What is the "cache" used for in implementations of forward propagation and backward propagation in our tutorials?

- a. 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.
- b. 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. ✔
- c. It is used to cache the intermediate values of the cost function during training.
- d. It is used to keep track of the hyperparameters that we are searching over, to speed up computation.

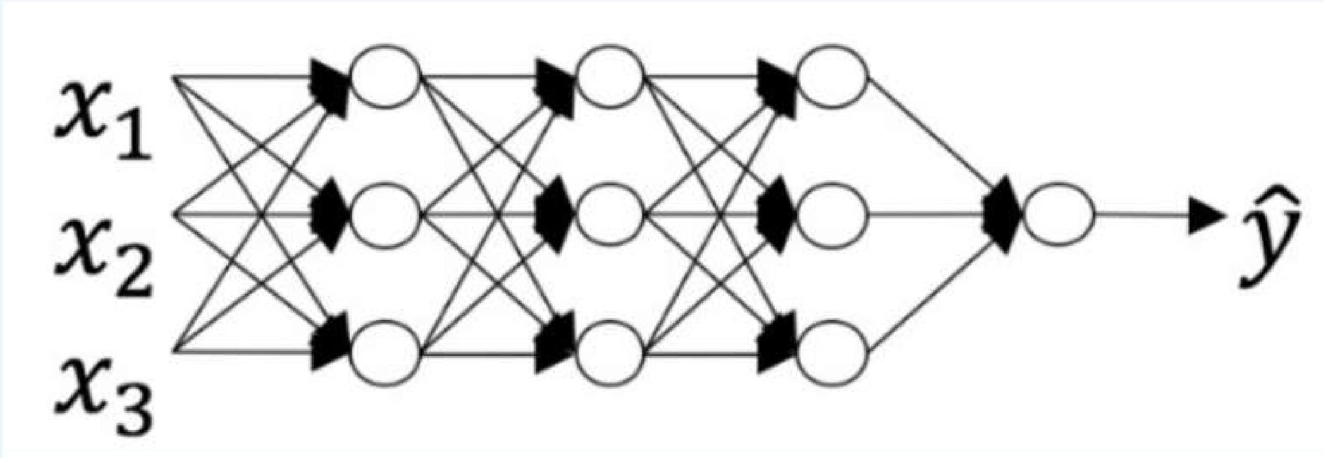
Which of the following are "hyperparameters"? (False answers lose points.)

- [ ] a. activation values
- [x] b. learning rate ✔
- [x] c. number of layers L in the neural network ✔
- [ ] d. bias vectors
- [ ] e. weight matrices
- [x] f. number of iterations ✔
- [x] g. size of the hidden layers ✔

Which of the following statements is true?

- a. The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.
- b. The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.

Consider the following neural network:



- a. The number of layers L is 4. The number of hidden layers is 3. ✔
- b. The number of layers L is 5. The number of hidden layers is 4.
- c. The number of layers L is 3. The number of hidden layers is 4.
- d. The number of layers L is 3. The number of hidden layers is 3.

Vectorization allows you to compute forward propagation in an L-layer neural network without an explicit for-loop (or any other explicit iterative loop) over the layers l=1, 2, ..., L. True/False?

Select one:

- True
- False ✔

What happens when you increase the strength of regularization on the network weights?

- a. Doubling the weighting of the regularization term should roughly result in doubling the weights
- b. Weights are pushed toward becoming smaller (closer to 0)
- c. Gradient descent taking bigger steps in each iteration (proportional to weighting of the regularization)
- d. Weights are pushed toward becoming greater (closer to 0)