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**

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**

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 these is a correct vectorized implementation of forward propagation for layer l, where $1 \leq l \leq L$? **1 / 1 point**

↙↗ **Expand**

✓ **Correct**

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**

↙↗ **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?

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 layer. You decide to initialize the weights to small random numbers and the biases to zero. The first hidden layer's neurons will perform different **1 / 1 point**

computations from each other even in the first iteration. True/False?

Expand

Correct

7. A single output and single layer neural network that uses the sigmoid function as activation is equivalent to the logistic regression. True/False **1 / 1 point**

Expand

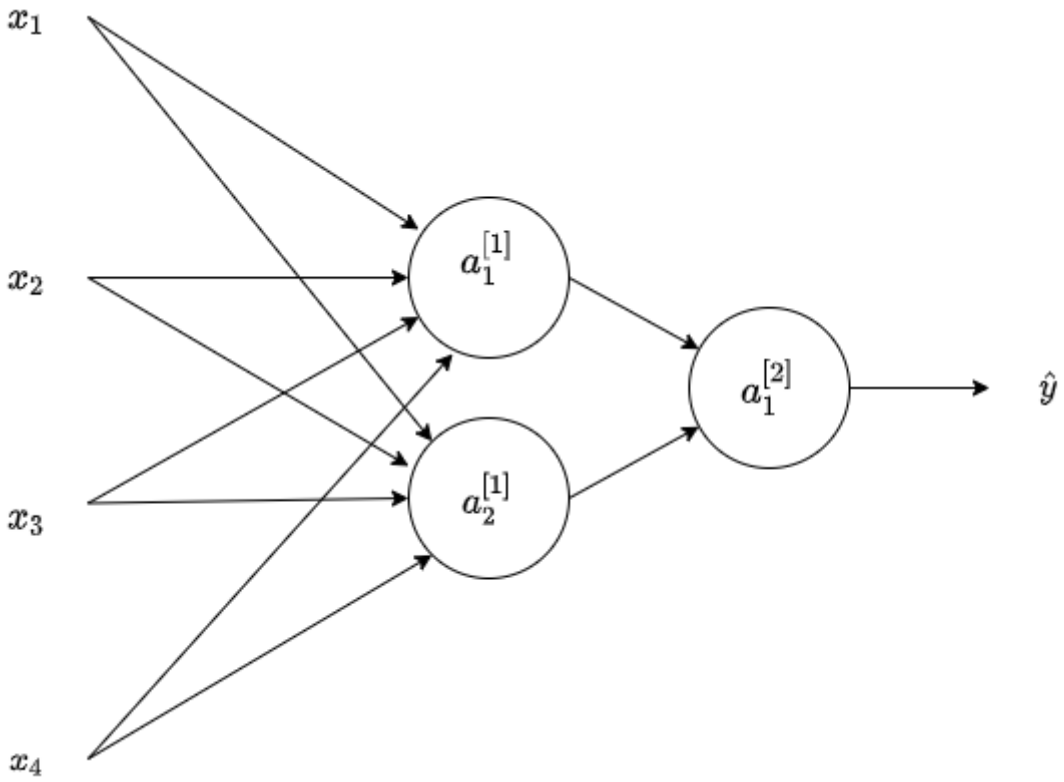
Correct
Yes. The logistic regression model can be expressed by $\hat{y} = \sigma\left(W\,x + b\right)$.
This is the same as $a^{[1]} = \sigma(W^{[1]}\,X + b)$.

8. You have built a network using the tanh activation for all the hidden units. You initialize the weights to relatively large values, using np.random.randn(..,..)*1000. What will happen? **1 / 1 point**

Expand

Correct
Yes. tanh becomes flat for large values; this leads its gradient to be close to zero. This slows down the optimization algorithm.

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

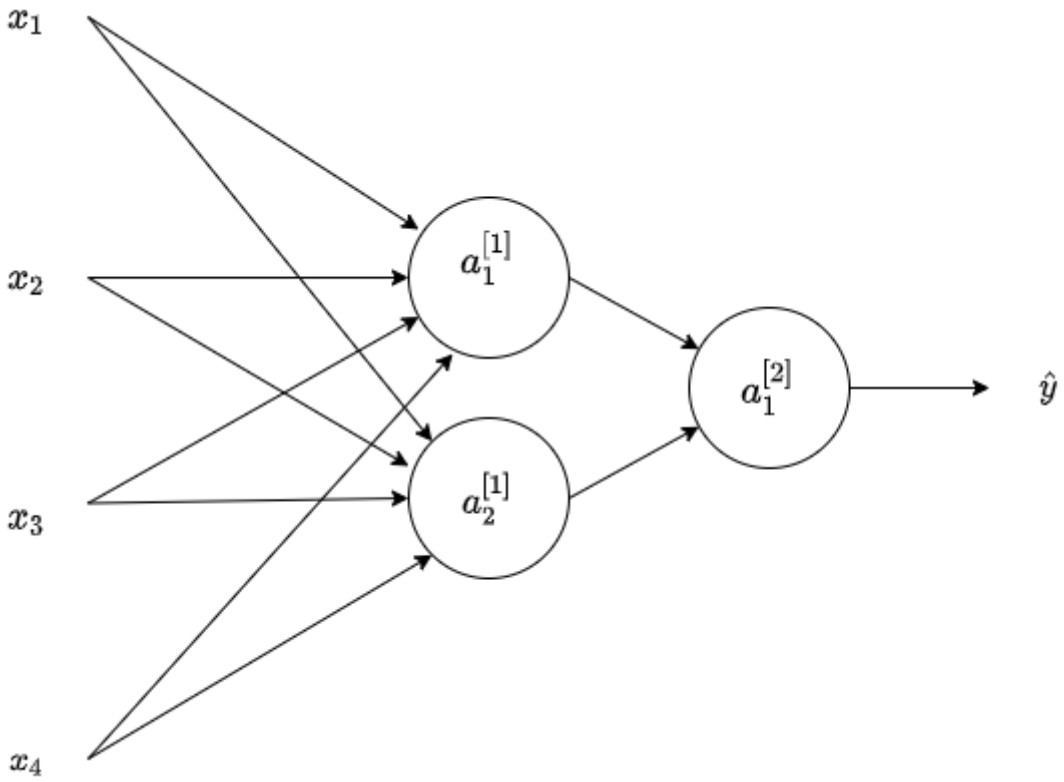


Which of the following statements are True? (Check all that apply).

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 $A^{[1]}$?

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.