

Week 3 - Shallow Neural Networks

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

Ans:

- X is a matrix in which each column is one training example.
 - $a_4^{[2]}$ is the activation output by the 4th neuron of the 2nd layer
 - $a^{[2](12)}$ denotes the activation vector of 2nd layer & 12th sample.
 - $a^{[2]}$ denotes the activation vector of the 2nd layer.
2. 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?

Ans: True

- <https://stats.stackexchange.com/a/101563/169377>

3. Which of these is a correct vectorized implementation of forward propagation for layer l , where $1 \leq l \leq L$?

Ans:

- $Z^{[l]} = W^{[l]}A^{[l-1]} + b^{[l]}$
- $A^{[l]} = g(Z^{[l]})$

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?

Ans: Sigmoid

5. Consider the following code. What will be `B.shape`?

```
A = np.random.randn(4,3)
B = np.sum(A, axis = 1, keepdims = True)
```

Ans: `B.shape = (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?

Ans: 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.

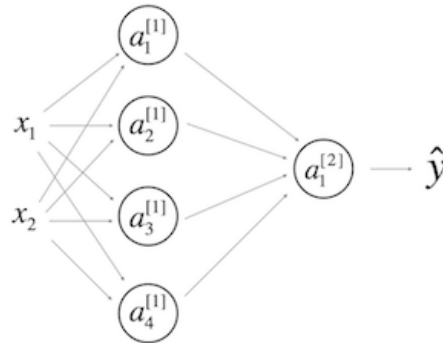
7. Logistic regression's weights w should be initialized randomly rather than to all zeros, because if you initialize to all zeros, then logistic regression will fail to learn a useful decision boundary because it will fail to “break symmetry”, True/False?

Ans: True

8. You have built a network using the tanh activation for all the hidden units. You initialize the weights to relative large values. What will happen?

Ans: 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 statements are true?



Ans:

- $b^{[1]}$ 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 $A^{[1]}$?

Ans: $Z^{[1]}$ and $A^{[1]}$ are $(4, m)$