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**

- [ ] X is a matrix in which each row is one training example.
- [x] X is a matrix in which each column is one training example.

Correct

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

Correct

- [ ] $a_4^{[2]}$ is the activation output of the 2^{nd} layer for the 4^{th} training example
- [x] $a_4^{[2]}$ is the activation output by the 4^{th} neuron of the 2^{nd} layer

Correct

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

Correct

Expand

Correct
Great, you got all the right answers.

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

Expand

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.

3. Which of these is a correct vectorized implementation of forward propagation for layer l, where $1 \leq l \leq L$? **1 / 1 point**

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

$A^{[l+1]} = g^{[l+1]}(Z^{[l]})$

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

Expand

Correct

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).

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

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 with one hidden layer and tanh as activation function for the hidden layer. You decide to initialize the weights to small random numbers and the biases to zero. The first hidden layer's neurons will perform different computations from each other even in the first iteration. True/False? **1 / 1 point**

- False No. Since the weights are most likely different, each neuron will do a different computation.
- True Yes. Since the weights are most likely different, each neuron will do a different computation.

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 $\hat{y} = \sigma\left(W\,x + b\right)$. 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

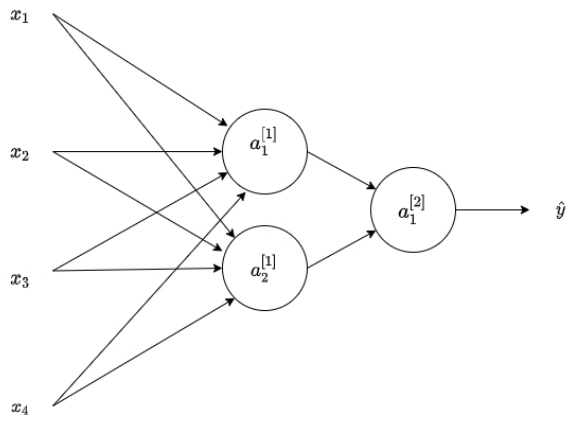
- ( ) They are only used in the case of regression problems, such as predicting house prices.
- ( ) They cause several problems in practice because they have no derivative at 0. That is why Leaky ReLU was invented.
- (●) They are the go to option when you don't know what activation function to choose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

Expand

Correct

9. Consider the following 1 hidden layer neural network:

0 / 1 point



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

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

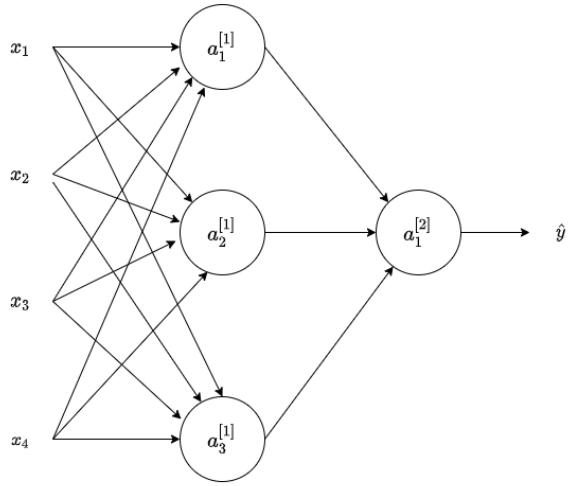
- [x] $b^{[1]}$ will have shape (2, 1).

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

Expand

Incorrect
You didn't select all the correct 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, 1)
- $Z^{[1]}$ and $A^{[1]}$ are (3, 1)
- $Z^{[1]}$ and $A^{[1]}$ are (4, m)
- $Z^{[1]}$ and $A^{[1]}$ are (3, 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.