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**

- [ ] $a^{[2](12)}$ denotes activation vector of the 12^{th} layer on the 2^{nd} training example.
- [ ] X is a matrix in which each row is one training example.
- [x] $a^{[2](12)}$ denotes the activation vector of the 2^{nd} layer for the 12^{th} training example.

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

  ✓ Correct
- [x] X is a matrix in which each column is one 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

Expand

Correct
Great, you got all the right answers.

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

Expand

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

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

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

Expand

Correct

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 **1 / 1 point**

- True
- False

Expand

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

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).

- (3,)
- (4, 1)
- (1, 3)
- (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 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. Logistic regression's weights 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

Expand

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. Which of the following are true about the tanh function? **1 / 1 point**

- ☑ 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 tanh is mathematically a shifted version of the sigmoid function.

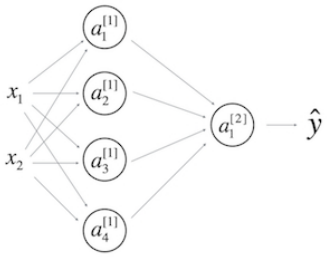
  Correct
  Yes. You can see the shape of both is very similar but tanh passes through the origin.

- ☐ The derivative at $c = 0$ is not well defined.
- ☐ For large values the slope is larger.
- ☐ The slope is zero for negative values.

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).

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

  Correct

- ☑ $b^{[2]}$ will have shape (1, 1)

  Correct

- ☑ $b^{[1]}$ will have shape (4, 1)

  Correct

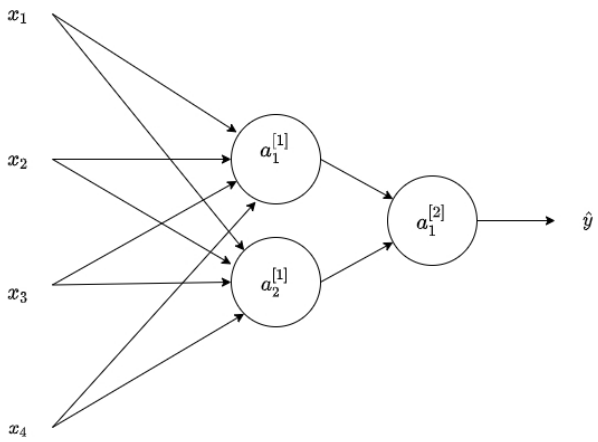
- [ ] $b^{[2]}$ will have shape (4, 1)
- [ ] $W^{[2]}$ will have shape (4, 1)
- [ ] $W^{[1]}$ will have shape (2, 4)
- [x] $W^{[1]}$ will have shape (4, 2)

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, m)
- (•) $Z^{[1]}$ and $A^{[1]}$ are (2, m)
- ( ) $Z^{[1]}$ and $A^{[1]}$ are (2, 1)
- ( ) $Z^{[1]}$ and $A^{[1]}$ are (4, 1)

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.