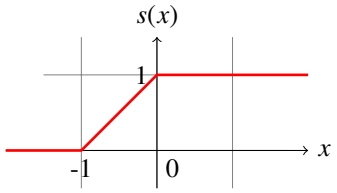
WEEKLY PARTICIPATION 6: THE EXPRESSIVITY OF RELU NEURAL NETWORKS

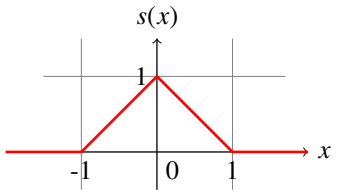
The key reason two hidden layer neural networks with nonlinear monotonic activation functions have the universal approximation property is that a small number of such neurons can be used to express a spike function, then we can approximate the function using a superposition of those spikes (see Lecture 16).

In this participation, you will verify that you can form a spike using a two-hidden layer ReLU feedforward neural network.

Let $\sigma(x) = x_+$ denote the ReLU activation function. Write the function s given below using a one hidden layer ReLU neural network containing 2 neurons.



Express the spike f function below in terms of s, and give a two-hidden layer ReLU neural network containing 5 neurons that computes f.



Note that this two-hidden layer ReLU neural network computed a piecewise linear function. In fact, *every* ReLU neural network, regardless of depth or width, computes a piecewise linear function. So another way to interpret the universal approximation property of ReLU networks is that one can approximate any nice function arbitarily well by using a piecewise linear function.