Your grade: 90%

Your latest: **90%** • Your highest: **90%**
To pass you need at least 80%. We keep your highest score.

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 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.
- ( ) 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.

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.

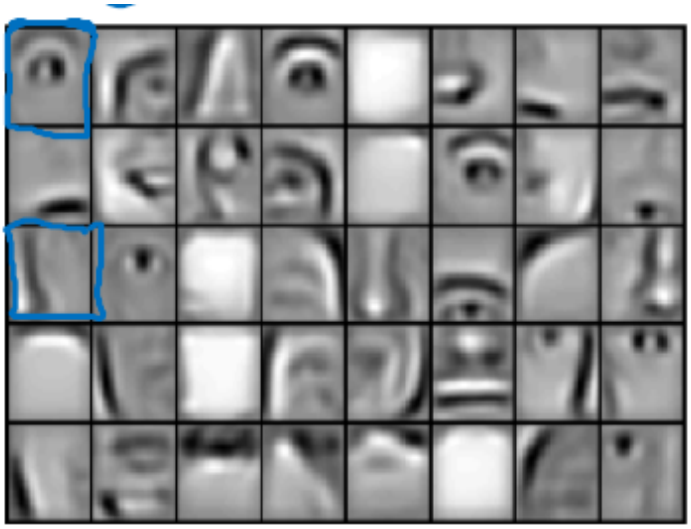
2. During the backpropagation process, we use gradient descent to change the hyperparameters. True/False? **1 / 1 point**

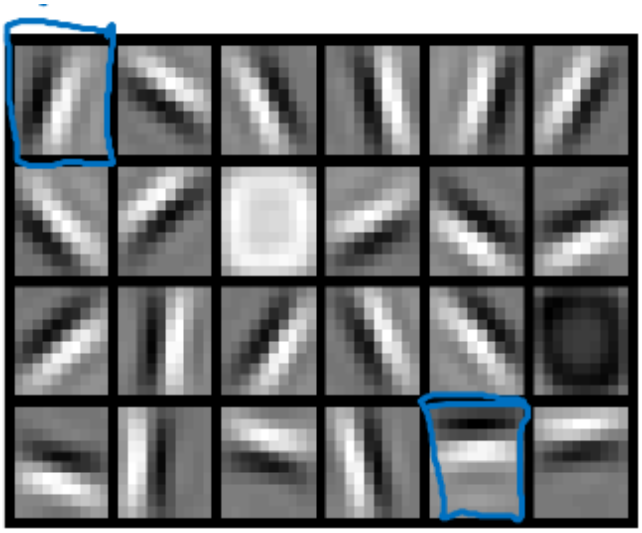
- ( ) True

False

Correct
Correct. During backpropagation, we use gradient descent to compute new values of $W^{[l]}$ and $b^{[l]}$. These are the parameters of the network.

3. Which of the following is more likely related to the early layers of a deep neural network? **1 / 1 point**





Correct
Yes. The early layer of a neural network usually computes simple features such as edges and lines.

4. We can not use vectorization to calculate $da^{[l]}$ in backpropagation, we must use a for loop over all the examples. True/False? **1 / 1 point**

- True
- False

Correct
Correct. We can use vectorization in backpropagation to calculate $dA^{[l]}$ for each layer. This computation is done over all the training examples.

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**

- [x]

```
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(1, L):
 Z[i] = W[i]*A[i-1] + b[i]
 A[i] = g(Z[i])
```

- [ ]

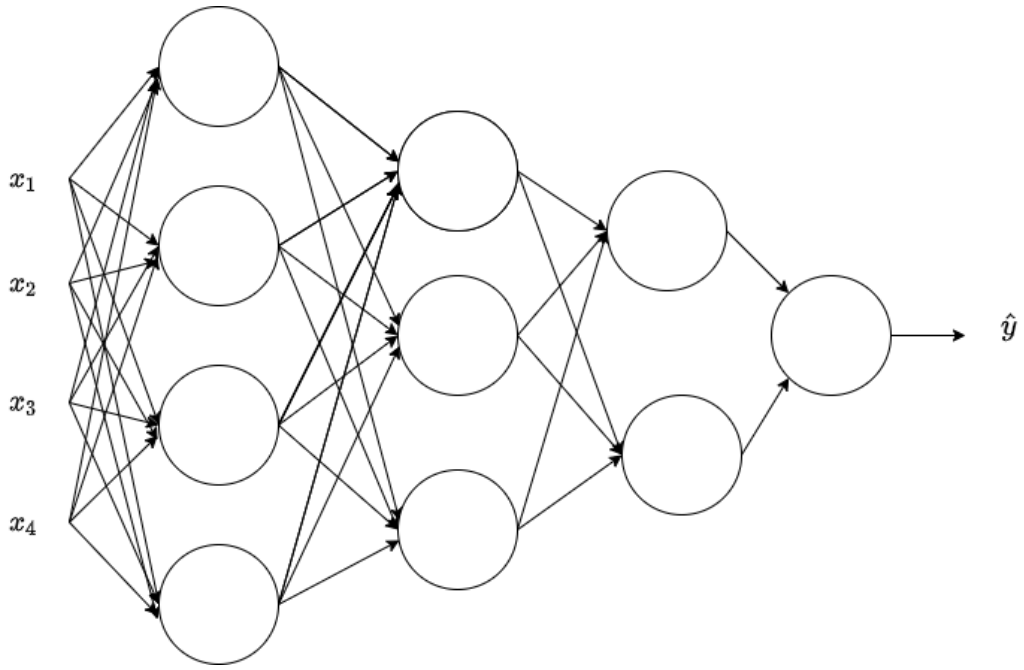
```
for i in range(L):
 Z[i+1] = W[i+1]*A[i+1] + b[i+1]
 A[i+1] = g(Z[i+1])
```

- [ ]

```
for i in range(L):
 Z[i] = W[i]*X + b[i]
 A[i] = g(Z[i])
```

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 $n^{[4]}$?

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

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

7. During forward propagation, to calculate $A^{[l]}$, you use the activation function $g^{[l]}$ with the values of $Z^{[l]}$. **1 point**

True/False: During backward propagation, you calculate $dA^{[l]}$ from $Z^{[l]}$.

- False
- (selected) True

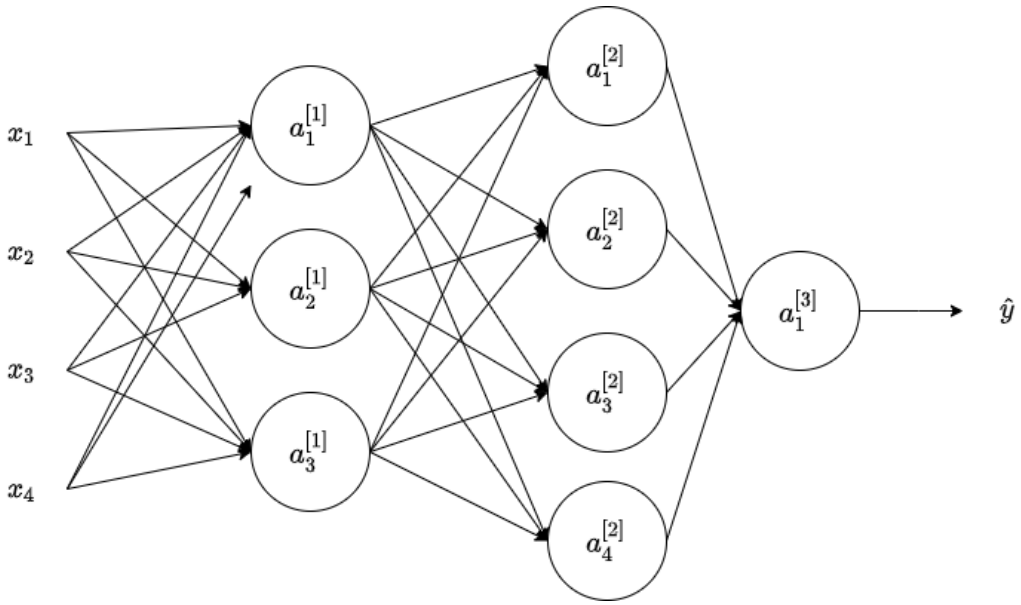
Incorrect
Incorrect. During backward propagation we are interested in computing $dW^{[l]}$ and $db^{[l]}$. For that we use g'^{L}, $dZ^{[l]}$, $Z^{[l]}$, and $W^{[l]}$.

8. A shallow neural network with a single hidden layer and 6 hidden units can compute any function that a neural network with 2 hidden layers and 6 hidden units can compute. True/False? **1 / 1 point**

- ○ True
- ◉ False

Correct
Correct. As seen during the lectures there are functions you can compute with a "small" L-layer deep neural network that shallower networks require exponentially more hidden units to compute.

9. Consider the following 2 hidden layers 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, 3)
- [x] $b^{[1]}$ will have shape (3, 1)

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

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

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

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

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

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

10. Whereas the previous question used a specific network, in the general case what is the dimension of $b^{[l]}$, the bias vector associated with layer l?

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

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

Correct
True. $b^{[l]}$ is a column vector with the same number of rows as units in the respective layer.