**1.Question 1**

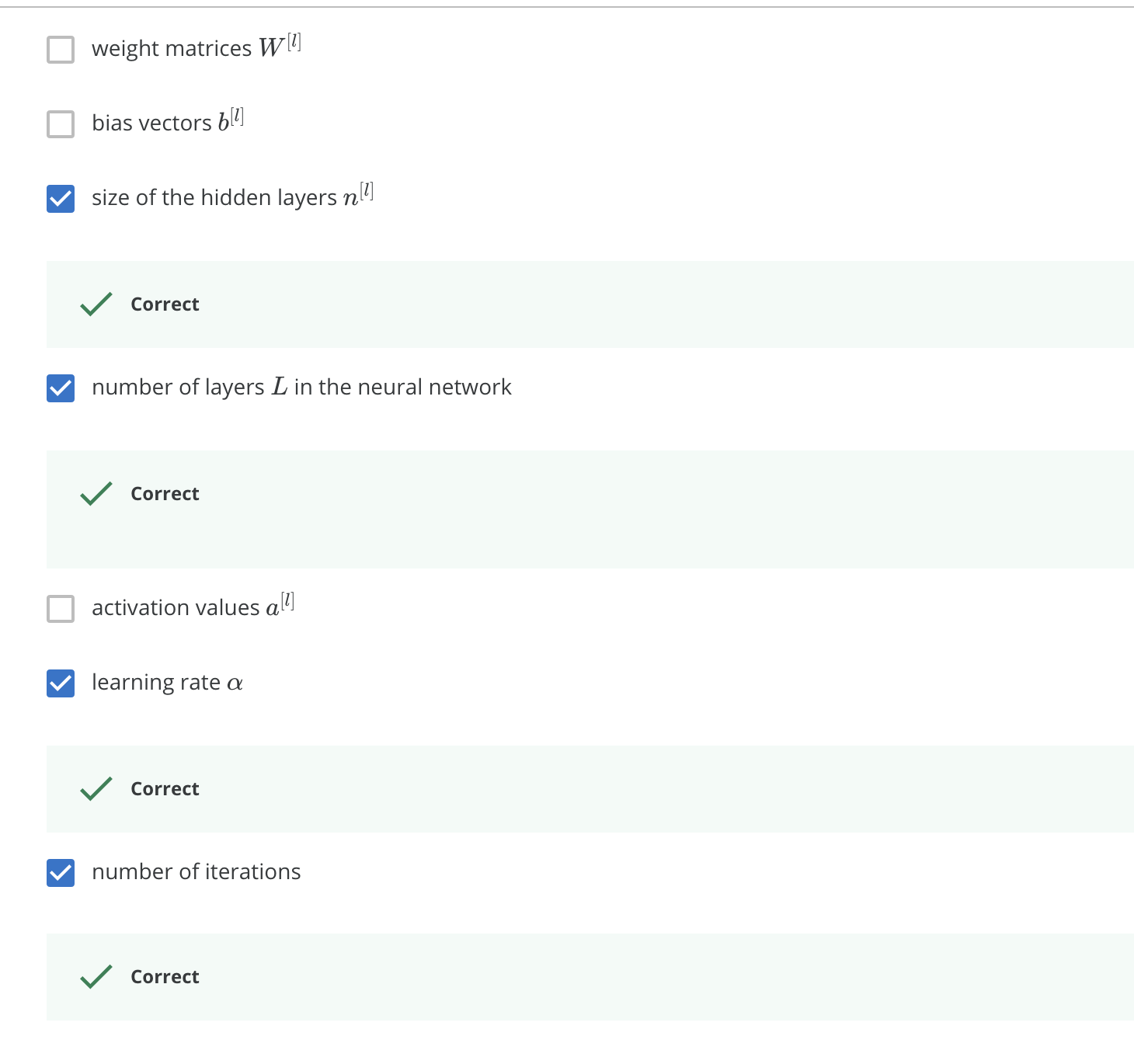
**What is the "cache" used for in our implementation of forward propagation and backward propagation?**

We use it to pass variables computed during forward propagation to the corresponding backward propagation step. It contains useful values for backward propagation to compute derivatives.

Correct, the "cache" records values from the forward propagation units and sends it to the backward propagation units because it is needed to compute the chain rule derivatives.

**2.Question 2**

**Among the following, which ones are "hyperparameters"? (Check all that apply.)**



**3.Question 3**

**Which of the following statements is true?**

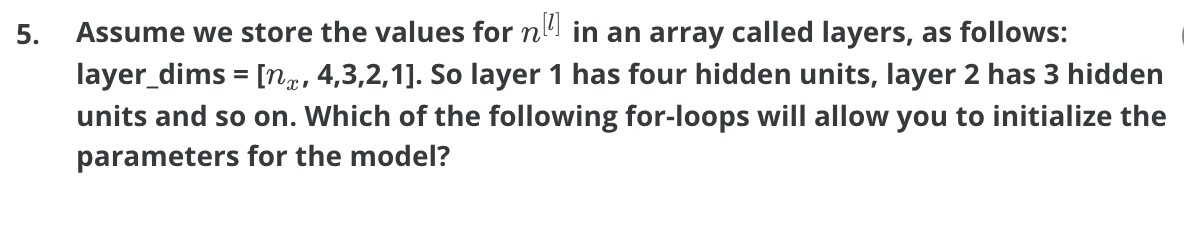
The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.

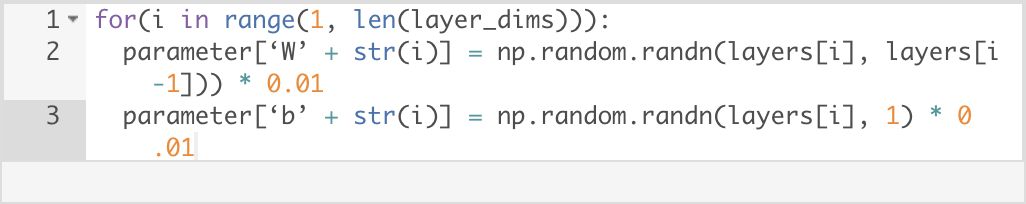
**4.Question 4**

**Vectorization allows you to compute forward propagation in an L*L*-layer neural network without an explicit for-loop (or any other explicit iterative loop) over the layers l=1, 2, …,L. True/False?**

False

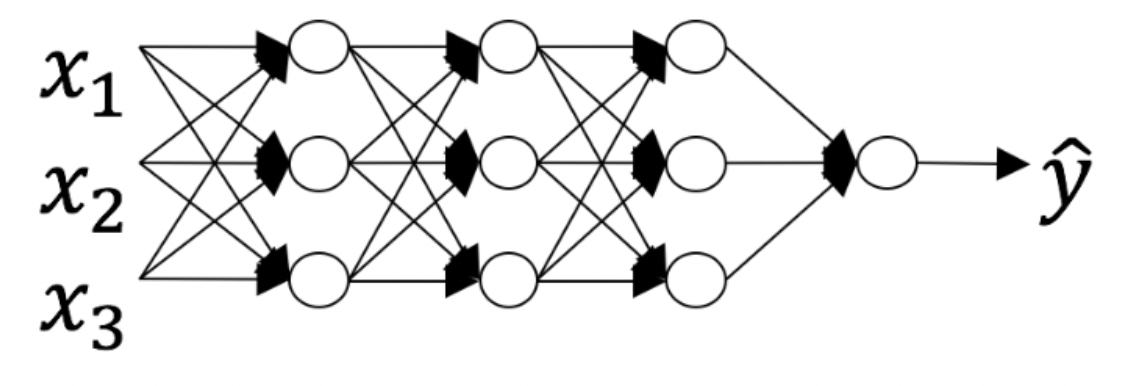
Forward propagation propagates the input through the layers, although for shallow networks we may just write all the lines (a^{[2]} = g^{[2]}(z^{[2]})*a*[2]=*g*[2](*z*[2]), z^{[2]}= W^{[2]}a^{[1]}+b^{[2]}*z*[2]=*W*[2]*a*[1]+*b*[2], ...) in a deeper network, we cannot avoid a for loop iterating over the layers: (a^{[l]} = g^{[l]}(z^{[l]})*a*[*l*]=*g*[*l*](*z*[*l*]), z^{[l]} = W^{[l]}a^{[l-1]} + b^{[l]}*z*[*l*]=*W*[*l*]*a*[*l*−1]+*b*[*l*], ...).





**6.Question 6**

**Consider the following neural network.**



The number of layers L is 4. The number of hidden layers is 3

Yes. As seen in lecture, the number of layers is counted as the number of hidden layers + 1. The input and output layers are not counted as hidden layers.

**7.Question 7**

**During forward propagation, in the forward function for a layer l*l* you need to know what is the activation function in a layer (Sigmoid, tanh, ReLU, etc.). During backpropagation, the corresponding backward function also needs to know what is the activation function for layer l*l*, since the gradient depends on it. True/False?**

True

Yes, as you've seen in the week 3 each activation has a different derivative. Thus, during backpropagation you need to know which activation was used in the forward propagation to be able to compute the correct derivative.

**8.Question 8**

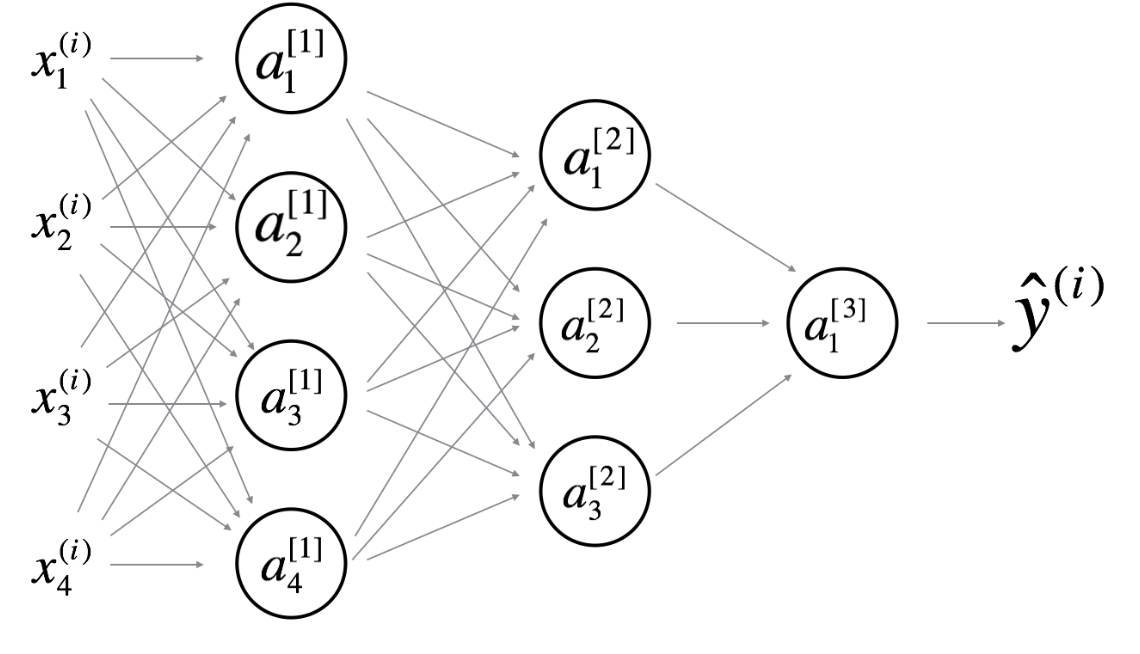
**There are certain functions with the following properties:**

**(i) To compute the function using a shallow network circuit, you will need a large network (where we measure size by the number of logic gates in the network), but (ii) To compute it using a deep network circuit, you need only an exponentially smaller network. True/False?**

True

**9.Question 9**

**Consider the following 2 hidden layer neural network:**



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

­ 