Your grade: 96.66%

Your latest: **96.66%** • Your highest: **96.66%** • To pass you need at least 80%. We keep your highest score.

Next item →

1. Which of the following are true? (Check all that apply.) 0.7 / 1 point

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

  ⊗ **This should not be selected**
  No. The vectors $w_k^{[j]}$ are column vectors.

- [ ] W_1 is a matrix with rows equal to the parameter vectors of the first layer.
- [ ] $w_3^{[4]}$ is the column vector of parameters of the fourth layer and third neuron.
- [x] $W^{[1]}$ is a matrix with rows equal to the transpose of the parameter vectors of the first layer.

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

- [ ] $w_3^{[4]}$ is the column vector of parameters of the third layer and fourth neuron.

2. The sigmoid function is only mentioned as an activation function for historical reasons. The tanh is always preferred without exceptions in all the layers of a Neural Network. True/False?

- ○ True
- ◉ False

✓ **Correct**
Yes. Although the tanh almost always works better than the sigmoid function when used in hidden layers, thus is always proffered as activation function, the exception is for the output layer in classification problems.

3. Which of the following is a correct vectorized implementation of forward propagation for layer 2?

- ○ $Z^{[2]} = W^{[2]} X + b^{[2]}$, $A^{[2]} = g^{[2]}(Z^{[2]})$
- ○ $Z^{[1]} = W^{[1]} X + b^{[1]}$, $A^{[1]} = g^{[1]}(Z^{[1]})$
- ○ $Z^{[2]} = W^{[2]} A^{[1]} + b^{[2]}$, $A^{[2]} = g(Z^{[2]})$
- ◉ $Z^{[2]} = W^{[2]} A^{[1]} + b^{[2]}$, $A^{[2]} = g^{[2]}(Z^{[2]})$

✓ **Correct**
Yes. The elements of layer two are represented using a superscript in brackets.

4. The use of the ReLU activation function is becoming more rare because the ReLU function has no derivative for $c = 0$. True/False?

- ◉ False
- ○ True

✓ **Correct**
Yes. Although the ReLU function has no derivative at $c = 0$ this rarely causes any problems in practice. Moreover it has become the default activation function in many cases, as explained in the lectures.

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).

- ○ (4,)
- ◉ (4, 1)
- ○ (3,)
- ○ (1, 3)

Correct
Yes, we use (keepdims = True) to make sure that A.shape is (4,1) and not (4,). It makes our code more robust.

6. Suppose you have built a neural network with one hidden layer and tanh as activation function for the hidden layers. Which of the following is a best option to initialize the weights?

- ○ Initialize the weights to large random numbers.
- ○ Initialize all weights to 0.
- ◉ Initialize the weights to small random numbers.
- ○ Initialize all weights to a single number chosen randomly.

Correct
The use of random numbers helps to "break the symmetry" between all the neurons allowing them to compute different functions. When using small random numbers the values $z^{[k]}$ will be close to zero thus the activation values will have a larger gradient speeding up the training process.

7. A single output and single layer neural network that uses the sigmoid function as activation is equivalent to the logistic regression. True/False

- ○ False
- ◉ True

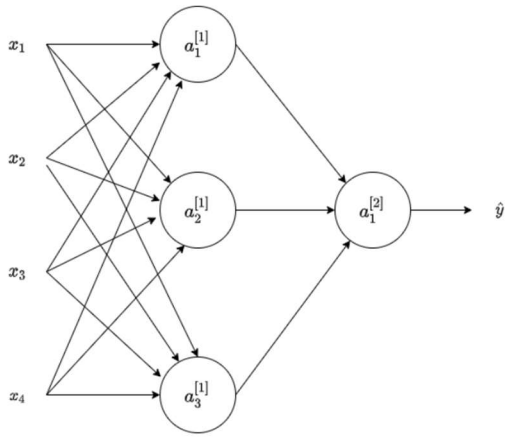
Correct
Yes. The logistic regression model can be expressed by $\hat{y} = \sigma\left(W\,x + b\right)$. This is the same as $a^{[1]} = \sigma(W^{[1]}\,X + b)$.

8. Which of the following is true about the ReLU activation functions?

- ○ They cause several problems in practice because they have no derivative at 0. That is why Leaky ReLU was invented.
- ○ They are only used in the case of regression problems, such as predicting house prices.
- ◉ They are the go to option when you don't know what activation function to choose for hidden layers.
- ○ They are increasingly being replaced by the tanh in most cases.

Correct

9. Consider the following 1 hidden layer neural network:



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

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

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, 3).
- [x] $W^{[1]}$ will have shape (3, 4).

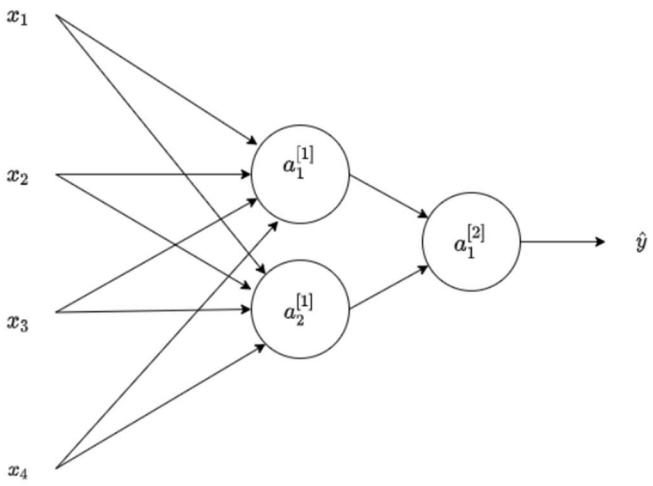
Correct
Yes. The number of rows in $W^{[k]}$ is the number of neurons in the k-th layer and the number of columns is the number of inputs of the layer.

- [x] $b^{[1]}$ will have shape (3, 1).

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

- [ ] $b^{[2]}$ will have shape (3, 1)
- [ ] $b^{[1]}$ will have shape (1, 3)

10. Consider the following 1 hidden layer neural network:



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

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

Correct
Yes. The $Z^{[1]}$ and $A^{[1]}$ are calculated over a batch of training examples. The number of columns in $Z^{[1]}$ and $A^{[1]}$ is equal to the number of examples in the batch, m. And the number of rows in $Z^{[1]}$ and $A^{[1]}$ is equal to the number of neurons in the first layer.