Congratulations! You passed!

Grade received 100% **Latest Submission Grade** 100% **To pass** 80% or higher

Go to next item

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

1 / 1 point

- [ ] $w_3^{[4]}$ is the column vector of parameters of the third layer and fourth neuron.
- [ ] $a^{[3](2)}$ denotes the activation vector of the second layer for the third example.
- [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.

- [ ] $w_3^{[4]}$ is the row vector of parameters of the fourth layer and third neuron.
- [ ] $a_3^{[2]}$ denotes the activation vector of the second layer for the third example.

Expand

Correct
Great, you got all the right answers.

2. The tanh activation is not always better than sigmoid activation function for hidden units because the mean of its output is closer to zero, and so it centers the data, making learning complex for the next layer. True/False?

1 / 1 point

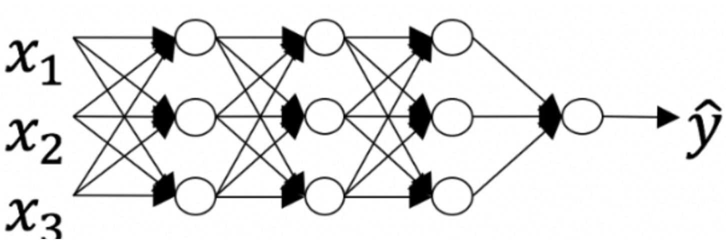
- ( ) True
- (•) False

Expand

Correct
Yes. As seen in lecture the output of the tanh is between -1 and 1, it thus centers the data which makes the learning simpler for the next layer.

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



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

Expand

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

- [x] False
- [ ] True

Expand

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(4, 5)

y = np.sum(x, axis=1)

#+end_src

What will be y.shape?

- [ ] (5,)
- [x] (4,)
- [ ] (1, 5)
- [ ] (4, 1)

Expand

Correct
Yes. By using axis=1 the sum is computed over each row of the array, thus the resulting array is a column vector with 4 entries. Since the option keepdims was not used the array doesn't keep the second dimension.

6. Suppose you have built a neural network with one hidden layer and tanh as activation function for the hidden layers. Which of the following is a best option to initialize the weights?

1 / 1 point

- ○ Initialize all weights to 0.
- ○ Initialize the weights to large random numbers.
- ◉ Initialize the weights to small random numbers.
- ○ Initialize all weights to a single number chosen randomly.

Expand

Correct
The use of random numbers helps to "break the symmetry" between all the neurons allowing them to compute different functions. When using small random numbers the values $z^{[k]}$ will be close to zero thus the activation values will have a larger gradient speeding up the training process.

7. Using linear activation functions in the hidden layers of a multilayer neural network is equivalent to using a single layer. True/False?

1 / 1 point

- ○ False
- ◉ True

Expand

Correct
Yes. When the identity or linear activation function $g(c) = c$ is used the output of composition of layers is equivalent to the computations made by a single layer.

8. Which of the following are true about the tanh function?

1 / 1 point

- ☑ For large values the slope is close to zero.

  Correct
  Yes. We can see in the graph of the $y = tanh(c)$ how as the values of c increase the curve becomes flatter.

- ☑ The tanh is mathematically a shifted version of the sigmoid function.

  Correct
  Yes. You can see the shape of both is very similar but tanh passes through the origin.

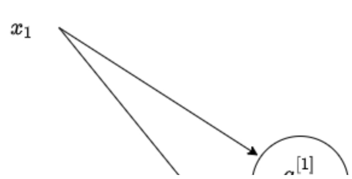
- ☐ For large values the slope is larger.
- ☐ The slope is zero for negative values.
- ☐ The derivative at $c = 0$ is not well defined.

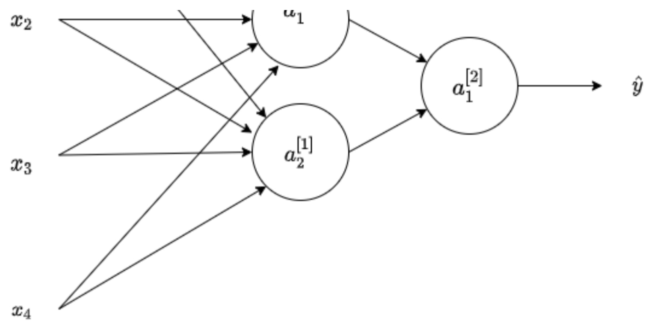
Expand

Correct
Great, you got all the right answers.

9. Consider the following 1 hidden layer neural network:

1 / 1 point





Which of the following statements are True? (Check all that apply).

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

Correct
Yes. The number of rows in $W^{[k]}$ is the number of neurons in the k-th layer and the number of columns is the number of inputs of the layer.

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

Correct
Yes. The number of rows in $W^{[k]}$ is the number of neurons in the k-th layer and the number of columns is the number of inputs of the layer.

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

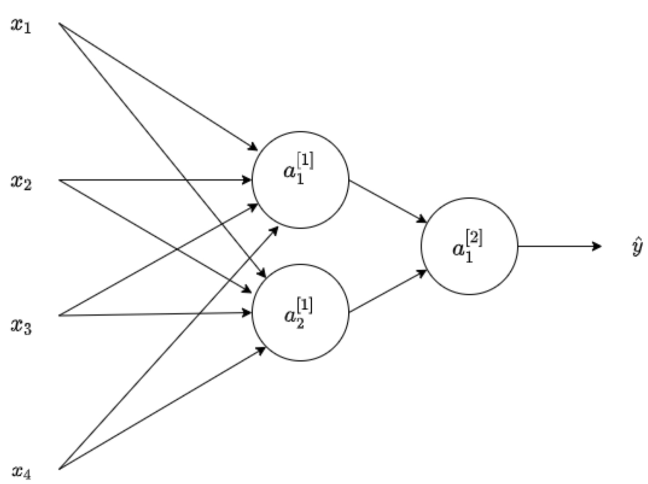
Correct
Yes. $b^{[k]}$ is a column vector and has the same number of rows as neurons in the k-th layer.

Expand

Correct
Great, you got all the right answers.

10. Consider the following 1 hidden layer neural network:

1 / 1 point



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

Expand

Correct

Yes. 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.