Congratulations! You passed!

Grade received 100% **Latest Submission Grade** 100% **To pass** 80% or higher

Go to next item

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

- [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 row vector of parameters of the fourth layer and third neuron.
- [ ] 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 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 neuron of that layer.

Expand

Correct
Great, you got all the right answers.

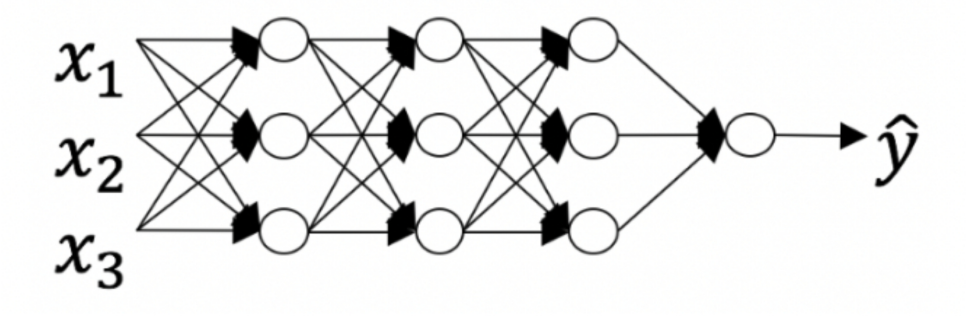
2. In which of the following cases is the linear (identity) activation function most likely used? **1 / 1 point**

- ( ) The linear activation function is never used.
- ( ) As activation function in the hidden layers.
- ( ) For binary classification problems.
- (•) When working with regression problems.

Expand

Correct
Yes. In problems such as predicting the price of a house it makes sense to use the linear activation function as output.

3. Which of the following represents the activation output of the second neuron of the third layer applied to the fourth example? **1 / 1 point**



- $a_2^{[3](4)}$
- $a_2^{[4](3)}$
- $a_4^{[3](2)}$
- $a_3^{[4]2}$

Expand

Correct
Yes. The superscript in brackets indicates the layer number, the superscript in parenthesis represents the number of examples, and the subscript the number of the neuron.

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? **1 / 1 point**

- ReLU
- sigmoid
- tanh
- Leaky ReLU

Expand

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: **1 / 1 point**

#+begin_src python

x = np.random.rand(4, 5)

y = np.sum(x, axis=1)

#+end_src

What will be y.shape?

- (4,)
- (1, 5)
- (5,)
- (4, 1)

Expand

Correct
Yes. By using axis=1 the sum is computed over each row of the array, thus the resulting array is a column vector with 4 entries. Since the option keepdims was not used the array doesn't keep the second dimension.

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? **1 / 1 point**

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

Expand

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. Using linear activation functions in the hidden layers of a multilayer neural network is equivalent to using a single layer. True/False? **1 / 1 point**

- (●) True
- ( ) False

Expand

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.

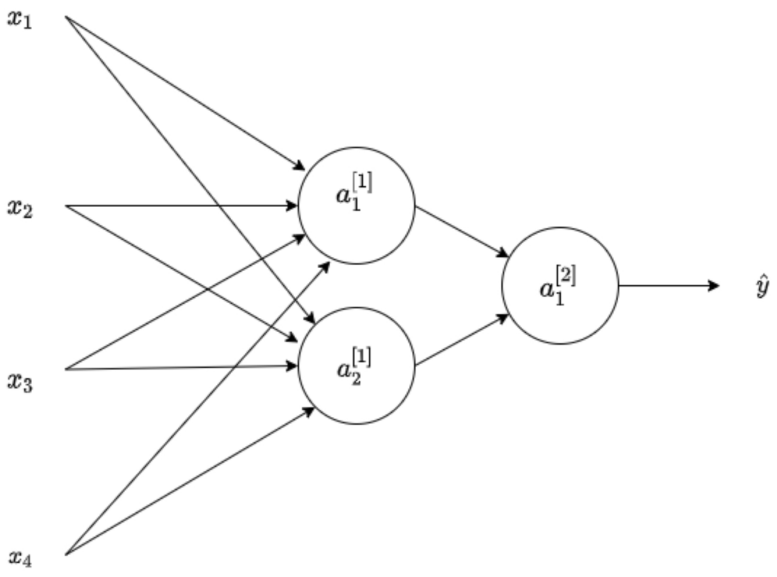
8. Which of the following is true about the ReLU activation functions? **1 / 1 point**

- (●) They are the go to option when you don't know what activation function to choose for hidden layers.
- ( ) 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 increasingly being replaced by the tanh in most cases.

Expand

Correct

9. Consider the following 1 hidden layer neural network: **1 / 1 point**



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

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

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

  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.

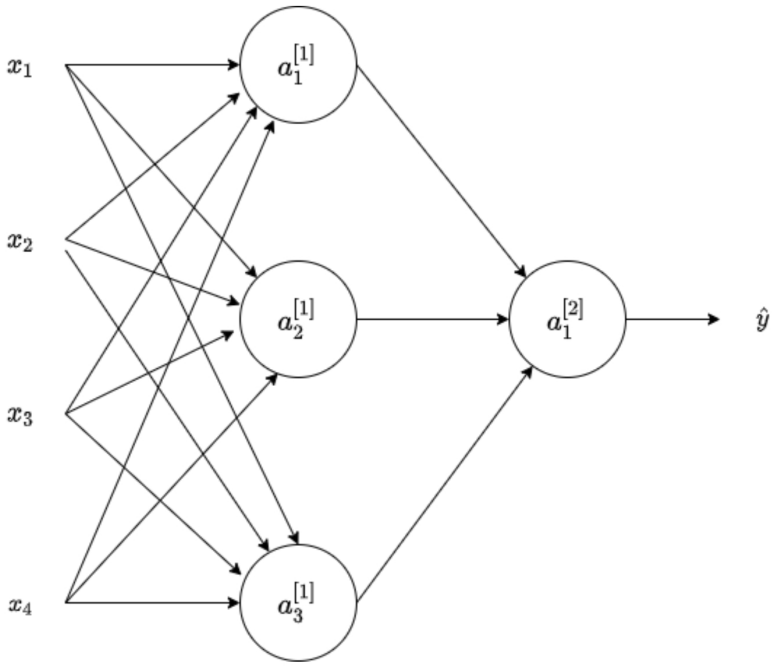
- [ ] $b^{[1]}$ will have shape (4, 2)
- [ ] $W^{[1]}$ will have shape (4, 2).

Expand

Correct
Great, you got all the right answers.

10. Consider the following 1 hidden layer neural network:

1 / 1 point



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

- $Z^{[1]}$ and $A^{[1]}$ are (4, m)
- $Z^{[1]}$ and $A^{[1]}$ are (3, 1)
- $Z^{[1]}$ and $A^{[1]}$ are (3, m)
- $Z^{[1]}$ and $A^{[1]}$ are (4, 1)

Expand

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.