Your grade: 80%

Your latest: **80%** • Your highest: **80%**
To pass you need at least 80%. We keep your highest score.

Next item →

1. Which of the following are true? (Check all that apply.) **1 / 1 point**

- [ ] $w_3^{[4]}$ is the row vector of parameters of the fourth layer and third neuron.
- [ ] $w_3^{[4]}$ is the column vector of parameters of the third layer and fourth neuron.
- [x] $w_3^{[4]}$ is the column vector of parameters of the fourth layer and third neuron.

Correct
Yes. The vector $w_j^{[i]}$ is the column vector of parameters of the i-th layer and j-th neuron of that layer.

- [x] $a^{[2]}$ denotes the activation vector of the second layer.

Correct
Yes. In our convention $a^{[j]}$ denotes the activation function of the j-th layer.

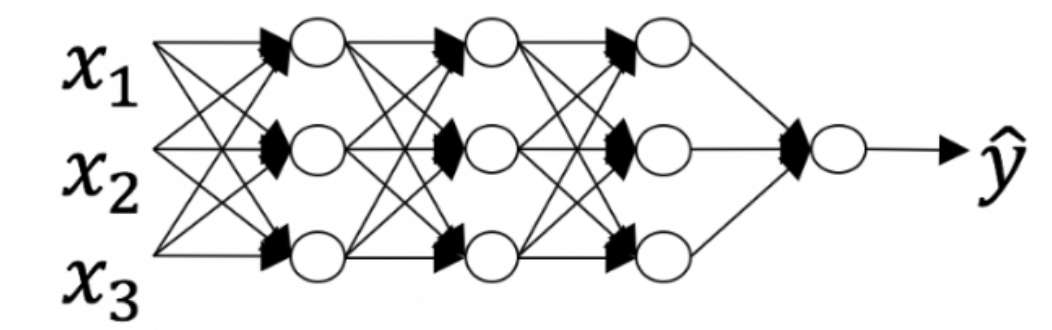
- [ ] $a^{[3](2)}$ denotes the activation vector of the second layer for the third example.
- [ ] $a_3^{[2]}$ denotes the activation vector of the second layer for the third example.

2. The sigmoid function is only mentioned as an activation function for historical reasons. The tanh is always preferred without exceptions in all the layers of a Neural Network. True/False? **1 / 1 point**

- ○ True
- ◉ False

Correct
Yes. Although the tanh almost always works better than the sigmoid function when used in hidden layers, thus is always proffered as activation function, the exception is for the output layer in classification problems.

3. Which of the following represents the activation output of the second neuron of the third layer applied to the fourth example? **1 / 1 point**



- ◉ $a_2^{[3](4)}$
- ○ $a_2^{[4](3)}$
- ○ $a_4^{[3](2)}$
- ○ $a_3^{[4]2}$

Correct

Yes. The superscript in brackets indicates the layer number, the superscript in parenthesis represents the number of examples, and the subscript the number of the neuron.

4. When building a binary classifier for recognizing cats (y=1) vs raccoons (y=0). Is better to use the sigmoid function as activation function for the hidden layers. True/False

1 / 1 point

- ○ True
- ◉ False

Correct
Yes. Using tanh almost always works better than the sigmoid function for hidden layers.

5. Consider the following code:

1 / 1 point

```
#+begin_src python

x = np.random.rand(3, 2)

y = np.sum(x, axis=0, keepdims=True)

#+end_src
```

What will be y.shape?

- ○ (3,)
- ◉ (1, 2)
- ○ (3, 1)
- ○ (2,)

Correct

Yes. By choosing the axis=0 the sum is computed over each column of the array, thus the resulting array is a row vector with 2 entries. Since the option keepdims=True is used the first dimension is kept, thus (1, 2).

6. Suppose you have built a neural network. You decide to initialize the weights and biases to be zero. Which of the following statements is true? **1 / 1 point**

- ○ Each neuron in the first hidden layer will compute the same thing, but neurons in different layers will compute different things, thus we have accomplished “symmetry breaking” as described in the lecture.
- ○ Each neuron in the first hidden layer will perform the same computation in the first iteration. But after one iteration of gradient descent they will learn to compute different things because we have “broken symmetry”.
- ◉ Each neuron in the first hidden layer will perform the same computation. So even after multiple iterations of gradient descent, each neuron in the layer will be computing the same thing as other neurons.
- ○ The first hidden layer’s neurons will perform different computations from each other even in the first iteration; their parameters will thus keep evolving in their own way.

Correct

7. Logistic regression’s weights w should be initialized randomly rather than to all zeros, because if you initialize to all zeros, then logistic regression will fail to learn a useful decision boundary because it will fail to “break symmetry”, True/False? **1 point**

- ◉ True

- False

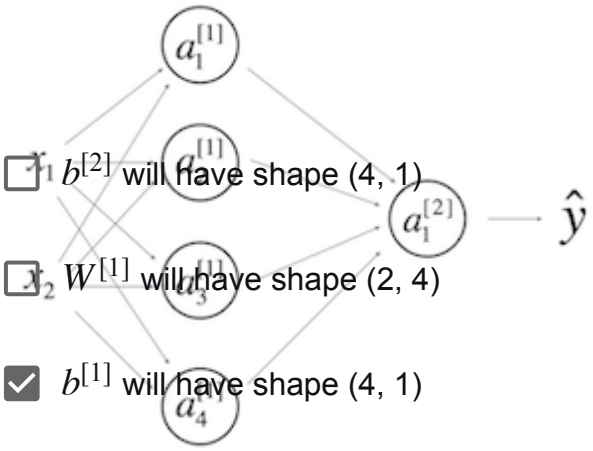
Incorrect
No, Logistic Regression doesn't have a hidden layer. If you initialize the weights to zeros, the first example x fed in the logistic regression will output zero but the derivatives of the Logistic Regression depend on the input x (because there's no hidden layer) which is not zero. So at the second iteration, the weights' values follow x's distribution and are different from each other if x is not a constant vector.

8. Which of the following is true about the ReLU activation functions? **1 / 1 point**

- They are only used in the case of regression problems, such as predicting house prices.
- They are increasingly being replaced by the tanh in most cases.
- They cause several problems in practice because they have no derivative at 0. That is why Leaky ReLU was invented.
- They are the go to option when you don't know what activation function to choose for hidden layers.

Correct

9. Consider the following 1 hidden layer neural network: **1 / 1 point**



- [ ] $b^{[2]}$ will have shape (4, 1)
- [ ] $W^{[1]}$ will have shape (2, 4)
- [x] $b^{[1]}$ will have shape (4, 1)

Correct

- [ ] $b^{[1]}$ will have shape (2, 1)
- [x] $W^{[2]}$ will have shape (1, 4)

Correct

- [x] $W^{[1]}$ will have shape (4, 2)

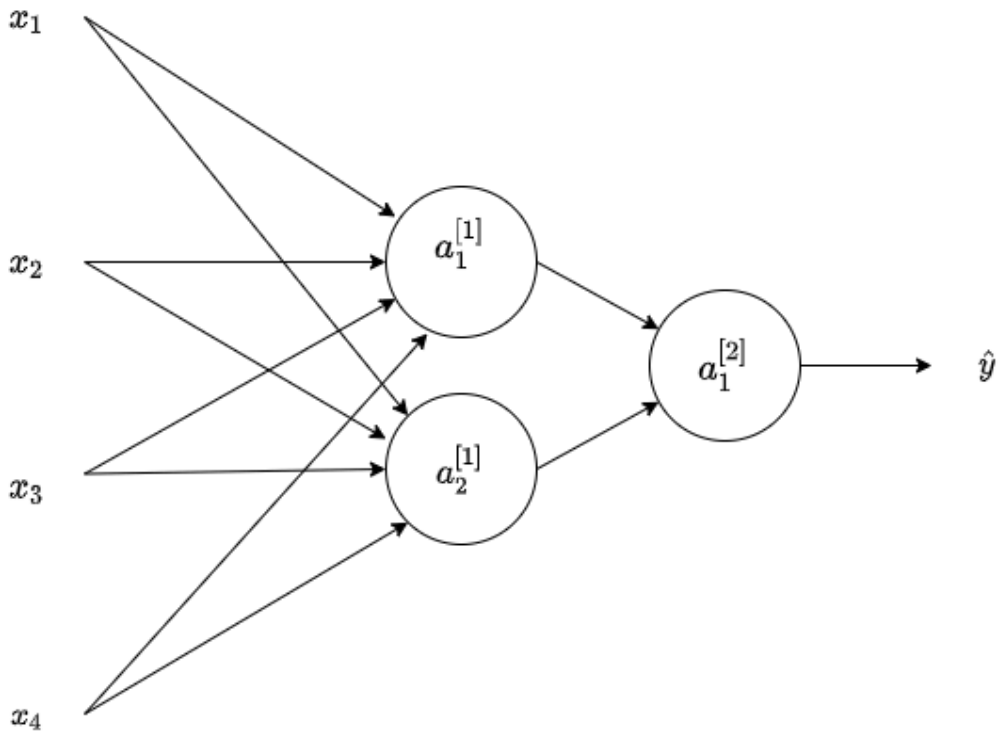
Correct

- [x] $b^{[2]}$ will have shape (1, 1)

Correct

- [ ] $W^{[2]}$ will have shape (4, 1)

10. Consider the following 1 hidden layer neural network:



What are the dimensions of $Z^{[1]}$ and $A^{[1]}$?

- (●) $Z^{[1]}$ and $A^{[1]}$ are (4, m)
- ( ) $Z^{[1]}$ and $A^{[1]}$ are (2, 1)
- ( ) $Z^{[1]}$ and $A^{[1]}$ are (2, m)
- ( ) $Z^{[1]}$ and $A^{[1]}$ are (4, 1)

⊗ **Incorrect**

No. The $Z^{[1]}$ and $A^{[1]}$ are calculated over a batch of training examples. The number of columns in $Z^{[1]}$ and $A^{[1]}$ is equal to the number of examples in the batch, m. And the number of rows in $Z^{[1]}$ and $A^{[1]}$ is equal to the number of neurons in the first layer.