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**

- [ ] $a_3^{[2]}$ denotes the activation vector of the second layer for the third example.
- [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.
- [ ] $a^{[3](2)}$ denotes the activation vector of the second layer for the third example.
- [ ] $w_3^{[4]}$ is the row vector of parameters of the fourth layer and third neuron.
- [ ] $w_3^{[4]}$ is the column vector of parameters of the third layer and fourth neuron.
- [x] $a^{[2]}$ denotes the activation vector of the second layer.

  Correct
  Yes. In our convention $a^{[j]}$ denotes the activation function of the j-th layer.

Expand

Correct
Great, you got all the right answers.

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? **1 / 1 point**

- (•) False
- ( ) True

Expand

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 the following is a correct vectorized implementation of forward propagation for layer 2? **1 / 1 point**

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

Loading [MathJax]/jax/output/CommonHTML/jax.js

Expand

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

4. When building a binary classifier for recognizing cats (y=1) vs raccoons (y=0). Is better to use the sigmoid function as activation function for the hidden layers. True/False **1 / 1 point**

- (•) False
- ( ) True

Expand

Correct
Yes. Using tanh almost always works better than the sigmoid function for hidden layers.

5. Consider the following code: **1 / 1 point**

#+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, 2)
- (3, 1)
- (2,)

Expand

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 layers. Which of the following is a best option to initialize the weights? **1 / 1 point**

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

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. Logistic regression's weights 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? **1 / 1 point**

- True
- False

Expand

Correct
Yes, Logistic Regression doesn't have a hidden layer. If you initialize the weights to zeros, the first example x fed into 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.

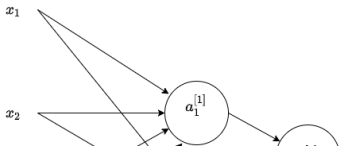
8. Which of the following is true about the ReLU activation functions? **1 / 1 point**

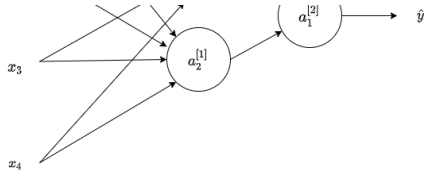
- They cause several problems in practice because they have no derivative at 0. That is why Leaky ReLU was invented.
- They are increasingly being replaced by the tanh in most cases.
- They are the go to option when you don't know what activation function to choose for hidden layers.
- They are only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

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

Correct
Yes. The number of rows in

$W^{[k]}$

- [x] $b^{[1]}$

$b^{[1]}$ 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.

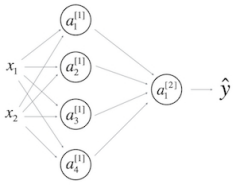
- [ ] $$W^[1]$$ will have shape (4, 2).
- [x] $$W^[1]$$ will have shape (2, 4).

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.

Expand

Correct
Great, you got all the right answers.

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



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

$Z^{[1]}$ and

Expand

Correct