Congratulations! You passed!

Grade received 100%

Latest Submission Grade 100%

To pass 80% or higher

Go to next item

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

1 / 1 point

- [ ] $a_4^{[2]}$ is the activation output of the 2^{nd} layer for the 4^{th} training example
- [x] $a^{[2]}$ denotes the activation vector of the 2^{nd} layer.

Correct

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

Correct

- [ ] 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

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

Correct

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

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

- (•) False
- ( ) True

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

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

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

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

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

Expand

Correct

4. You are building a binary classifier for recognizing cucumbers (y=1) vs. watermelons (y=0). Which one of these activation functions would you recommend using for the output layer?

1 / 1 point

- ReLU
- tanh
- sigmoid
- Leaky ReLU

Expand

Correct
Yes. Sigmoid outputs a value between 0 and 1 which makes it a very good choice for binary classification. You can classify as 0 if the output is less than 0.5 and classify as 1 if the output is more than 0.5. It can be done with tanh as well but it is less convenient as the output is between -1 and 1.

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)
- (4, 1)
- (4,)
- (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 layers. Which of the following is a best option to initialize the weights?

1 / 1 point

- Initialize the weights to small random numbers.
- Initialize the weights to large random numbers.
- Initialize all weights to a single number chosen randomly.
- Initialize all weights to 0.

Expand

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. 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

Expand

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 are true about the tanh function?

1 / 1 point

- [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.

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

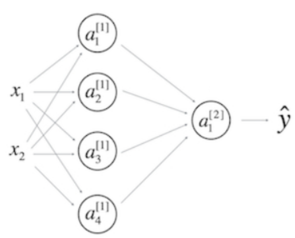
- [ ] The slope is zero for negative values.
- [ ] For large values the slope is larger.

Expand

Correct
Great, you got all the right answers.

9. Consider the following 1 hidden layer neural network:

1 / 1 point



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

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

  Correct

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

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

  Correct

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

  Correct

☐ $$W^{[2]}$$ will have shape (4, 1)

☑ $$W^{[2]}$$ will have shape (1, 4)

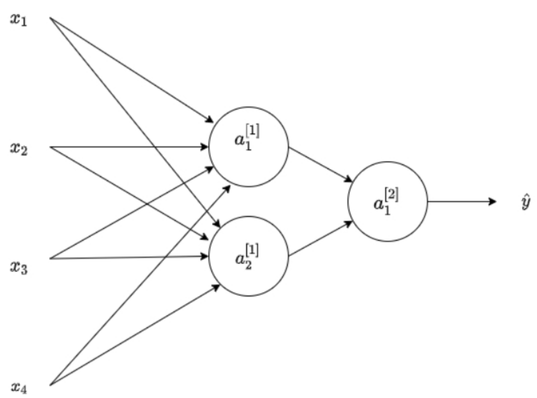
✓ Correct

Expand

Correct
Great, you got all the right 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 (2, 1)
- ◉ $Z^{[1]}$ and $A^{[1]}$ are (2, m)
- ○ $Z^{[1]}$ and $A^{[1]}$ are (4, 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.