

✔ Congratulations! You passed!

Grade received 80% To pass 80% or higher

Go to next item

Shallow Neural Networks

Latest Submission Grade 80%

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

0.6666666666666666

/ 1 point

☐ $a_3^{[2]}$ denotes the activation vector of the second layer for the third example.

✔ 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 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.

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

✘ This should not be selected

No. The vector $w_j^{[i]}$ is the column vector of parameters of the jth neuron in the i-th layer.

☐ $a^{[2]}$ 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.

- ☐ • $Z^{[l]} = W^{[l]}A^{[l]} + b^{[l]}$
- $A^{[l+1]} = g^{[l+1]}(Z^{[l]})$
- ☐ • $Z^{[l]} = W^{[l-1]}A^{[l]} + b^{[l-1]}$
- $A^{[l]} = g^{[l]}(Z^{[l]})$
- ☐ • $Z^{[l]} = W^{[l]}A^{[l]} + b^{[l]}$
- $A^{[l+1]} = g^{[l]}(Z^{[l]})$
- ☒ • $Z^{[l]} = W^{[l]}A^{[l-1]} + b^{[l]}$
- $A^{[l]} = g^{[l]}(Z^{[l]})$

✔ Correct

☒ False

☐ True

✔ Correct

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

5. Consider the following code:

```
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).

☒ (4, 1)

☐ (3,)

☐ (1, 3)

✔ 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 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?

0 / 1 point

☐ Initialize the weights to small random numbers.

☐ Initialize all weights to a single number chosen randomly.

☐ Initialize all weights to 0.

✘ Incorrect

No. In this case all neurons in the 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.

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

☒ True

☐ False

✔ 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 is true about the ReLU activation functions?

1 / 1 point

☐ They are increasingly being replaced by the tanh in most cases.

☐ They are only used in the case of regression problems, such as predicting house prices.

☐ 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.

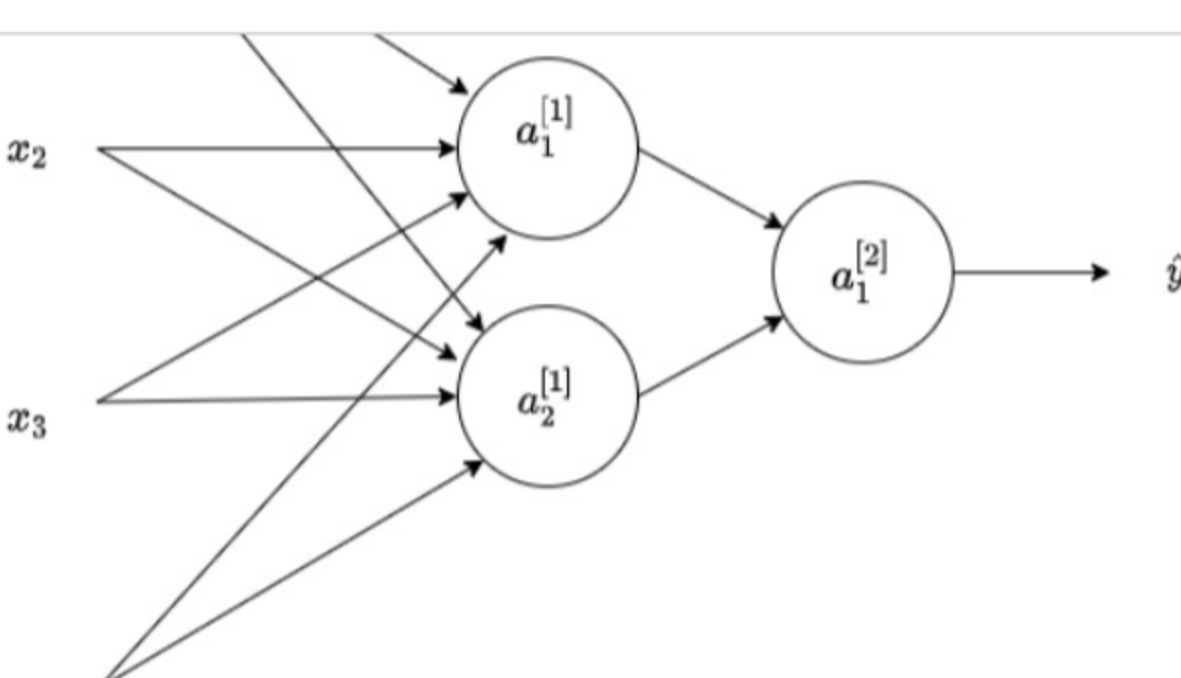
✔ Correct

9. Consider the following 1 hidden layer neural network:

0.3333333333333333

/ 1 point

x_1



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

← Back

Shallow Neural Networks

Graded Quiz • 20 min

Due

Feb 28, 2:59 PM WIB

✔ 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 (2, 4).

☐ $b^{[1]}$ will have shape (4, 2)

☒ $W^{[1]}$ will have shape (4, 2).

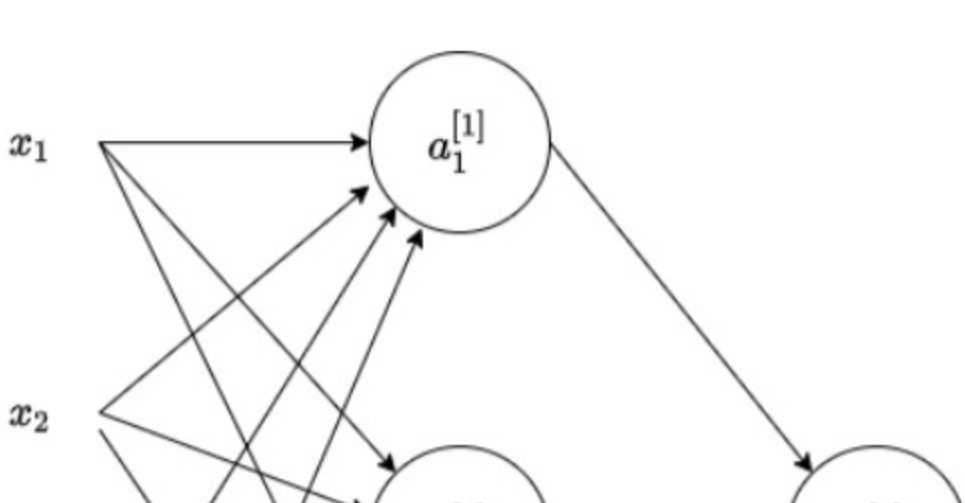
✘ This should not be selected

No. 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.

☒ $W^{[2]}$ 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 (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)

✔ 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.