

Leaky ReLU



Yes. Sigmoid outputs a value between 0 and 1 which makes it a very good choice res. Sigmolo outputs a value between u and I winto makes it a very good choice for binary classification. You can classify as 0 if the output is less than 0.5 and classify as 1 if the output is more than 0.5. It can be done with tanh as well but it is less convenient as the output is between -1 and 1.

tanh

~

8. You have built a network using the tanh activation for all the hidden units. You initialize the weights to relative large values, using np.random.randn(....)\*1000. What will happen?

1/1 point

- It doesn't matter. So long as you initialize the weights randomly gradient descent is not affected by whether the weights are large or small.
- This will cause the inputs of the tanh to also be very large, thus causing gradients to also become large. You therefore have to set  $\alpha$  to be very small to prevent divergence; this will slow down learning.
- This will cause the inputs of the tanh to also be very large, causing the units to be "highly activated" and thus speed up learning compared to if the weights had to start from small values.
- This will cause the inputs of the tanh to also be very large, thus causing gradients to be close to zero. The optimization algorithm will thus become slow.

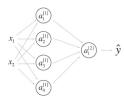
## Correct

Yes. tanh becomes flat for large values, this leads its gradient to be close to zero. This slows down the optimization algorithm.

~

9. Consider the following 1 hidden layer neural network:



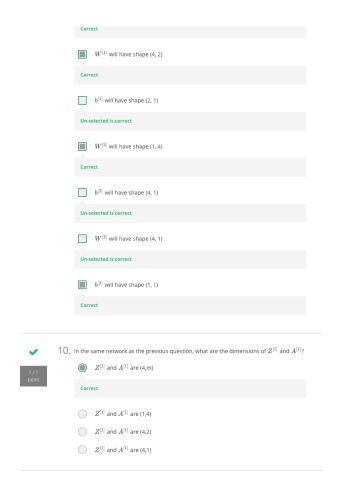


Which of the following statements are True? (Check all that apply).

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

Un-selected is correct

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



6 P P