grade 100%



TO PASS 80% or higher

## **Key concepts on Deep Neural Networks**

Autors submissions values 10096  1. What is the "cache" used for in our implementation of forward propagation and backward propagation?  ○ We use it to pass variables computed during backward propagation to the corresponding forward propagation step. It contains useful values for forward propagation to compute activations.  ② We use it to pass variables computed during forward propagation to the corresponding backward propagation step. It is used to least the intermediate values of the cost function during training.  ○ It is used to keep track of the hyperparameters that we are searching over, to speed up computation.  ✓ carrect  Correct, the "cache" records values from the forward propagation units and sends it to the backward propagation units because it is needed to compute the chain rule derivatives.  2. Among the following, which ones are "hyperparameters"? (Check all that apply.)  ✓ carrect  ✓ correct  ✓ size of the hidden layers n. I'll  ✓ correct  ✓ correct  ✓ correct  ✓ correct  ✓ correct  Only the deport layers of a neural network are typically computing more complex features of the input than the earlier layers.  ○ The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.  ✓ correct  ✓ correct  On the desper layers of a neural network are typically computing more complex features of the input than the deeper layers.  ✓ correct  ✓ correct  ✓ correct  On the desper layers of a neural network are typically computing more complex features of the input than the deeper layers.  ✓ correct		- <b>,</b>
<ul> <li>We use it to pass variables computed during backward propagation to the corresponding forward propagation step. It contains useful values for forward propagation to compute activations.</li> <li>We use it to pass variables computed during forward propagation to compute during the contains useful values for backward propagation to compute derivatives.</li> <li>It is used to cache the intermediate values of the cost function during training.</li> <li>It is used to keep track of the hyperparameters that we are searching over, to speed up computation.</li> <li>✓ correct</li> <li>Correct. the "cache" records values from the forward propagation units and sends it to the backward propagation units because it is needed to compute the chain rule derivatives.</li> <li>2. Among the following, which ones are "hyperparameters"? (Check all that apply.)</li> <li>✓ learning rate α</li> <li>✓ correct</li> <li>✓ size of the hidden layers n.<sup>[1]</sup></li> <li>✓ correct</li> <li>In blas vectors b.<sup>[7]</sup></li> <li>✓ number of tierations</li> <li>✓ correct</li> <li>In weight matrices W.<sup>[7]</sup></li> <li>✓ number of layers L in the neural network</li> <li>✓ correct</li> <li>In the deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.</li> <li>✓ The deeper layers of a neural network are typically computing more complex features of the input than the deeper layers.</li> <li>✓ Correct</li> <li>✓ Correct</li> <li>On the earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.</li> <li>✓ Correct</li> <li< td=""><td></td><td></td></li<></ul>		
step. It contains useful values for forward propagation to compute deviations.  ● We use it to pass variables computed during forward propagation to the corresponding backward propagation step. It contains useful values for backward propagation to compute derivatives.  ○ It is used to cache the intermediate values of the cost function during training.  ○ It is used to keep track of the hyperparameters that we are searching over, to speed up computation.  ✓ Carrect  Correct, the "cache" records values from the forward propagation units and sends it to the backward propagation units because it is needed to compute the chain rule derivatives.  2. Among the following, which ones are "hyperparameters"? (Check all that apply.)  ✓ Carrect  ✓ size of the hidden layers n <sup>   </sup> ✓ correct  □ blas vectors b <sup>   </sup> ✓ number of iterations  ✓ correct  □ weight matrices W <sup>   </sup> ✓ number of layers L in the neural network  ✓ correct  □ activation values a <sup>   </sup> 3. Which of the following statements is true?  ⑤ The deeper layers of a neural network are typically computing more complex features of the input than the deeper layers.  ✓ Carrect  ✓ Carrect	1.	What is the "cache" used for in our implementation of forward propagation and backward propagation?
step. It contains useful values for backward propagation to compute derivatives.  It is used to cache the intermediate values of the cost function during training.  It is used to keep track of the hyperparameters that we are searching over, to speed up computation.  ✓ correct  Correct, the "cache" records values from the forward propagation units and sends it to the backward propagation units because it is needed to compute the chain rule derivatives.  2. Among the following, which ones are "hyperparameters"? (Check all that apply.)  ✓ correct  ✓ correct  ✓ size of the hidden layers n <sup>[1]</sup> ✓ correct  □ bias vectors b <sup>[1]</sup> ☑ number of iterations  ✓ correct  □ weight matrices W <sup>[1]</sup> ☑ number of layers L in the neural network  ✓ correct  □ activation values a <sup>[1]</sup> 3. Which of the following statements is true?  ⑥ The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.  ○ The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.  ✓ correct		
<ul> <li>It is used to keep track of the hyperparameters that we are searching over, to speed up computation.</li> <li>✓ correct</li></ul>		
<ul> <li>Correct         Correct, the "cache" records values from the forward propagation units and sends it to the backward propagation units because it is needed to compute the chain rule derivatives.</li> <li>Among the following, which ones are "hyperparameters"? (Check all that apply.)</li> <li>Learning rate α</li> <li>✓ correct</li> <li>size of the hidden layers n<sup>[j]</sup></li> <li>✓ correct</li> <li>bias vectors b<sup>[j]</sup></li> <li>number of iterations</li> <li>✓ correct</li> <li>weight matrices W<sup>[j]</sup></li> <li>number of layers L in the neural network</li> <li>✓ correct</li> <li>activation values a<sup>[j]</sup></li> <li>Which of the following statements is true?</li> <li>The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.</li> <li>The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.</li> <li>✓ correct</li> <li>Vectorization allows you to compute forward propagation in an L-layer neural network without an explicit for-loop (or an other explicit iterative loop) over the layers [=1, 2,, L. True/False?</li> <li>True</li> </ul>		It is used to cache the intermediate values of the cost function during training.
Correct. The "cache" records values from the forward propagation units and sends it to the backward propagation units because it is needed to compute the chain rule derivatives.  2. Among the following, which ones are "hyperparameters"? (Check all that apply.)  □ learning rate α  □ correct □ size of the hidden layers n <sup>[0]</sup> □ number of iterations □ weight matrices W <sup>[1]</sup> □ number of layers <i>L</i> in the neural network □ activation values a <sup>[0]</sup> □ the deeper layers of a neural network are typically computing more complex features of the input than the earlier layers. □ the earlier layers of a neural network are typically computing more complex features of the input than the deeper layers. □ the earlier layers of a neural network are typically computing more complex features of the input than the deeper layers. □ the earlier layers of a neural network are typically computing more complex features of the input than the deeper layers. □ the earlier layers of a neural network are typically computing more complex features of the input than the deeper layers. □ the earlier layers of a neural network are typically computing more complex features of the input than the deeper layers. □ the earlier layers of a neural network are typically computing more complex features of the input than the deeper layers. □ the earlier layers of a neural network are typically computing more complex features of the input than the deeper layers. □ the earlier layers of a neural network are typically computing more complex features of the input than the deeper layers. □ the earlier layers of a neural network are typically computing more complex features of the input than the deeper layers. □ the earlier layers of a neural network are typically computing more complex features of the input than the deeper layers. □ the earlier layers of a neural network are typically computing more complex features of the input than the deeper layers. □ the earlier layers of a neural network are typically computing more complex features of the input than the d		It is used to keep track of the hyperparameters that we are searching over, to speed up computation.
<ul> <li>☑ learning rate α</li> <li>✓ correct</li> <li>☑ size of the hidden layers n<sup>[I]</sup></li> <li>✓ correct</li> <li>☐ bias vectors b<sup>[I]</sup></li> <li>☑ number of iterations</li> <li>✓ correct</li> <li>☐ weight matrices W<sup>[I]</sup></li> <li>☑ number of layers L in the neural network</li> <li>✓ correct</li> <li>☐ activation values a<sup>[I]</sup></li> <li>3. Which of the following statements is true?</li> <li>⑥ The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.</li> <li>○ The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.</li> <li>✓ correct</li> <li>4. Vectorization allows you to compute forward propagation in an L-layer neural network without an explicit for-loop (or an other explicit iterative loop) over the layers I=1, 2,, L. True/False?</li> <li>○ True</li> </ul>		Correct, the "cache" records values from the forward propagation units and sends it to the backward
<ul> <li>✓ correct</li> <li>☑ size of the hidden layers n<sup>[I]</sup></li> <li>✓ correct</li> <li>□ bias vectors b<sup>[I]</sup></li> <li>☑ number of iterations</li> <li>✓ correct</li> <li>□ weight matrices W<sup>[I]</sup></li> <li>☑ number of layers L in the neural network</li> <li>✓ correct</li> <li>□ activation values a<sup>[I]</sup></li> <li>3. Which of the following statements is true?</li> <li>④ The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.</li> <li>○ The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.</li> <li>✓ correct</li> <li>4. Vectorization allows you to compute forward propagation in an L-layer neural network without an explicit for-loop (or an other explicit iterative loop) over the layers i=1, 2,, L. True/False?</li> <li>○ True</li> </ul>	2.	Among the following, which ones are "hyperparameters"? (Check all that apply.)
<ul> <li>✓ correct</li> <li>bias vectors b<sup>[i]</sup></li> <li>✓ number of iterations</li> <li>✓ correct</li> <li>weight matrices W<sup>[i]</sup></li> <li>✓ number of layers L in the neural network</li> <li>✓ correct</li> <li>activation values a<sup>[i]</sup></li> <li>3. Which of the following statements is true?</li> <li>The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.</li> <li>The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.</li> <li>✓ correct</li> <li>4. Vectorization allows you to compute forward propagation in an L-layer neural network without an explicit for-loop (or an other explicit iterative loop) over the layers i=1, 2,,L. True/False?</li> <li>True</li> </ul>		lacksquare learning rate $lpha$
<ul> <li>✓ correct</li> <li>□ bias vectors b<sup>[I]</sup></li> <li>✓ number of iterations</li> <li>✓ correct</li> <li>□ weight matrices W<sup>[I]</sup></li> <li>✓ number of layers L in the neural network</li> <li>✓ correct</li> <li>□ activation values a<sup>[I]</sup></li> <li>3. Which of the following statements is true?</li> <li>④ The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.</li> <li>○ The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.</li> <li>✓ correct</li> <li>4. Vectorization allows you to compute forward propagation in an L-layer neural network without an explicit for-loop (or an other explicit iterative loop) over the layers l=1, 2,,L. True/False?</li> <li>○ True</li> </ul>		✓ Correct
<ul> <li>bias vectors b<sup>[I]</sup></li> <li>✓ number of iterations</li> <li>✓ correct</li> <li>weight matrices W<sup>[I]</sup></li> <li>number of layers L in the neural network</li> <li>✓ correct</li> <li>activation values a<sup>[I]</sup></li> <li>Which of the following statements is true?</li> <li>The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.</li> <li>The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.</li> <li>✓ correct</li> <li>Vectorization allows you to compute forward propagation in an L-layer neural network without an explicit for-loop (or an other explicit iterative loop) over the layers i=1, 2,,L. True/False?</li> <li>True</li> </ul>		$m{arphi}$ size of the hidden layers $n^{[l]}$
<ul> <li>✓ correct</li> <li>weight matrices W<sup>[I]</sup></li> <li>number of layers L in the neural network</li> <li>✓ correct</li> <li>activation values a<sup>[I]</sup></li> <li>Which of the following statements is true?</li> <li>The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.</li> <li>The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.</li> <li>✓ correct</li> <li>Vectorization allows you to compute forward propagation in an L-layer neural network without an explicit for-loop (or an other explicit iterative loop) over the layers l=1, 2,,L. True/False?</li> <li>True</li> </ul>		✓ Correct
<ul> <li>✓ correct</li> <li>weight matrices W<sup>[l]</sup></li> <li>In unmber of layers L in the neural network</li> <li>✓ correct</li> <li>activation values a<sup>[l]</sup></li> <li>Which of the following statements is true?</li> <li>The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.</li> <li>The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.</li> <li>✓ correct</li> <li>Vectorization allows you to compute forward propagation in an L-layer neural network without an explicit for-loop (or an other explicit iterative loop) over the layers !=1, 2,,L. True/False?</li> <li>True</li> </ul>		$lacksquare$ bias vectors $b^{[l]}$
<ul> <li>weight matrices W<sup>[l]</sup></li> <li>✓ number of layers <i>L</i> in the neural network</li> <li>✓ correct</li> <li>activation values a<sup>[l]</sup></li> <li>Which of the following statements is true?</li> <li>The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.</li> <li>The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.</li> <li>✓ correct</li> <li>Vectorization allows you to compute forward propagation in an <i>L</i>-layer neural network without an explicit for-loop (or an other explicit iterative loop) over the layers  =1, 2,,L. True/False?</li> <li>True</li> </ul>		✓ number of iterations
<ul> <li>✓ correct</li> <li>activation values a <sup>[I]</sup></li> <li>Which of the following statements is true?</li> <li>The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.</li> <li>The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.</li> <li>✓ correct</li> <li>Vectorization allows you to compute forward propagation in an <i>L</i>-layer neural network without an explicit for-loop (or an other explicit iterative loop) over the layers l=1, 2,,L. True/False?</li> <li>True</li> </ul>		✓ Correct
<ul> <li>✓ correct</li> <li>activation values a<sup>[I]</sup></li> <li>Which of the following statements is true?</li> <li>The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.</li> <li>The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.</li> <li>✓ correct</li> <li>Vectorization allows you to compute forward propagation in an <i>L</i>-layer neural network without an explicit for-loop (or an other explicit iterative loop) over the layers l=1, 2,,L. True/False?</li> <li>True</li> </ul>		$lacksquare$ weight matrices $W^{[l]}$
<ul> <li>□ activation values a<sup>[I]</sup></li> <li>3. Which of the following statements is true?</li> <li>⑤ The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.</li> <li>○ The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.</li> <li>✓ correct</li> <li>4. Vectorization allows you to compute forward propagation in an <i>L</i>-layer neural network without an explicit for-loop (or an other explicit iterative loop) over the layers l=1, 2,,L. True/False?</li> <li>○ True</li> </ul>		lacksquare number of layers $L$ in the neural network
<ul> <li>Which of the following statements is true?</li> <li>The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.</li> <li>The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.</li> <li>Correct</li> <li>Vectorization allows you to compute forward propagation in an <i>L</i>-layer neural network without an explicit for-loop (or an other explicit iterative loop) over the layers l=1, 2,,L. True/False?</li> <li>True</li> </ul>		✓ Correct
<ul> <li>The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.</li> <li>The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.</li> <li>Correct</li> <li>Vectorization allows you to compute forward propagation in an <i>L</i>-layer neural network without an explicit for-loop (or an other explicit iterative loop) over the layers l=1, 2,,L. True/False?</li> <li>True</li> </ul>		$igsqcup$ activation values $a^{[l]}$
layers.  The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.  Correct  Vectorization allows you to compute forward propagation in an <i>L</i> -layer neural network without an explicit for-loop (or an other explicit iterative loop) over the layers l=1, 2,,L. True/False?  True	3.	Which of the following statements is true?
layers.  ✓ correct  4. Vectorization allows you to compute forward propagation in an <i>L</i> -layer neural network without an explicit for-loop (or an other explicit iterative loop) over the layers l=1, 2,,L. True/False?  ○ True		
4. Vectorization allows you to compute forward propagation in an <i>L</i> -layer neural network without an explicit for-loop (or an other explicit iterative loop) over the layers l=1, 2,,L. True/False? True		•
other explicit iterative loop) over the layers l=1, 2,,L. True/False?  True		✓ Correct
○ True	4.	

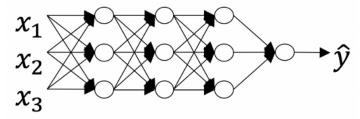


Forward propagation propagates the input through the layers, although for shallow networks we may just write all the lines  $(a^{[2]}=g^{[2]}(z^{[2]}),z^{[2]}=W^{[2]}a^{[1]}+b^{[2]},...)$  in a deeper network, we cannot avoid a for loop iterating over the layers:  $(a^{[l]}=g^{[l]}(z^{[l]}),z^{[l]}=W^{[l]}a^{[l-1]}+b^{[l]},...)$ .

5. Assume we store the values for  $n^{[l]}$  in an array called layer\_dims, as follows: layer\_dims =  $[n_x, 4, 3, 2, 1]$ . So layer 1 has four 1/1 point hidden units, layer 2 has 3 hidden units and so on. Which of the following for-loops will allow you to initialize the parameters for the model?

```
for i in range(1, len(layer_dims)/2):
0
                 parameter['\u00fc'\u00e4' + str(i)] = np.random.randn(layer_dims[i], layer_dims[i-1]) *
parameter['\u00fc' + str(i)] = np.random.randn(layer_dims[i], 1) * 0.01
0
             for i in range(1, len(layer_dims)/2):
                 for i in range(1, len(layer_dims)):
    parameter['W' + str(i)] = np.random.randn(layer_dims[i-1], layer_dims[i]) *
    parameter['b' + str(i)] = np.random.randn(layer_dims[i], 1) * 0.01
0
             for i in range(1, len(layer_dims)):
•
                 ✓ Correct
```

6. Consider the following neural network.



How many layers does this network have?

- lacktriangle The number of layers L is 4. The number of hidden layers is 3.
- $\bigcirc$  The number of layers L is 3. The number of hidden layers is 3.
- $\bigcirc$  The number of layers L is 4. The number of hidden layers is 4.
- The number of layers L is 5. The number of hidden layers is 4.

✓ Correct

Yes. As seen in lecture, the number of layers is counted as the number of hidden layers + 1. The input and output layers are not counted as hidden layers.

7. During forward propagation, in the forward function for a layer l you need to know what is the activation function in a layer (Sigmoid, tanh, ReLU, etc.). During backpropagation, the corresponding backward function also needs to know what is the activation function for layer l, since the gradient depends on it. True/False?

True

Yes, as you've seen in the week 3 each activation has a different derivative. Thus, during backpropagation you need to know which activation was used in the forward propagation to be able to compute the correct

8. There are certain functions with the following properties:

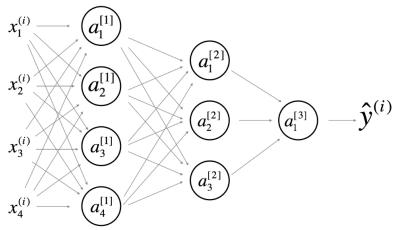
1 / 1 point

(i) To compute the function using a shallow network circuit, you will need a large network (where we measure size by the number of logic gates in the network), but (ii) To compute it using a deep network circuit, you need only an exponentially smaller network. True/False?

True

9. Consider the following 2 hidden layer neural network:

1 / 1 point



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

 $lacksquare W^{[1]}$  will have shape (4, 4)

 $\checkmark$  correct Yes. More generally, the shape of  $W^{[l]}$  is  $(n^{[l]}, n^{[l-1]})$ .

 $lacksquare b^{[1]}$  will have shape (4, 1)

 $\checkmark$  correct Yes. More generally, the shape of  $b^{[l]}$  is  $oldsymbol{(n^{[l]},1)}.$ 

 $\ \ \ \ \ \ b^{[1]}$  will have shape (3, 1)

 $lacksquare W^{[2]}$  will have shape (3, 4)

 $\checkmark$  Correct Yes. More generally, the shape of  $W^{[l]}$  is  $\left(n^{[l]},n^{[l-1]}
ight)$ .

 $\ \ \ \ \ \ b^{[2]}$  will have shape (1, 1)

lacksquare  $b^{[2]}$  will have shape (3, 1)

 $\checkmark$  Correct Yes. More generally, the shape of  $b^{[l]}$  is  $(n^{[l]},1)$ .

 $\qquad \qquad W^{[3]}$  will have shape (3, 1)

lacksquare  $b^{[3]}$  will have shape (1, 1)

Yes. More generally, the shape of  $b^{[l]}$  is  $(n^{[l]},1)$ .

 $lacksquare W^{[3]}$  will have shape (1, 3)

 $\checkmark$  Correct Yes. More generally, the shape of  $W^{[l]}$  is  $(n^{[l]}, n^{[l-1]}).$ 

10. Whereas the previous question used a specific network, in the general case what is the dimension of W^{[i]}, the weight matrix associated with layer *l*?

 $W^{[l]}$  has shape  $(n^{[l+1]}, n^{[l]})$   $W^{[l]}$  has shape  $(n^{[l]}, n^{[l-1]})$   $W^{[l]}$  has shape  $(n^{[l]}, n^{[l+1]})$