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

- [ ] $a_4^{[2]}$ is the activation output of the 2^{nd} layer for the 4^{th} training example
- [x] $a^{[2](12)}$ denotes the activation vector of the 2^{nd} layer for the 12^{th} training example.
- [x] $a_4^{[2]}$ is the activation output by the 4^{th} neuron of the 2^{nd} layer
- [x] $a^{[2]}$ denotes the activation vector of the 2^{nd} layer.
- [ ] X is a matrix in which each row is one training example.
- [ ] $a^{[2](12)}$ denotes activation vector of the 12^{th} layer on the 2^{nd} training example.
- [x] X is a matrix in which each column is one training example.

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 point

- (•) True
- ( ) False

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

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

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 point

- (•) False
- ( ) True

5. Consider the following code: 1 point

#+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, 2)
- ( ) (2,)
- ( ) (3,)
- ( ) (3, 1)

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 point

- (•) True

  Yes. Since the weights are most likely different, each neuron will do a different computation.
- ( ) False

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

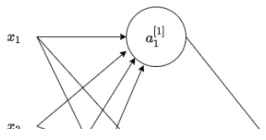
7. Using linear activation functions in the hidden layers of a multilayer neural network is equivalent to using a single layer. True/False? 1 point

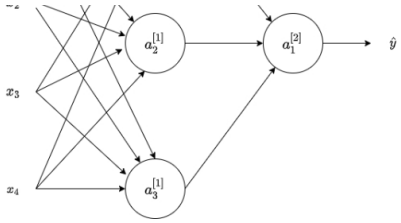
- (•) True
- ( ) False

8. Which of the following is true about the ReLU activation functions? 1 point

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

9. Consider the following 1 hidden layer neural network: 1 point

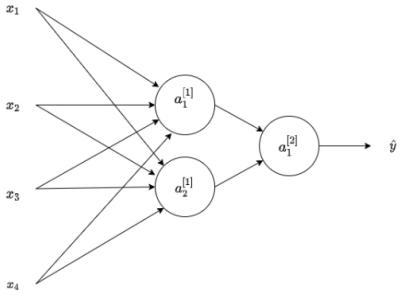




Which of the following statements are True? (Check all that apply).

- [x] $b^{[1]}$ will have shape (3, 1).
- [x] $b^{[2]}$ will have shape (1,1)
- [x] $W^{[1]}$ will have shape (3, 4).
- [ ] $W^{[1]}$ will have shape (4, 3).
- [ ] $b^{[2]}$ will have shape (3, 1)
- [ ] $b^{[1]}$ will have shape (1, 3)

10. Consider the following 1 hidden layer neural network: 1 point



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

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