



Why

1

Definition

A sequence of functions (g_1, \dots, g_ℓ) is *composable* if g_i is composable with g_{i-1} for $i = 2, \dots, \ell$. In this case we write $g_\ell \circ g_{\ell-1} \circ \dots \circ g_2 \circ g_1$. For example, we write $g_3 \circ g_2 \circ g_1$ for (g_1, g_2, g_3) .

A *neural network* (or *feedforward neural network*) from \mathbf{R}^n to \mathbf{R}^m is a sequence of composable functions (g_1, \dots, g_ℓ) , $\text{dom } g_1 = \mathbf{R}^n$, $\text{ran } g_\ell \subset \mathbf{R}^m$, satisfying

$$g_i(\xi) = h_i(A_i \xi + b_i)$$

for some conforming matrices A_i , vectors b_i and functions h_i .

The *i*th *layer* of the neural network is the *i*th function g_i . The *i*th *activation* of the neural network is the function h_i . A *neural network* is called *deep* if its number of layers is larger than 3.

We call the composition of the layers of the neural network the *network predictor* (or just *predictor*). We also call it *the function* of the network.²

A *multi-layer perceptron* (*MLP*) is a neural network with 2 layers (1 *hidden layer*) and for which A_i and b_i have unrestricted nonzero entries.

¹Future editions will include. Future editions may change the name of this sheet to *computation networks*, or may add prerequisite sheet on computation graphs.

²Many authorities refer to a neural network as a function. Strictly speaking that is true for us, as well, since a sequence is a function. But the meaning of the common use is in reference to the *network predictor*.

