rate η=1/2, what is the updated weight $W21$ (new)? Write your answer to three decimal places.
Note: Please use the computation graph method. One can calculate the gradients 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 in the computation graph. This question does not need coding and the answer can be easily obtained analytically.

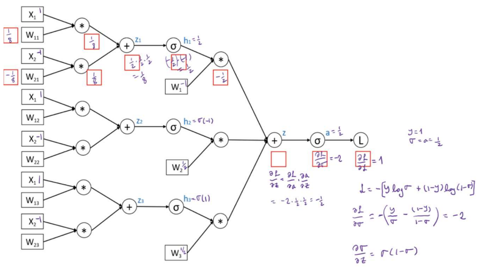
Hint: You may use the property of $\frac{\partial\, \sigma(z)}{\partial\, z} = \sigma\,(1-\sigma)$

Calculate new weight using the old weight and learning learning as follows:

$$W21 \leftarrow W21 - \eta \frac{\partial L}{\partial W21}$$

0.562

Correct



$$W21 \leftarrow W21 - \eta \frac{\partial L}{\partial W21} = 1/2 - (1/2)\,(-1/8) = 9/16$$