Neural Networks and Deep Learning

Week 4: Key Concepts on Deep Neural Networks

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

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

Comment: The cache records values from the forward propagation units and are used in backward propagation units because it is needed to compute the chain rule derivatives.

2 Version 1: During the backpropagation process, we use gradient descent to change the hyperparameters. True/False?

Answer: False.

Comment: During backpropagation, we use gradient descent to compute new values of $W^{[l]}$ and $b^{[l]}$. These are the parameters of the network.

2 Version 2: Among the following which ones are hyperparameters? (Check all that apply)

Answer: number of iterations; number of layers L in the neural network; learning rate α; size of the hidden layers $n^{[l]}$.

3 Which of the following statements is true?

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

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 $l = 1, 2, \ldots, L$. True/False?

Answer: False.

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

5 Assume we store the values for $n^{[l]}$ in an array called layer dims, as follows, layer dims $= [n_x, 4, 3, 2, 1]$. So layer 1 has four hidden units, layer 2 has 3 hidden units and so on. Which of the following for loops will allow you to initialize the parameters for the model?

Answer:

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

6 Omitted. Too easy. The answer is directly from the lecture notes.

7 Version 1: During forward propagation, for the value of $A^{[l]}$ the value is used of $Z^{[l]}$ with the activation function $g^{[l]}$. During backward propagation we calculated $dA^{[l]}$ form $Z^{[l]}$.

Answer: False.

Comment: During backward propagation we are interested in computing $dW^{[l]}$ and $db^{[l]}$. For that we use $g'^{L}, dZ^{[L]}, Z^{[L]}$ and $W^{[L]}$.

7 Version 2: 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?

Answer: True.

Comment: 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 Version 1: A shallow neural network with a single hidden layer and 6 hidden units can compute any function that a neural network with 2 hidden layers and 6 hidden units can compute. True/False?

Answer: False.

8 Version 2: There are certain functions with the following properties:

- 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
- To compute it using a deep network circuit, you need only an exponentially smaller network. True/False.

Answer: True.

9 Version 1: Consider the following 2 hidden layers neural network:

- $x_1, \ldots, x_4$
- 3 neurons
- 4 neurons
- $a_1^{[3]}$
- $\hat{y}$

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

Answer: $b^{[1]}$ shape (3,1); $W^{[2]}$ shape (4,3); $W^{[1]}$ shape (3,4)

Comment: In general, shape of $b^{[l]}$ is $(n^{[l]}, 1)$; shape of $W^{[l]}$ is $(n^{[l]}$. See lecture notes for more details.

9 Version 2: Consider the following 2 hidden layers neural network:

- $x_1, \ldots, x_4$
- 4 neurons
- 3 neurons

- $a_1^{[3]}$
- $\hat{y}$

Answer: $b^{[1]}$ shape (4,1); $W^{[2]}$ shape (3,4); $W^{[1]}$ shape (4,4); $b^{[2]}$ shape (3,1); $b^{[3]}$ shape (1,1); $W^{32]}$ shape (1,3).

10 Whereas the previous question used a specific network, in the general case what is the dimension of $W^{[l]}$, the weight matrix associated with layer l?

Answer: $W^{[l]}$ has shape $(n^{[l]}, n^{[l+1]})$.