Your grade: 100%

Your latest: **100%** • Your highest: **100%** • 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**

- ○ 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.
- ◉ 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.

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. Among the following, which ones are "hyperparameters"? (Check all that apply.) **1 / 1 point**

- ☑ size of the hidden layers $n^{[l]}$

Correct

- ☐ weight matrices $W^{[l]}$
- ☐ bias vectors $b^{[l]}$
- ☐ activation values $a^{[l]}$
- ☑ learning rate α

Correct

- ☑ number of layers L in the neural network

Correct

- ☑ number of iterations

Correct

3. Which of the following statements is true? **1 / 1 point**

- ◉ 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

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**

- ◉ False
- ○ True

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**

◉

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

○

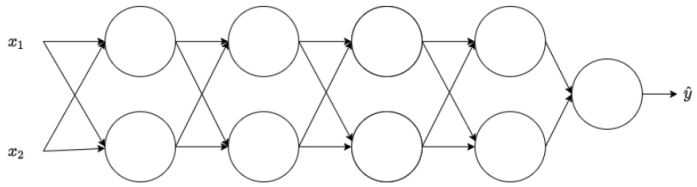
```
for i in range(L):
    Z[i+1] = W[i+1]*A[i+1] + b[i+1]
    A[i+1] = g(Z[i+1])
```

```
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])
```

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**



How many layers does this network have?

- The number of layers L is 4.
- The number of layers L is 2.
- The number of layers L is 6
- **The number of layers L is 5.**

Correct
Yes. The number of layers is the number of hidden layers + 1.

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? **1 / 1 point**

- False
- **True**

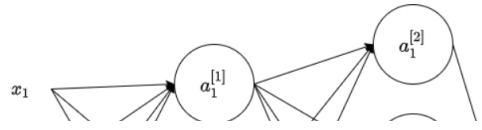
Correct
Yes, as you've seen in week 3 each activation has a different derivative. Thus, during backpropagation you need to know which activation was used in the forward propagation to be able to compute the correct derivative.

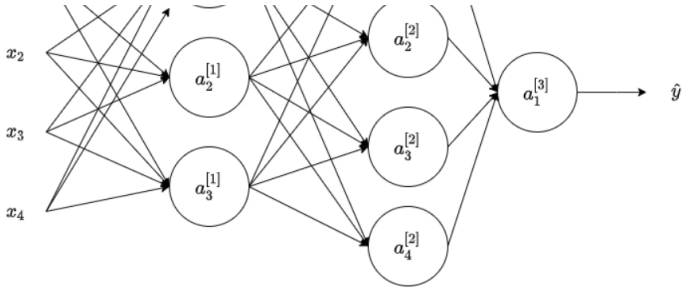
8. A shallow neural network with a single hidden layer and 6 hidden units can compute any function that a neural network with 2 hidden layers and 6 hidden units can compute. True/False? **1 / 1 point**

- 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).

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

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

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

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

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

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 $b^{[l]}$, the bias vector associated with layer l? **1 / 1 point**

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

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