Shallow Neural Networks

Quiz20 minutes • 20 min

Submit your assignment

Due April 25, 1:59 PM +07Apr 25, 1:59 PM +07

Attempts 3 every 8 hours

Try again

Receive grade

To Pass 80% or higher

Your grade

85.14%

View Feedback

We keep your highest score

Like

Dislike

Report an issue

Back

Shallow Neural Networks

Graded Quiz • 20 min

DueApr 25, 1:59 PM +07

Congratulations! You passed!

Grade received 85.14%

To pass 80% or higher

Go to next item

Shallow Neural Networks

Latest Submission Grade 85.14%

1.

Question 1

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

0.7142857142857143 / 1 point

☑ a^{[2]}_4$a_{4}^{[2]}$ is the activation output by the 4^{th}4^{th} neuron of the 2^{nd}2^{nd} layer

Correct

☐ a^{[2](12)}$a^{[2](12)}$ denotes activation vector of the 12^{th}12^{th} layer on the 2^{nd}2^{nd} training example.

☑ a^{[2]}$a^{[2]}$ denotes the activation vector of the 2^{nd}2^{nd} layer.

Correct

☐ a^{[2](12)}$a^{[2](12)}$ denotes the activation vector of the 2^{nd}2^{nd} layer for the 12^{th}12^{th} training example.

☑ a^{[2]}_4$a_{4}^{[2]}$ is the activation output of the 2^{nd}2^{nd} layer for the 4^{th}4^{th} training example

This should not be selected

☐

XX is a matrix in which each row is one training example.

☑

XX is a matrix in which each column is one training example.

Correct

2.

Question 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

◉

False

○

True

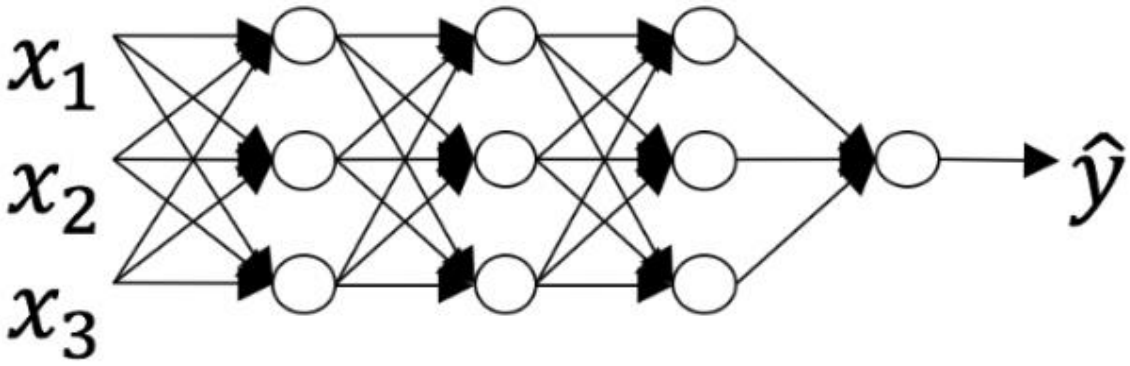
Correct

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.

Question 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^{[3](4)}_2$a^{2[3](4)}$

○

a^{[4](3)}_2$a^{2[4](3)}$

○

a^{[3](2)}_4$a^{4[3](2)}$

○

a^{[4]{2}}_3$a^{3[4]2}$

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.

Question 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

0 / 1 point

- (•) True
- ( ) False

Incorrect

No. Using tanh almost always works better than the sigmoid function for hidden layers.

5.

Question 5

Consider the following code:

#+begin_src python

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

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

#+end_src

What will be y.shape?

1 / 1 point

- ( ) (3, 1)
- ( ) (3,)
- ( ) (2,)
- (•) (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.

Question 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 all weights to a single number chosen randomly.
- (•) Initialize the weights to small random numbers.

Initialize the weights to large random numbers.

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.

Question 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 / 1 point

True

False

Correct

Yes, Logistic Regression doesn't have a hidden layer. If you initialize the weights to zeros, the first example x fed into 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.

Question 8

Which of the following are true about the tanh function?

0.8 / 1 point

The slope is zero for negative values.

The derivative at $c=0$ is not well defined.

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.

For large values the slope is larger.

This should not be selected

No. 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.

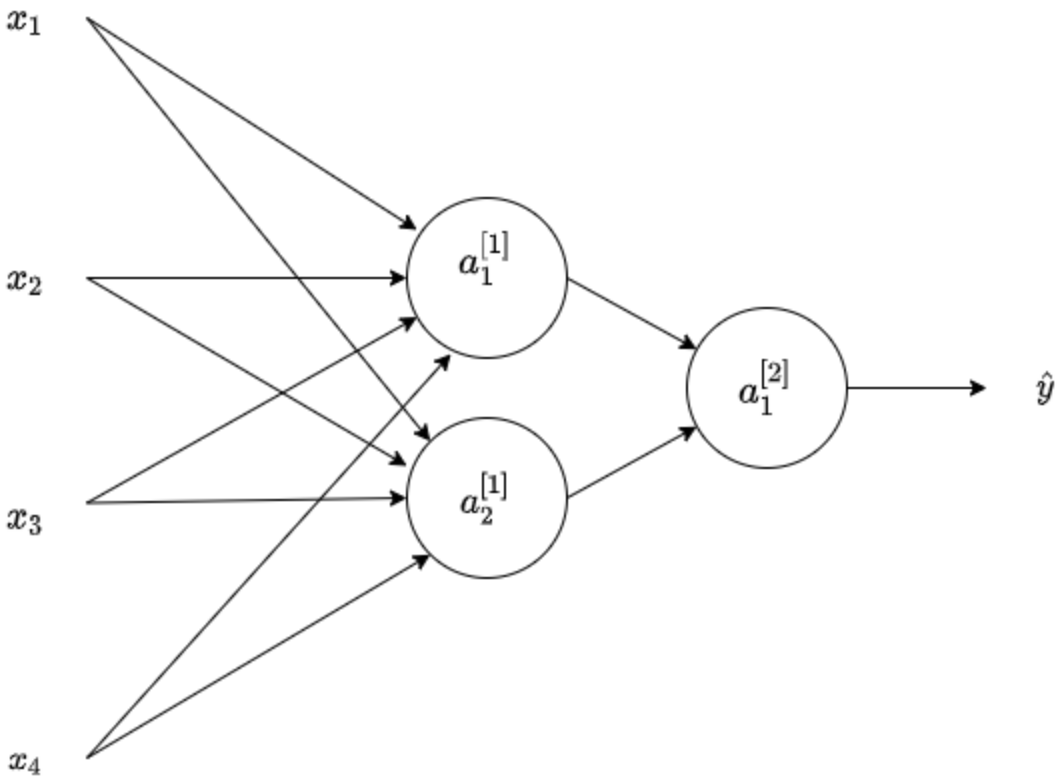
Correct

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

9.

Question 9

Consider the following 1 hidden layer neural network:



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

1 / 1 point

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

Correct

Yes. The number of rows in W^{[k]}$W_{[k]}$ 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]}$b_{[1]}$ will have shape (2, 1).

Correct

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

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

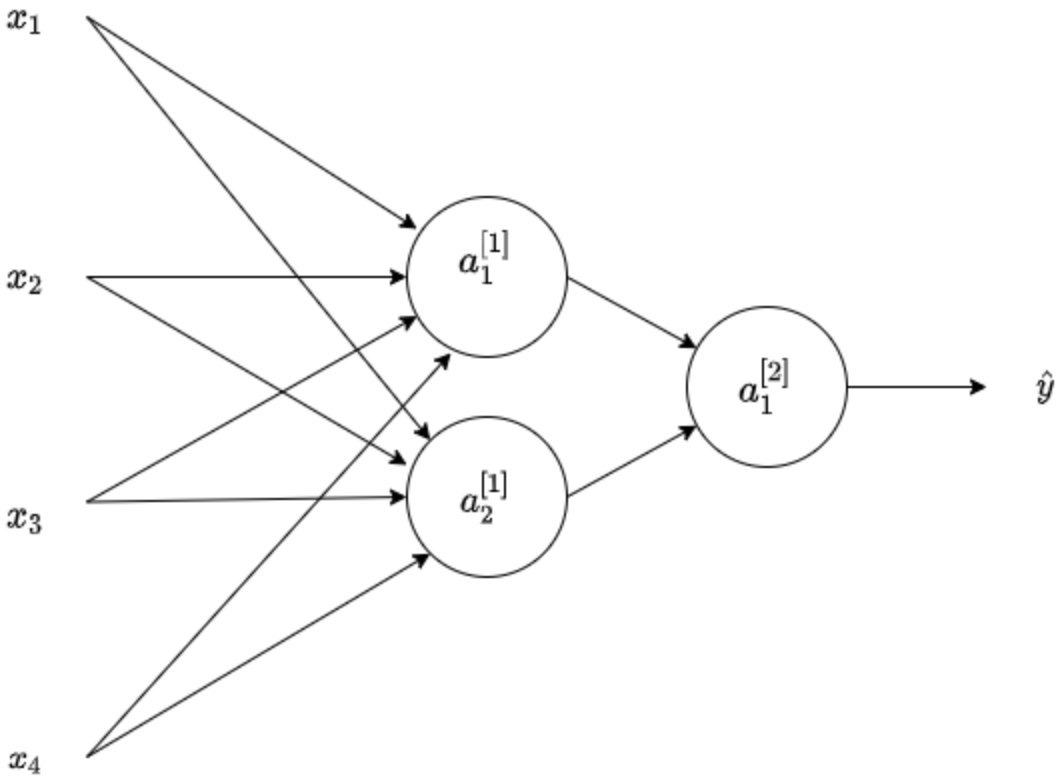
Correct

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

10.

Question 10

Consider the following 1 hidden layer neural network:



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

1 / 1 point

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

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

Z^{[1]}$Z^{[1]}$ and A^{[1]}$A^{[1]}$ are (4, 1)

Z^{[1]}$Z^{[1]}$ and A^{[1]}$A^{[1]}$ are (2, 1)

Correct

Yes. The Z^{[1]}$Z^{[1]}$ and A^{[1]}$A^{[1]}$ are calculated over a batch of training examples. The number of columns in Z^{[1]}$Z^{[1]}$ and A^{[1]}$A^{[1]}$ is equal to the number of examples in the batch, m. And the number of rows in Z^{[1]}$Z^{[1]}$ and A^{[1]}$A^{[1]}$ is equal to the number of neurons in the first layer.