Question 1

What does a neuron compute?

A. A neuron computes a function g that scales the input x linearly (Wx + b)
B. A neuron computes a linear function (z = Wx + b) followed by an activation function
C. A neuron computes the mean of all features before applying the output to an activation function
D. A neuron computes an activation function followed by a linear function (z = Wx + b)

Question 2

Suppose img is a (32,32,3) array, representing a 32x32 image with 3 color channels red, green and blue. How do you reshape this into a column vector?

A. x = img.reshape((32*32*3,1))
B. x = img.reshape((32*32,3))
C. x = img.reshape((1,32*32,*3))
D. x = img.reshape((3,32*32))

Question 3

The tanh activation usually works better than sigmoid activation function for hidden units because the mean of its output is closer to zero, and so it centers the data better for the next layer. True/False?

A. True
B. False

Question 4

You are building a binary classifier for recognizing cucumbers (y=1) vs. watermelons (y=0). Which one of these activation functions would you recommend using for the output layer?

A. ReLU
B. Leaky ReLU
C. sigmoid
D. tanh

Question 5

Suppose you have built a neural network. You decide to initialize the weights and biases to be zero. Which of the following statements is true?

A. **Each neuron in the first hidden layer will perform the same computation. So even after multiple iterations of gradient descent each neuron in the layer will be computing the same thing as other neurons.**
B. Each neuron in the first hidden layer will perform the same computation in the first iteration. But after one iteration of gradient descent they will learn to compute different things because we have “broken symmetry”.
C. Each neuron in the first hidden layer will compute the same thing, but neurons in different layers will compute different things, thus we have accomplished “symmetry breaking” as described in lecture.
D. The first hidden layer’s neurons will perform different computations from each other even in the first iteration; their parameters will thus keep evolving in their own way.

Question 6

Logistic regression’s weights w 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?

A. True
B. False

Question 7

You have built a network using the tanh activation for all the hidden units. You initialize the weights to relative large values, using np.random.randn(..,..)*1000. What will happen?

A. **This will cause the inputs of the tanh to also be very large, thus causing gradients to be close to zero. The optimization algorithm will thus become slow.**
B. It doesn’t matter. So long as you initialize the weights randomly gradient descent is not affected by whether the weights are large or small.
C. This will cause the inputs of the tanh to also be very large, causing the units to be “highly activated” and thus speed up learning compared to if the weights had to start from small values.

D. This will cause the inputs of the tanh to also be very large, thus causing gradients to also become large. You therefore have to set α to be very small to prevent divergence; this will slow down learning.

.

Question 8

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

A. It is used to keep track of the hyperparameters that we are searching over, to speed up computation.
B. We use it to pass variables computed during backward propagation to the corresponding forward propagation step. It contains useful values for forward propagation to compute activations.
C. It is used to cache the intermediate values of the cost function during training.
D. 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.

Question 9

Which of the following statements is true?

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

Question 10

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?

A. True
B. False

Question 11

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 l, since the gradient depends on it. True/False?

A. **True**
B. False