Back

Shallow Neural Networks

Graded Quiz • 50 min

Due Jun 18, 11:59 PM PDT

Congratulations! You passed!

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

Go to next item

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

1 / 1 point

- [ ] $a_3^{[2]}$ denotes the activation vector of the second layer for the third example.
- [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 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] $w_3^{[4]}$ 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.
- [ ] $a^{[3](2)}$ denotes the activation vector of the second layer for the third example.

Expand

Correct
Great, you got all the right answers.

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

1 / 1 point

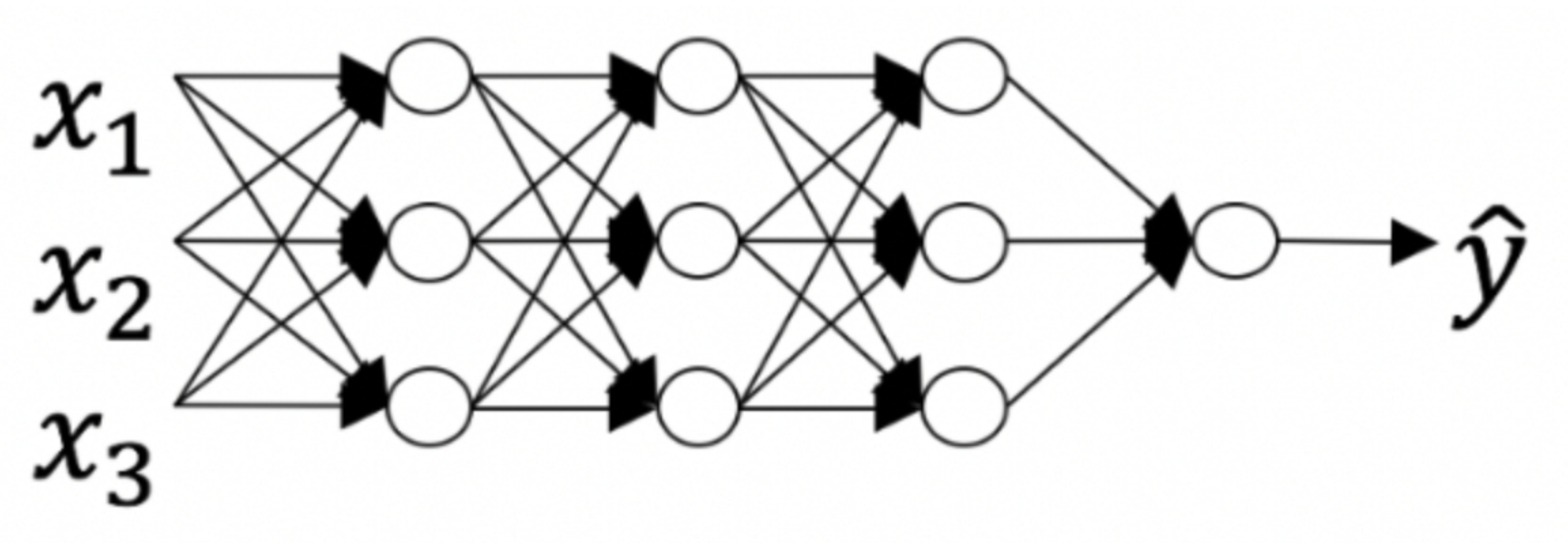
- (•) When working with regression problems.
- ( ) The linear activation function is never used.
- ( ) For binary classification problems.
- ( ) As activation function in the hidden layers.

Expand

Correct
Yes. In problems such as predicting the price of a house it makes sense to use the linear activation function as output.

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



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

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. 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

- ( ) True
- (•) False

Expand

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

A = np.random.randn(4,3)

B = np.sum(A, axis = 1, keepdims = True)

What will be B.shape? (If you're not sure, feel free to run this in python to find out).

- ( ) (1, 3)
- ( ) (3,)
- (•) (4, 1)
- ( ) (4,)

Expand

Correct
Yes, we use (keepdims = True) to make sure that A.shape is (4,1) and not (4,). It makes our code more robust.

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.
- ( ) 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 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.

Expand

Correct

7. A single output and single layer neural network that uses the sigmoid function as activation is equivalent to the logistic regression. True/False

1 / 1 point

- (•) True
- ( ) False

Expand

Correct
Yes. The logistic regression model can be expressed by $\hat{y} = \sigma\,(W\,x + b)$. This is the same as $a^{[1]} = \sigma(W^{[1]}\,X + b)$.

8. Which of the following is true about the ReLU activation functions?

1 / 1 point

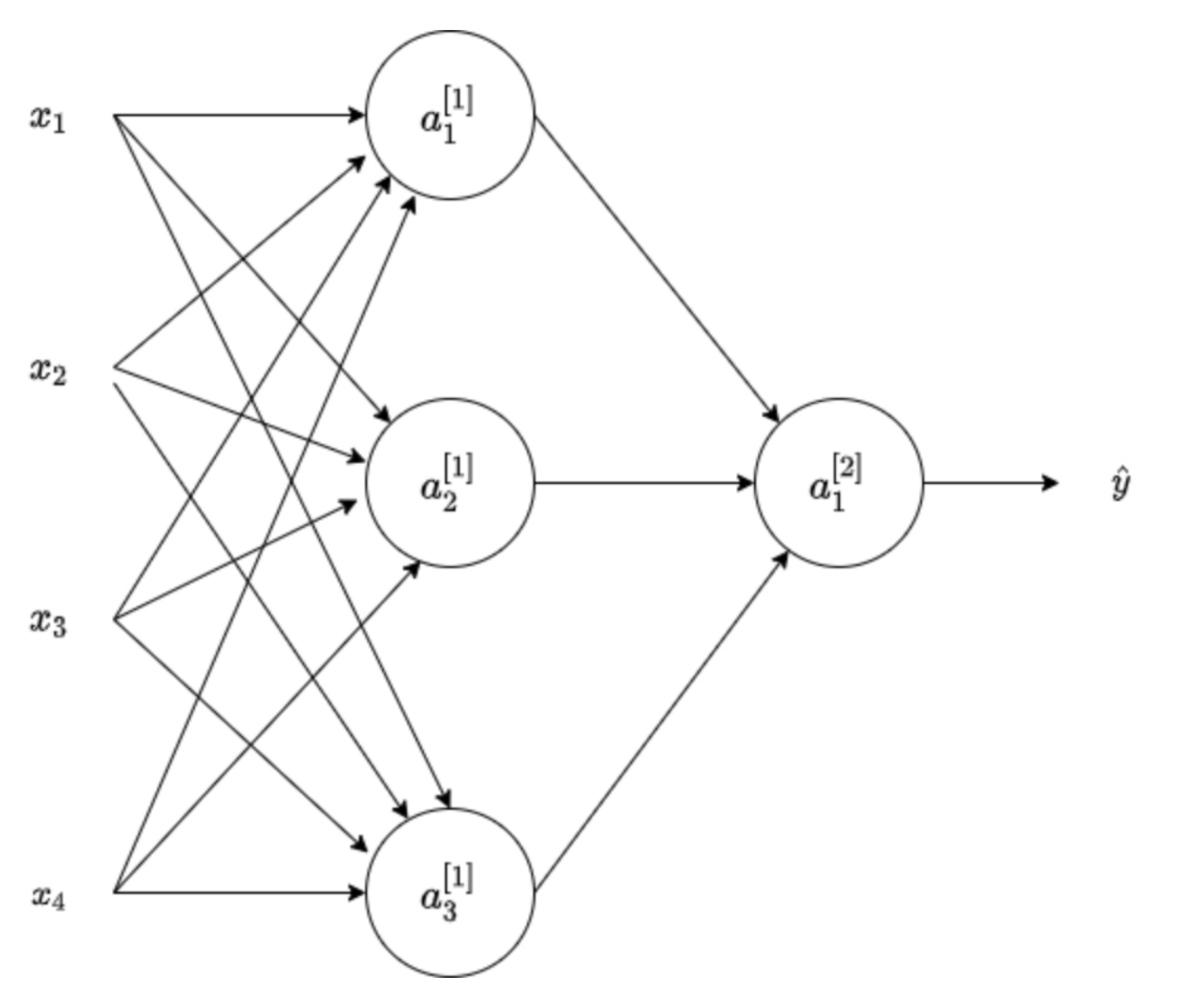
- ( ) They are only used in the case of regression problems, such as predicting house prices.
- (•) They are the go to option when you don't know what activation function to choose for hidden layers.
- ( ) They cause several problems in practice because they have no derivative at 0. That is why Leaky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

Expand

Correct

9. Consider the following 1 hidden layer neural network:

1 / 1 point



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

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

  Correct
  Yes. $b^{[k]}$ is a column vector and has the same number of rows as neurons in the k-th layer.
- [x] $b^{[1]}$ will have shape (3, 1).

  Correct
  Yes. $b^{[k]}$ is a column vector and has the same number of rows as neurons in the k-th layer.
- [ ] $W^{[1]}$ will have shape (4, 3).
- [x] $W^{[1]}$ will have shape (3, 4).

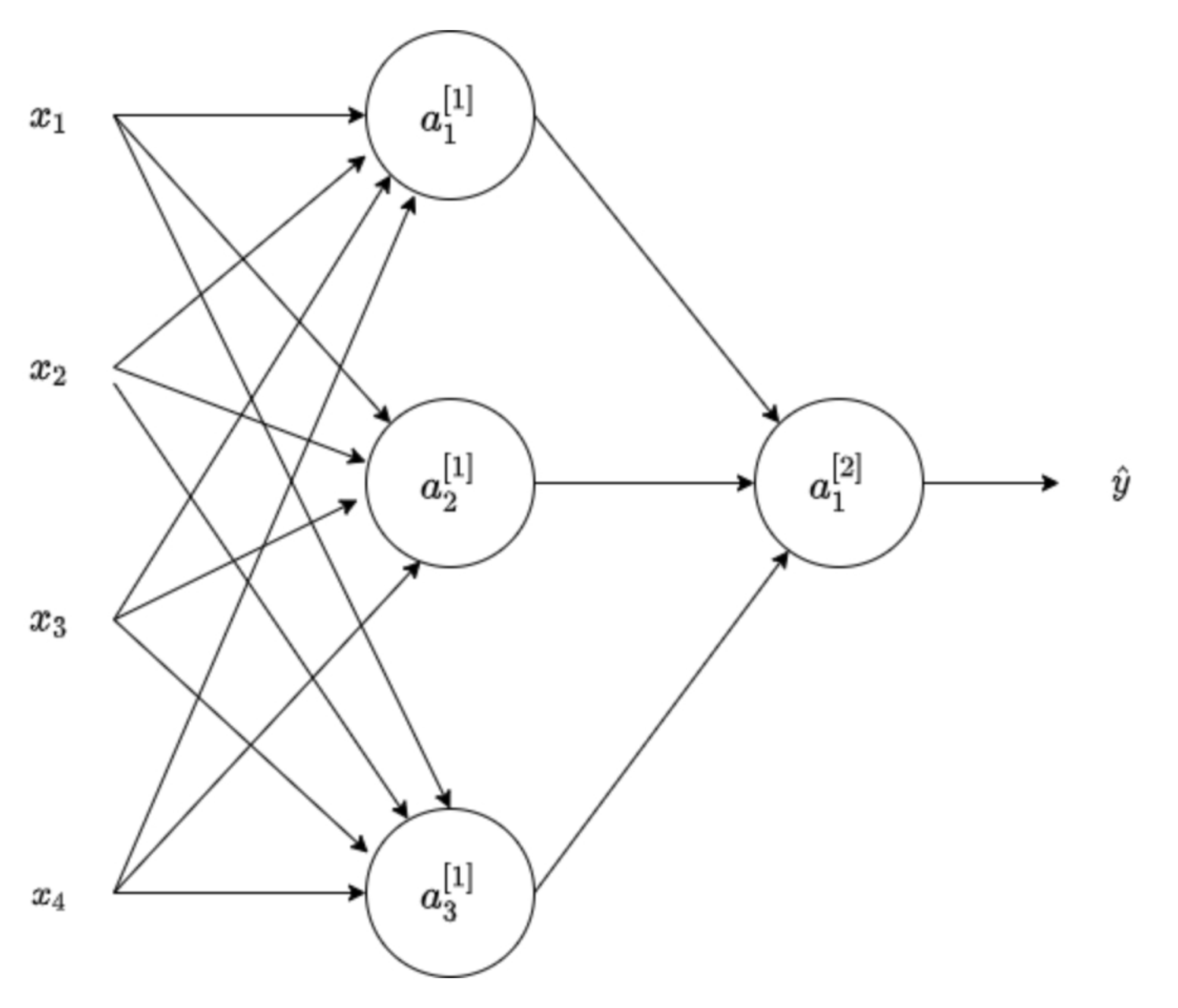
  Correct
  Yes. The number of rows in $W^{[k]}$ is the number of neurons in the k-th layer and the number of columns is the number of inputs of the layer.
- [ ] $b^{[1]}$ will have shape (1, 3)
- [ ] $b^{[2]}$ will have shape (3, 1)

Expand

Correct
Great, you got all the right answers.

10. Consider the following 1 hidden layer neural network:

0 / 1 point



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

- ( ) $Z^{[1]}$ and $A^{[1]}$ are (4, 1)
- ( ) $Z^{[1]}$ and $A^{[1]}$ are (3, m)
- (•) $Z^{[1]}$ and $A^{[1]}$ are (3, 1)
- ( ) $Z^{[1]}$ and $A^{[1]}$ are (4, m)

Expand

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.