Shallow Neural Networks

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

- [x] W_1 is a matrix with rows equal to the parameter vectors of the first layer.

  ! **This should not be selected**

  No. The notation convention is that the superscript number in brackets indicates the number of layers.

- [ ] $w_3^{[4]}$ is the row 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.

- [ ] $W^{[1]}$ 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.

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

Expand

Incorrect

You chose the extra incorrect 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

- ○ True
- ◉ False

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

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

Expand

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?

1 / 1 point

- ○ True
- ◉ False

Expand

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:

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,)
- ◉ (1, 2)
- ○ (3,)

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. You decide to initialize the weights and biases to be zero. Which of the following statements is true?

1 / 1 point

- ( ) The first hidden layer's neurons will perform different computations from each other even in the first iteration; their parameters will thus keep evolving in their own way.
- ( ) Each neuron in the first hidden layer will perform the same computation in the first iteration. But after one iteration of gradient descent they will learn to compute different things because we have "broken symmetry".
- ( ) Each neuron in the first hidden layer will compute the same thing, but neurons in different layers will compute different things, thus we have accomplished "symmetry breaking" as described in the lecture.
- (•) Each neuron in the first hidden layer will perform the same computation. So even after multiple iterations of gradient descent, each neuron in the layer will be computing the same thing as other neurons.

Expand

Correct

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

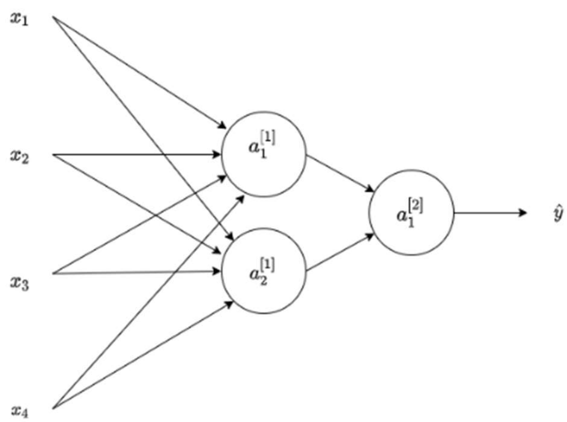
- ( ) True
- (•) False

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.

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



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

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

- [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.

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

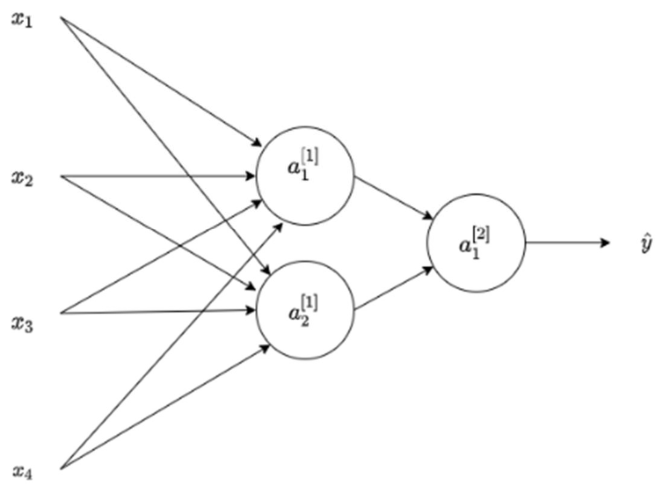
- [ ] $W^{[2]}$ will have shape (2, 1)
- [x] $b^{[1]}$ will have shape (2, 1).

Correct

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

10. Consider the following 1 hidden layer neural network:

1 / 1 point



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

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

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.