Your grade: 100%

Your latest: **100%** • Your highest: **100%** • 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.) 1 / 1 point

- [x] $a^{[2]}$ denotes the activation vector of the 2^{nd} layer.
  - Correct
- [x] $a_1^{[2]}$ is the activation output by the 4^{th} neuron of the 2^{nd} layer
  - Correct
- [ ] X is a matrix in which each row is one training example.
- [x] $a^{[2](12)}$ denotes the activation vector of the 2^{nd} layer for the 12^{th} training example.
  - Correct
- [x] X is a matrix in which each column is one training example.
  - Correct
- [ ] $a^{[2](12)}$ denotes activation vector of the 12^{th} layer on the 2^{nd} training example.
- [ ] $a_1^{[2]}$ is the activation output of the 2^{nd} layer for the 4^{th} training example

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

- ( ) True
- (•) False

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 $1 \leq l \leq L$? 1 / 1 point

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

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? 1 / 1 point

- ( ) Leaky ReLU
- (•) sigmoid
- ( ) 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: 1 / 1 point

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).

- ( ) (1, 3)
- ( ) (4,)
- (•) (4, 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? 1 / 1 point

- ( ) Initialize the weights to large random numbers.
- (•) Initialize the weights to small random numbers.
- ( ) Initialize all weights to 0.
- ( ) 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 $\approx^{[0]}$ will be close to zero thus the activation values will have a larger gradient speeding up the training process.

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? 1 / 1 point

- ( ) True
- (•) False

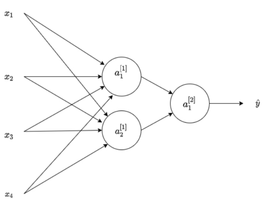
Correct
Yes, Logistic Regression doesn't have a hidden layer. If you initialize the weights to zeros, the first example x fed into the logistic regression will output zero but the derivatives of the Logistic Regression depend on the input x (because there's no hidden layer) which is not zero. So at the 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 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 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.

Correct

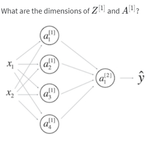
9. Consider the following 1 hidden layer neural network: 1 / 1 point



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

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

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



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

Correct