Congratulations! You passed!

Grade received 90% | Latest Submission Grade 90% | To pass 80% or higher

Go to next item

1. What is the "cache" used for in our implementation of forward propagation and backward propagation? 1 / 1 point

- 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.
- (selected) We use it to pass Z 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.

Expand

Correct
Correct, the "cache" records values from the forward propagation units and are used in backward propagation units because it is needed to compute the chain rule derivatives.

2. Which of the following are "parameters" of a neural network? (Check all that apply.) 1 / 1 point

- (checked) $W^{[l]}$ the weight matrices.
  - Correct
    Correct. The weight matrices and the bias vectors are the parameters of the network.
- (checked) $b^{[l]}$ the bias vector.
  - Correct
    Correct. The weight matrices and the bias vectors are the parameters of the network.
- ☐ L the number of layers of the neural network.
- ☐ $g^{[l]}$ the activation functions.

Expand

Correct
Great, you got all the right answers.

3. Considering the intermediate results below, which layers of a deep neural network are they likely to belong to? 1 / 1 point

- Input layer of the deep neural network.
- (selected) Later layers of the deep neural network.
- Early layers of the deep neural network.
- Middle layers of the deep neural network.

Expand

Correct
Correct. The deep layers of a neural network are typically computing more complex features such as the ones shown in the figure.

4. Vectorization allows us to compute $a^{[l]}$ for all the examples on a batch at the same time without using a for loop. True/False? 1 / 1 point

- False

- (●) True

Expand

Correct
Correct. Vectorization allows us to compute the activation for all the training examples at the same time, avoiding the use of a for loop.

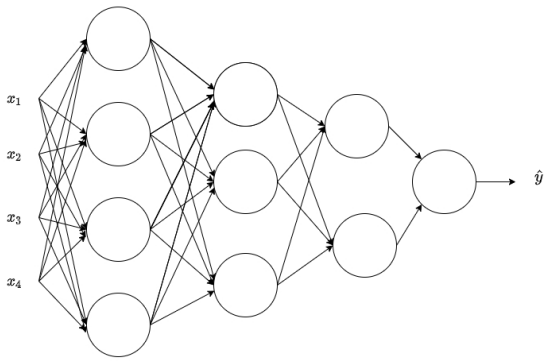
5. Suppose W[i] is the array with the weights of the i-th layer, b[i] is the vector of biases of the i-th layer, and g is the activation function used in all layers. Which of the following calculates the forward propagation for the neural network with L layers. **1 / 1 point**

- ( ) for i in range(L):
  Z[i] = W[i]*X + b[i]
  A[i] = g(Z[i])
- ( ) for i in range(1, L):
  Z[i] = W[i]*A[i-1] + b[i]
  A[i] = g(Z[i])
- (●) for i in range(1, L+1):
  Z[i] = W[i]*A[i-1] + b[i]
  A[i] = g(Z[i])
- ( ) for i in range(L):
  Z[i+1] = W[i+1]*A[i+1] + b[i+1]
  A[i+1] = g(Z[i+1])

Expand

Correct
Yes. Remember that the range omits the last number thus the range from 1 to L+1 gives the L necessary values.

6. Consider the following neural network: **1 / 1 point**



What are all the values of $n^{[0]}, n^{[1]}, n^{[2]}, n^{[3]}$ and $n^{[4]}$?

- ( ) 4, 4, 3, 2
- ( ) 4, 3, 2
- ( ) 4, 3, 2, 1
- (●) 4, 4, 3, 2, 1

Expand

Correct
Yes. The $n^{[l]}$ are the number of units in each layer, notice that $n^{[0]} = n_x$.

7. If L is the number of layers of a neural network then $dZ^{[L]} = A^{[L]} - Y$. True/False? **0 / 1 point**

- (●) False
- ( ) True

Expand

Incorrect
Incorrect. No. The gradient of the output layer depends on the difference between the value computed during the forward propagation process and the target values.

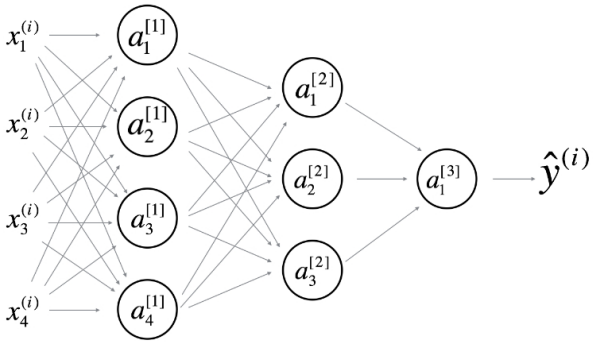
8. For any mathematical function you can compute with an L-layered deep neural network with N hidden units there is a shallow neural network that requires only $\log N$ units, but it is very difficult to train. **1 / 1 point**

- ○ True
- ◉ False

Expand

Correct
Correct. On the contrary, some mathematical functions can be computed using an L-layered neural network and a given number of hidden units; but using a shallow neural network the number of necessary hidden units grows exponentially.

9. Consider the following 2 hidden layer neural network: **1 / 1 point**



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

- ☐ $W^{[2]}$ will have shape (3, 1)
- ☑ $W^{[1]}$ will have shape (4, 4)

  ✓ **Correct**
  Yes. More generally, the shape of $W^{[l]}$ is $(n^{[l]}, n^{[l-1]})$.
- ☑ $b^{[1]}$ will have shape (4, 1)

  ✓ **Correct**
  Yes. More generally, the shape of $b^{[l]}$ is $(n^{[l]}, 1)$.
- ☑ $b^{[2]}$ will have shape (3, 1)

  ✓ **Correct**
  Yes. More generally, the shape of $b^{[l]}$ is $(n^{[l]}, 1)$.
- ☑ $b^{[3]}$ will have shape (1, 1)

  ✓ **Correct**
  Yes. More generally, the shape of $b^{[l]}$ is $(n^{[l]}, 1)$.
- ☐ $b^{[3]}$ will have shape (3, 1)
- ☐ $b^{[2]}$ will have shape (1, 1)
- ☐ $b^{[1]}$ will have shape (3, 1)
- ☑ $W^{[3]}$ will have shape (1, 3)

  ✓ **Correct**
  Yes. More generally, the shape of $W^{[l]}$ is $(n^{[l]}, n^{[l-1]})$.
- ☐ $W^{[1]}$ will have shape (3, 4)
- ☑ $W^{[2]}$ will have shape (3, 4)

  ✓ **Correct**
  Yes. More generally, the shape of $W^{[l]}$ is $(n^{[l]}, n^{[l-1]})$.
- ☐ $W^{[3]}$ will have shape (3, 1)

Expand

Correct
Great, you got all the right answers.

10. Whereas the previous question used a specific network, in the general case what is the dimension of W^{[l]}, the weight matrix associated with layer l? **1 / 1 point**

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

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

Expand

Correct
True