

Week 4 - Deep Neural Networks

1. What is the **cache** used for in our implementation of forward & backward propagation?

Ans: 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.

2. Among the following, which ones are hyper-parameters?

Ans:

- Size of the hidden layers $n^{[l]}$
- Learning rate α
- Number of iterations
- Number of layers L

http://colinraffel.com/wiki/neural_network_hyperparameters

3. Which of the following statements is true?

Ans: The deeper layers of a neural network are typically computing more complex features.

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

Ans: False. We cannot avoid for-loop iteration over the computations among layers

5. Assume we store the values for $n^{[l]}$ in an array. `layerdims = [nx, 4, 3, 2, 1]`. Which of the following for-loops will allow you to initialize the parameters for the model?

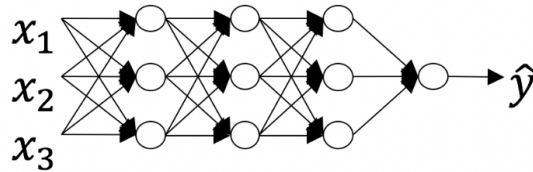
Ans:

```

for i in range(1, len(layer_dims)):
    parameter['W'+str(i)] = np.random.randn(layers[i],layers[i-1])*0.01
    parameter['b'+str(i)] = np.zeros(layers[i], 1)

```

6. Consider the neural network diagram. How many layers does this network have? **Ans:** The number of layers L is 4. The number of hidden layers is



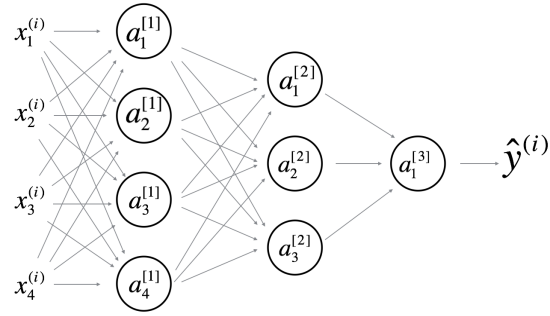
3. The input layer $L^{[0]}$ does not count; the number of layers is counted as the (number of hidden layers + 1). The input and output layers are not counted as hidden layers.
7. During forward propagation, in the forward function for a layer l you need to know what is the activation function in a layer (Sigmoid, tanh, ReLU, etc.). During backpropagation, the corresponding backward function also needs to know what is the activation function for layer l , since the gradient depends on it. True/False?

Ans: True

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) (ii)
 To compute it using a deep network circuit, you need only an exponentially smaller network. True/False?

Ans: True

9. Consider the following 2-hidden layer neural network. Which of the following statements are True? (Check all that apply).



Ans: The following are true

- $W^{[1]} = (4, 4)$ and $b^{[1]} = (4, 1)$
- $W^{[2]} = (3, 4)$ and $b^{[2]} = (3, 1)$
- $W^{[3]} = (1, 3)$ and $b^{[3]} = (1, 1)$

10. In the general case what is the dimension of $W^{[l]}$, the weight matrix of layer l ?

Ans: $(n^{[l]}, n^{[l-1]})$