Congratulations! You passed!

Grade received 90% **To pass** 80% or higher

Go to next item

Shallow Neural Networks

Latest Submission Grade 90%

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

1 / 1 point

- [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.

- [ ] $W^{[1]}$ is a matrix with rows equal to the parameter vectors of the first layer.
- [ ] $w_3^{[4]}$ is the column vector of parameters of the third layer and fourth neuron.
- [ ] $w_3^{[4]}$ is the row vector of parameters of the fourth layer and third neuron.
- [x] $W^{[1]}$ is a matrix with rows equal to the transpose of the parameter vectors of the first layer.

  Correct
  Yes. We construct $W^{[1]}$ stacking the parameter vectors $w_j^{[1]}$ of all the neurons of the first layer.

- [ ] W_1 is a matrix with rows equal to the parameter vectors of the first layer.

2. In which of the following cases is the linear (identity) activation function most likely used?

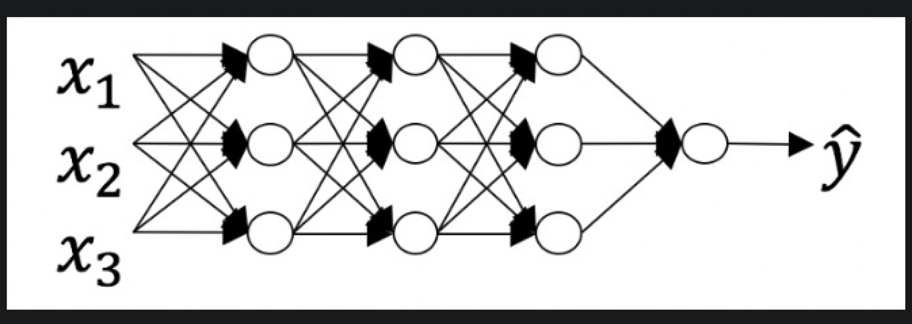
0 / 1 point

- ( ) The linear activation function is never used.
- ( ) As activation function in the hidden layers.
- ( ) When working with regression problems.
- (•) For binary classification problems.

  Incorrect
  No. In these problems we rather have an output restricted between 0 and 1.

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. The use of the ReLU activation function is becoming more rare because the ReLU function has no derivative for $c = 0$. True/False?

1 / 1 point

- (•) False
- ( ) True

  Correct
  Yes. Although the ReLU function has no derivative at $c = 0$ this rarely causes any problems in practice. Moreover it has become the default activation function in many cases, as explained in the lectures.

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?

- (•) (4,)
- ( ) (5,)
- ( ) (1, 5)
- ( ) (4, 1)

  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. 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 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.
- ( ) 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".

  Correct

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

  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.

1 / 1 point

Which of the following are true about the tanh function?

- [ ] The derivative at $c = 0$ is not well defined.
- [ ] The slope is zero for negative values.
- [x] 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.

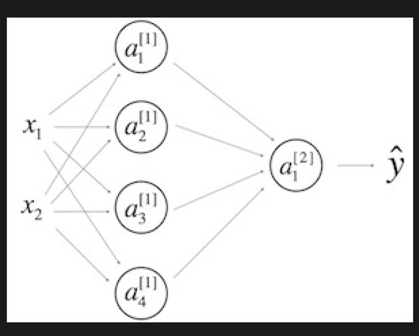
- [x] The tanh is mathematically a shifted version of the sigmoid function.

  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.

9. Consider the following 1 hidden layer neural network:

1 / 1 point



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

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

  Correct

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

  Correct

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

  Correct

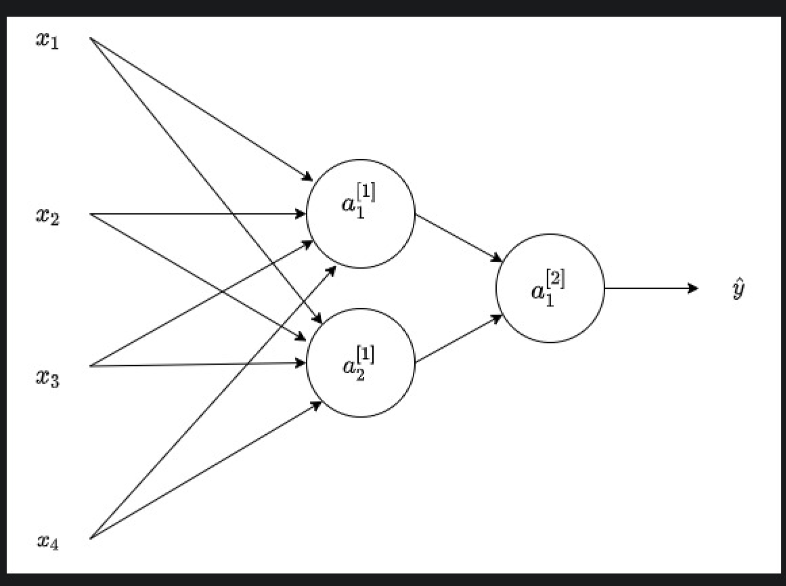
- [x] $W^{[1]}$ will have shape (4, 2)

  Correct

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

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

  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.