1. **Which of the following are true? (Check all that apply.)**
   - Answer :
     + X is a matrix in which each column is one training example.
     + a^[2]_4 is the activation output by the 4th neuron of the 2nd layer
     + a^[2](12) denotes the activation vector of the 2nd layer for the 12th training example.
     + a^[2] denotes the activation vector of the 2nd layer.

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?**
   - Answer : False

3. **Which of these is a correct vectorized implementation of forward propagation for layer l, where $1 \leq l \leq L$?**
   - Answer :
     $Z^{[l]} = W^{[l]} A^{[l-1]} + b^{[l]}$
     $A^{[l]} = g^{[l]}(Z^{[l]})$

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?**
   - Answer : Sigmoid

5. **Consider the following code:**
   A = np.random.randn(4,3)B = np.sum(A, axis = 1, keepdims = True)
   What will be B.shape? (If you're not sure, feel free to run this in python to find out).
   - Answer : (4,1)

6. **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?**
   - Answer : 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.

7. **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?**

- Answer : False

   (Logistic Regression doesn't have a hidden layer. If you initialize the weights to zeros, the first example x fed in the logistic regression will output zero but the derivatives of the Logistic Regression depend on the input x (because there's no hidden layer) which is not zero. So at the second iteration, the weights values follow x's distribution and are different from each other if x is not a constant vector.)

8. **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?**

   - Answer : 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.

9. **Consider the following 1 hidden layer neural network:**
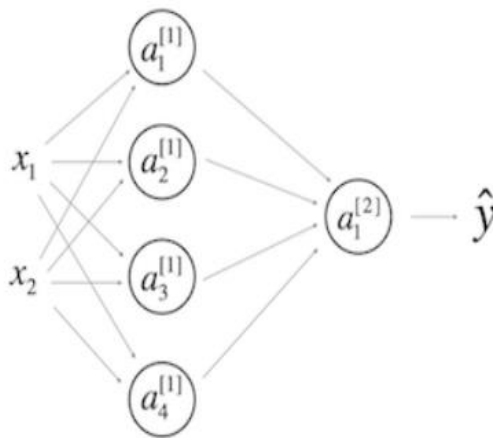
   Which of the following statements are True? (Check all that apply).

   - Answer :
      + b[1] will have shape (4, 1)
      + W[1] will have shape (4, 2)
      + W[2] will have shape (1, 4)
      + b[2] will have shape (1, 1)

10. **In the same network as the previous question, what are the dimensions of Z^[1] and A^[1]?**
   - Answer : Z[1] and A[1] are (4,m)