SHALLOW NEURAL NETWORK

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

- [x] $W^{[1]}$ is a matrix with rows equal to the transpose of the parameter vectors of the fir

  Correct
  Yes. We construct $W^{[1]}$ stacking the parameter vectors $w_j^{[1]}$ of all the neurons of th

- [ ] $w_3^{[4]}$ is the column vector of parameters of the third layer and fourth neuron.
- [ ] W_1 is a matrix with rows equal to the parameter vectors of the first layer.
- [x] $w_3^{[4]}$ is the column vector of parameters of the fourth layer and third neuron.

  Correct
  Yes. The vector $w_j^{[i]}$ is the column vector of parameters of the i-th layer and j-th ne

- [ ] $W^{[1]}$ is a matrix with rows equal to the parameter vectors of the first layer.
- [ ] $w_3^{[4]}$ is the row vector of parameters of the fourth layer and third neuron.

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?

- (•) False
- ( ) True

Correct
Yes. As seen in lecture the output of the tanh is between -1 and 1, it thus centers the data which makes the learning simpler for the next layer.

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

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

Correct

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?

- (●) sigmoid
- ( ) ReLU
- ( ) Leaky ReLU
- ( ) tanh

Correct
Yes. Sigmoid outputs a value between 0 and 1 which makes it a very good choice for binary classification. You can classify as 0 if the output is less than 0.5 and classify as 1 if the output is more than 0.5. It can be done with tanh as well but it is less convenient as the output is between -1 and 1.

5. Consider the following code:

```
#+begin_src python
x = np.random.rand(3, 2)
y = np.sum(x, axis=0, keepdims=True)
#+end_src
```

What will be y.shape?

- ( ) (3, 1)
- (●) (1, 2)
- ( ) (2,)
- ( ) (3,)

Correct
Yes. By choosing the axis=0 the sum is computed over each column of the array, thus the resulting array is a row vector with 2 entries. Since the option keepdims=True is used the first dimension is kept, thus (1, 2).

6. Suppose you have built a neural network with one hidden layer and tanh as activation function for the hidden laye You decide to initialize the weights to small random numbers and the biases to zero. The first hidden layer's neurons will perform different computations from each other even in the first iteration. True/False?

- (●) True

  Yes. Since the weights are most likely different, each neuron will do a different computation.

- ( ) False

  No. Since the weights are most likely different, each neuron will do a different computation.

Correct

7. Using linear activation functions in the hidden layers of a multilayer neural network is equivalent to using a single layer. True/False?

- ( ) False
- (●) True

Correct
Yes. When the identity or linear activation function $g(c) = c$ is used the output of composition of layers is equivalent to the computations made by a single layer.

8.

Which of the following are true about the tanh function?

- [x] The tanh is mathematically a shifted version of the sigmoid function.

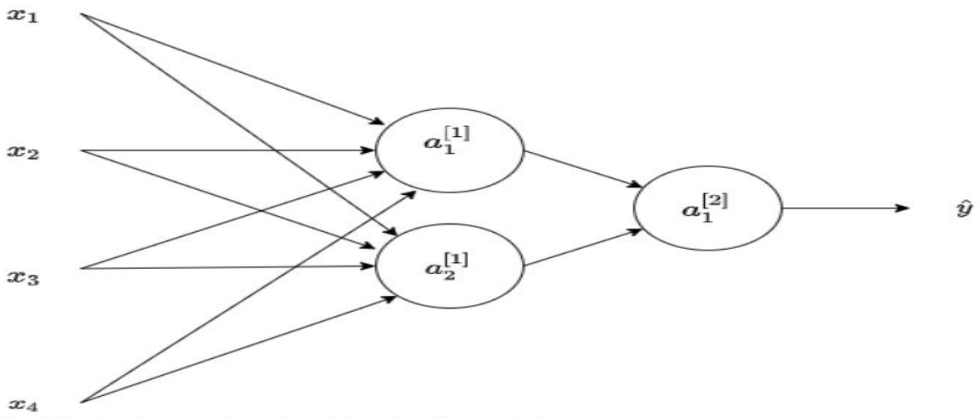
Correct
Yes. You can see the shape of both is very similar but tanh passes through the origin.

- [ ] For large values the slope is larger.
- [ ] The slope is zero for negative values.
- [x] For large values the slope is close to zero.

Correct
Yes. We can see in the graph of the $y = tanh(c)$ how as the values of c increase the curve becomes flatter.

- [ ] The derivative at $c = 0$ is not well defined.

9. Consider the following 1 hidden layer neural network:



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

- [ ] $W^{[2]}$ will have shape (2, 1)
- [ ] $b^{[1]}$ will have shape (4, 2)
- [x] $W^{[1]}$ will have shape (2, 4).

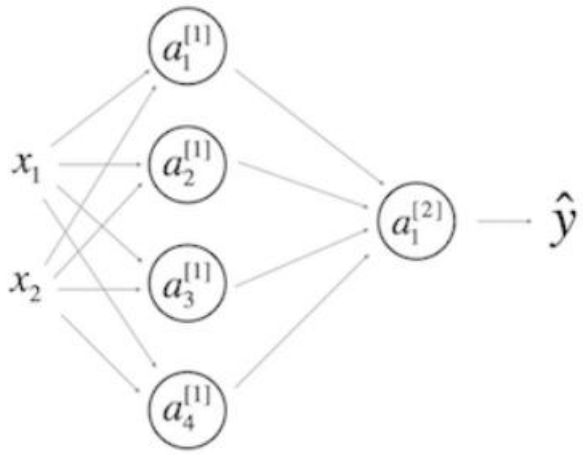
Correct
Yes. The number of rows in $W^{[k]}$ is the number of neurons in the k-th layer and the number of columns is the number of inputs of the layer.

- [x] $b^{[1]}$ will have shape (2, 1).

Correct
Yes. $b^{[k]}$ is a column vector and has the same number of rows as neurons in the k-th layer.

- [ ] $W^{[1]}$ will have shape (4, 2).
- [x] $W^{[2]}$ will have shape (1, 2)

10. What are the dimensions of $Z^{[1]}$ and $A^{[1]}$?



- (●) $Z^{[1]}$ and $A^{[1]}$ are (4,m)
- ( ) $Z^{[1]}$ and $A^{[1]}$ are (4,1)
- ( ) $Z^{[1]}$ and $A^{[1]}$ are (1,4)
- ( ) $Z^{[1]}$ and $A^{[1]}$ are (4,2)

✓ **Correct**