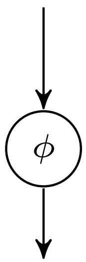
NEURAL NETWORKS

THE MOST IMPORTANT BIT

A neural network represents a function $f : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$.

Units

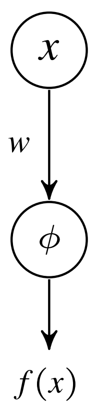
The basic building block is a **node** or **unit**:



- The unit has incoming and outgoing arrows. We think of each arrow as "transmitting" a signal.
- The signal is always a scalar.
- A unit represents a function ϕ.

We read the diagram as: A scalar value (say x) is transmitted to the unit, the function ϕ is applied, and the result $\phi(x)$ is transmitted from the unit along the outgoing arrow.

Weights

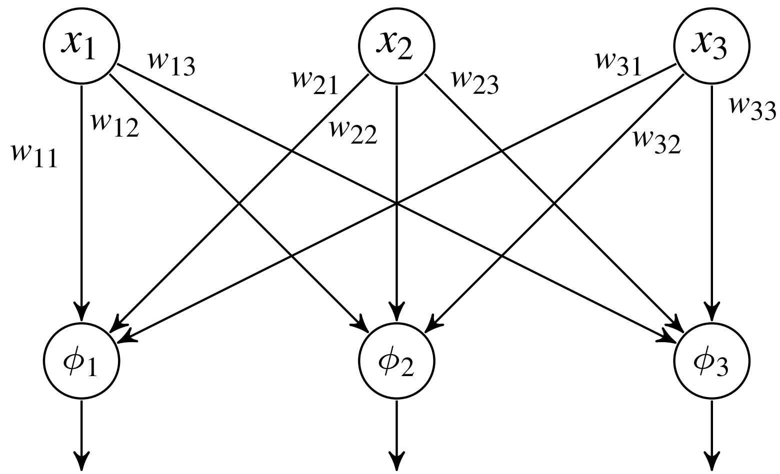


- If we want to "input" a scalar x, we represent it as a unit, too.
- We can think of this as the unit representing the constant function $g(x) = x$.
- Additionally, each arrow is usually inscribed with a (scalar) weight w.
- As the signal x passes along the edge, it is multiplied by the edge weight w.

The diagram above represents the function $f(x) := \phi(wx)$.

Reading Neural Networks

$$f : \mathbb{R}^3 \to \mathbb{R}^3 \quad \text{with input} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_2 \end{pmatrix}$$

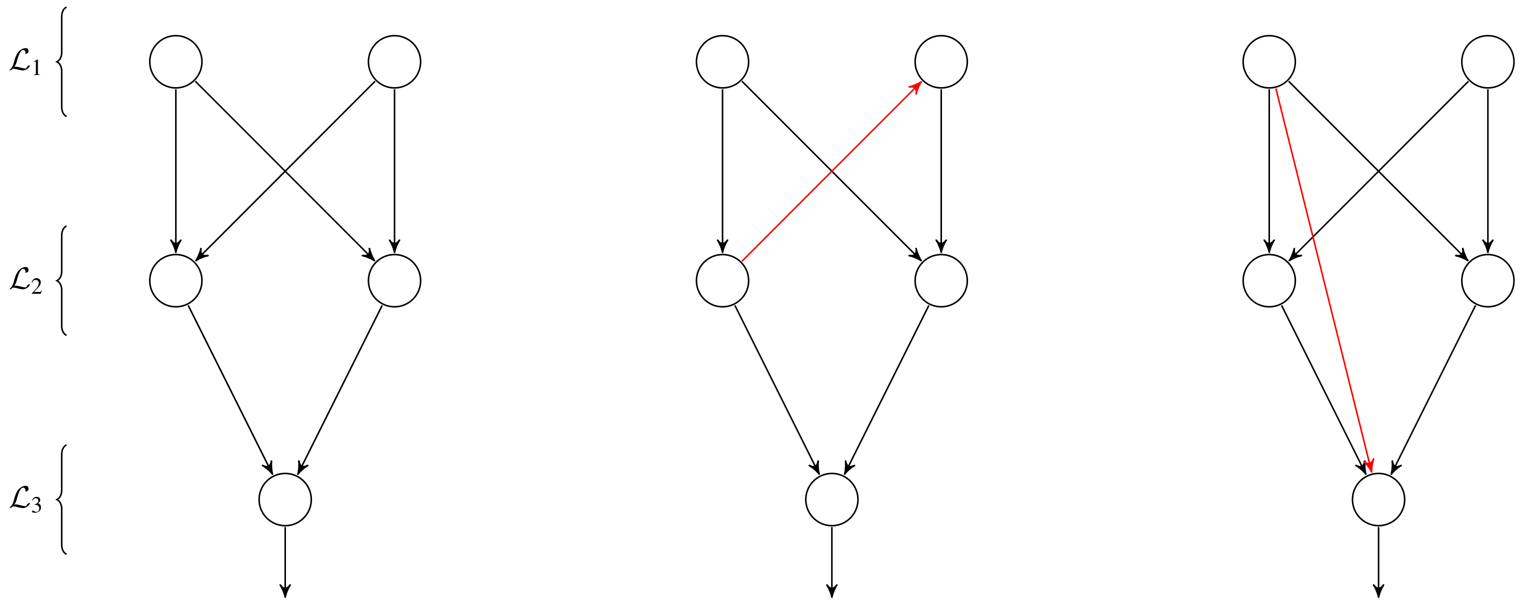


$f_1(\mathbf{x}) = \phi_1(\langle \mathbf{w}_1, \mathbf{x} \rangle)$ $\quad$ $f_2(\mathbf{x}) = \phi_2(\langle \mathbf{w}_2, \mathbf{x} \rangle)$ $\quad$ $f_3(\mathbf{x}) = \phi_3(\langle \mathbf{w}_3, \mathbf{x} \rangle)$

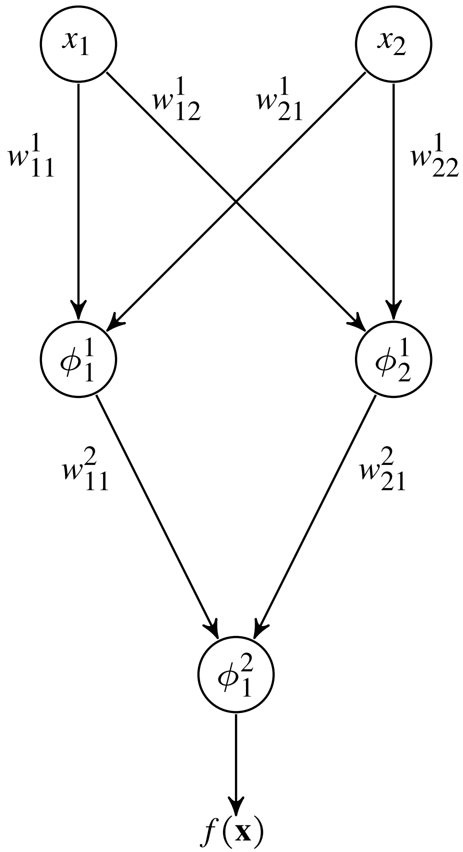
$$f(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ f_3(\mathbf{x}) \end{pmatrix} \quad \text{with} \quad f_i(\mathbf{x}) = \phi_i\Big(\sum_{j=1}^{3} w_{ij} x_j\Big)$$

A **feed-forward network** is a neural network whose units can be arranged into groups $\mathcal{L}_1, \ldots, \mathcal{L}_K$ so that connections (arrows) only pass from units in group $\mathcal{L}_k$ to units in group $\mathcal{L}_{k+1}$. The groups are called **layers**. In a feed-forward network:

- There are no connections within a layer.
- There are no backwards connections.
- There are no connections that skip layers, e.g. from $\mathcal{L}_k$ to units in group $\mathcal{L}_{k+2}$.



feed-forward | not feed-forward | not feed-forward



- This network computes the function

$$f(x_1,x_2) = \phi^2_1\Big(w^2_{11}\phi^1_1\big(w^1_{11}x_1+w^1_{21}x_2\big)+w^2_{12}\phi^1_2\big(w^1_{21}x_1+w^1_{22}x_2\big)\Big)$$

- Clearly, writing out f gets complicated fairly quickly as the network grows.

First shorthand: Scalar products

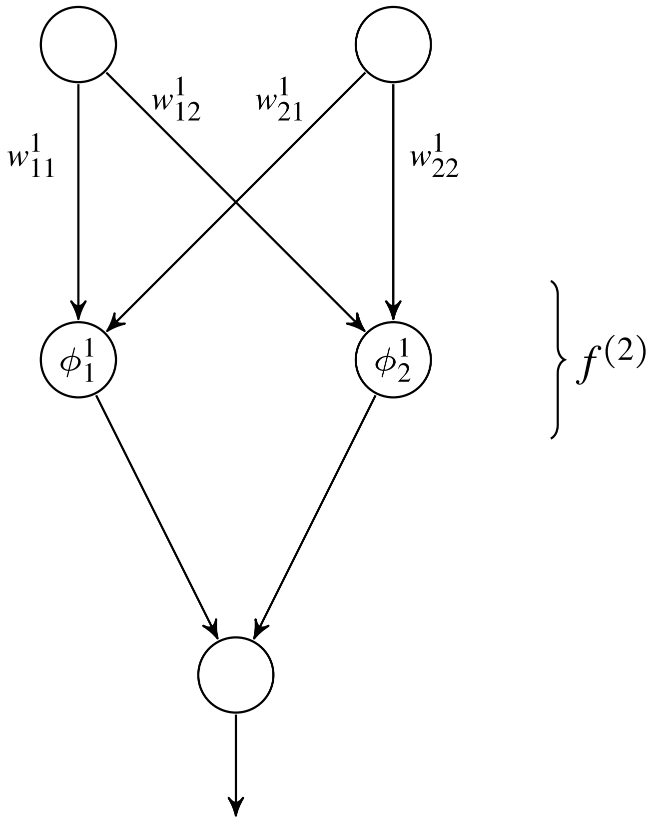
- Collect all weights coming into a unit into a vector, e.g.

$$\mathbf{w}^2_1 := (w^2_{11}, w^2_{21})$$

- Same for inputs: $\mathbf{x} = (x_1, x_2)$
- The function then becomes

$$f(\mathbf{x}) = \phi^2_1\left(\left\langle \mathbf{w}^2_1, \begin{pmatrix} \phi^1_1(\langle \mathbf{w}^1_1, \mathbf{x}\rangle) \\ \phi^1_2(\langle \mathbf{w}^1_2, \mathbf{x}\rangle) \end{pmatrix} \right\rangle\right)$$

LAYERS



- Each layer represents a function, which takes the output values of the previous layers as its arguments.
- Suppose the output values of the two nodes at the top are y_1, y_2.
- Then the second layer defines the (two-dimensional) function

$$f^{(2)}(\mathbf{y}) = \begin{pmatrix} \phi_1^1(\langle \mathbf{w}_1^1, \mathbf{y} \rangle) \\ \phi_2^1(\langle \mathbf{w}_2^1, \mathbf{y} \rangle) \end{pmatrix}$$

Composition of Functions

Basic composition

Suppose f and g are two function $\mathbb{R} \to \mathbb{R}$. Their **composition** $g \circ f$ is the function

$$g \circ f(x) := g(f(x)) \ .$$

For example:

$$f(x) = x + 1 \qquad g(y) = y^2 \qquad g \circ f(x) = (x+1)^2$$

We could combine the same functions the other way around:

$$f \circ g(x) = x^2 + 1$$

In multiple dimensions

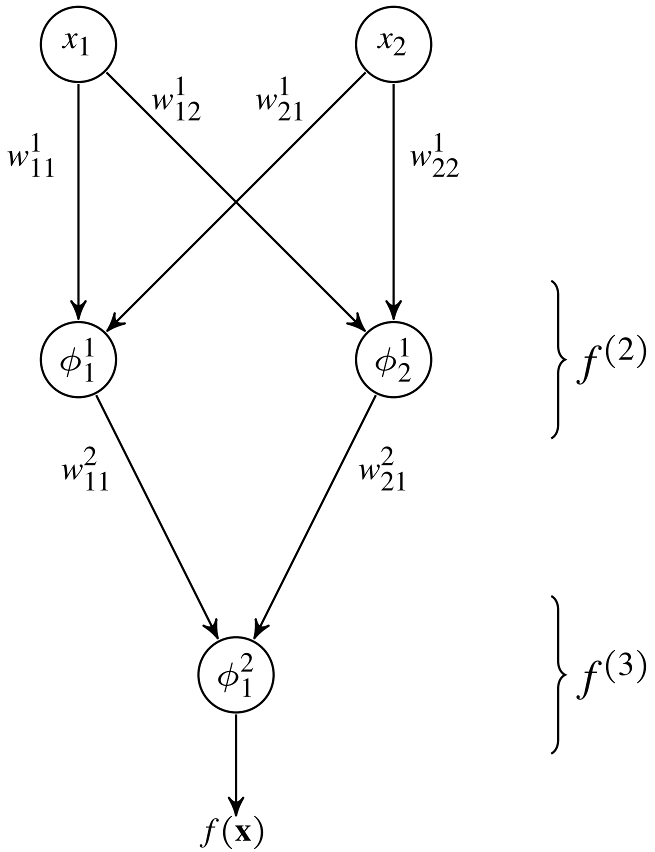
Suppose $f : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ and $g : \mathbb{R}^{d_2} \to \mathbb{R}^{d_3}$. Then

$$g \circ f(\mathbf{x}) = g(f(\mathbf{x})) \qquad \text{is a function } \mathbb{R}^{d_1} \to \mathbb{R}^{d_3} \ .$$

For example:

$$f(\mathbf{x}) = \langle \mathbf{x}, \mathbf{v} \rangle - c \qquad g(y) = \text{sgn}(y) \qquad g \circ f(\mathbf{x}) = \text{sgn}(\langle \mathbf{x}, \mathbf{v} \rangle - c)$$

LAYERS AND COMPOSITION



- As above, we write

$$f^{(2)}(\,\bullet\,) = \begin{pmatrix} \phi^1_1(\left\langle \mathbf{w}^1_1, \,\bullet\, \right\rangle) \\ \phi^1_2(\left\langle \mathbf{w}^1_2, \,\bullet\, \right\rangle) \end{pmatrix}$$

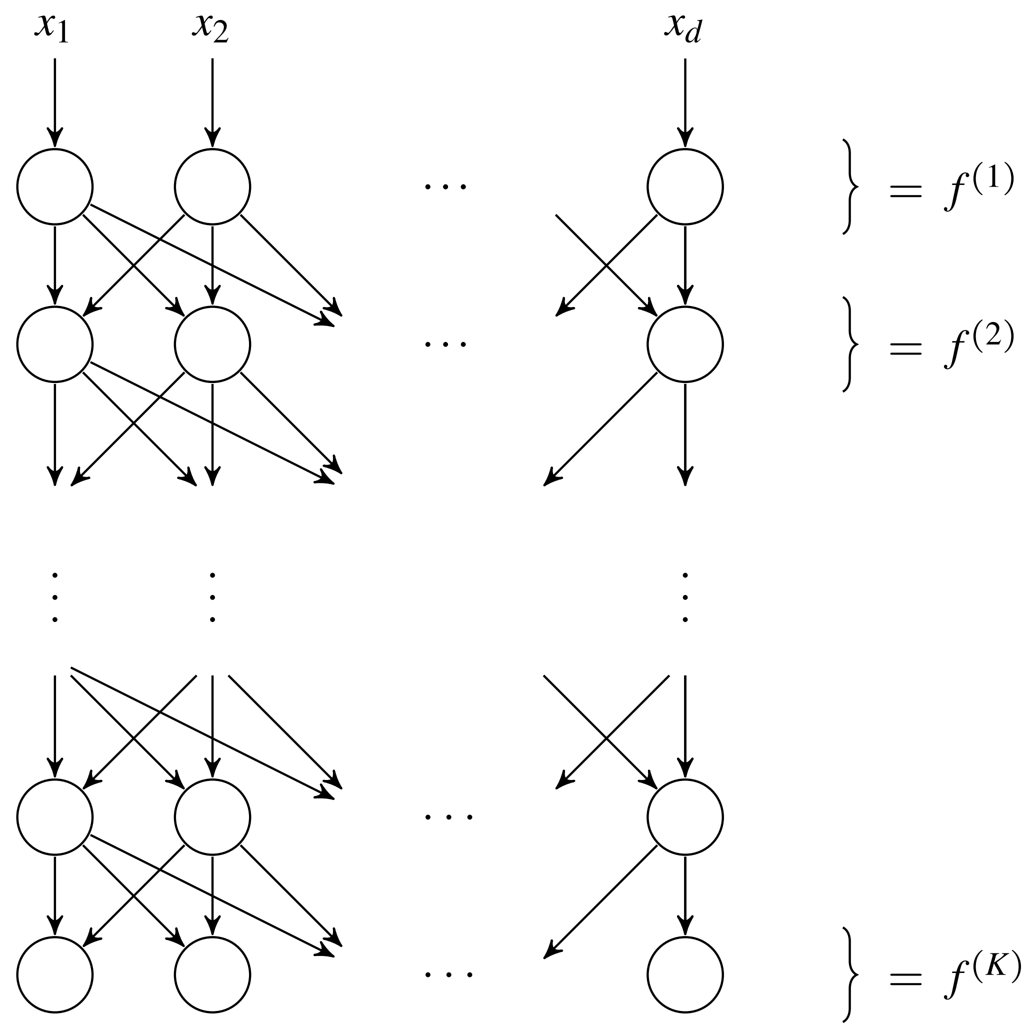
- The function for the third layer is similarly

$$f^{(3)}(\,\bullet\,) = \phi^2_1(\left\langle \mathbf{w}^2_1, \,\bullet\, \right\rangle)$$

- The entire network represents the function

$$f(\mathbf{x}) = f^{(3)}(f^{(2)}(\mathbf{x})) = f^{(3)} \circ f^{(2)}(\mathbf{x})$$

A feed-forward network represents a function as a composition of several functions, each given by one layer.



$$f(\mathbf{x}) \quad = \quad f^{(K)}(\cdots f^{(2)}(f^{(1)}(\mathbf{x}))) \quad = \quad f^{(K)} \circ \ldots \circ f^{(1)}(\mathbf{x})$$

General feed-forward networks

A feed-forward network with K layers represents a function

$$f(\mathbf{x}) \quad = \quad f^{(K)} \circ \ldots \circ f^{(1)}(\mathbf{x})$$

Each layer represents a function $f^{(k)}$. These functions are of the form:

$$f^{(k)}(\,\bullet\,) = \begin{pmatrix} \phi_1^{(k)}(\langle \mathbf{w}_1^{(k)}, \bullet \rangle) \\ \vdots \\ \phi_d^{(k)}(\langle \mathbf{w}_d^{(k)}, \bullet \rangle) \end{pmatrix} \qquad \text{typically:} \qquad \phi^{(k)}(x) = \begin{cases} \sigma(x) & \text{(sigmoid)} \\ \mathbb{I}\{\pm x > \tau\} & \text{(threshold)} \\ c & \text{(constant)} \\ x & \text{(linear)} \\ \max\{0, x\} & \text{(rectified linear)} \end{cases}$$

Dimensions

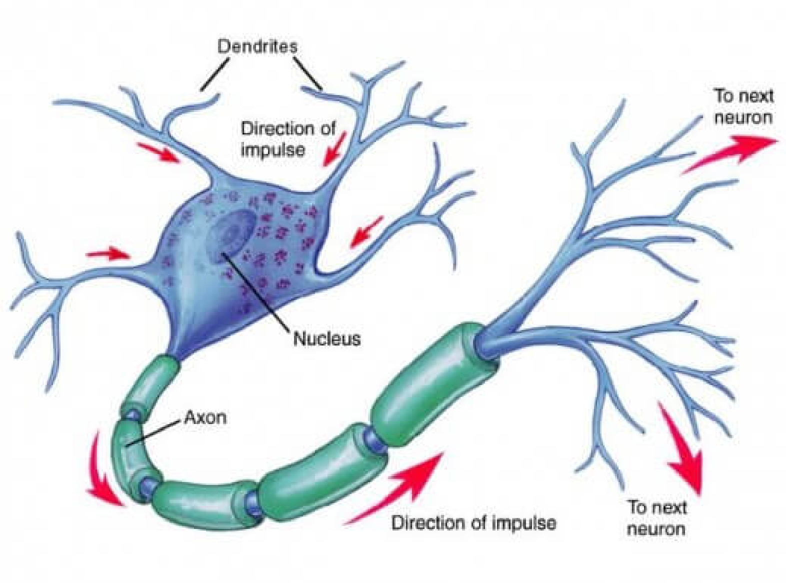
- Each function $f^{(k)}$ is of the form

$$f^{(k)} : \mathbb{R}^{d_k} \to \mathbb{R}^{d_{k+1}}$$

- d_k is the number of nodes in the kth layer. It is also called the *width* of the layer.
- We mostly assume for simplicity: $d_1 = \ldots = d_K =: d$.

ORIGIN OF THE NAME

If you look up the term “neuron” online, you will find illustrations like this:



This one comes from a web site called easyscienceforkids.com, which means it is likely to be scientifically more accurate than typical references to “neuron” and “neural” in machine learning.

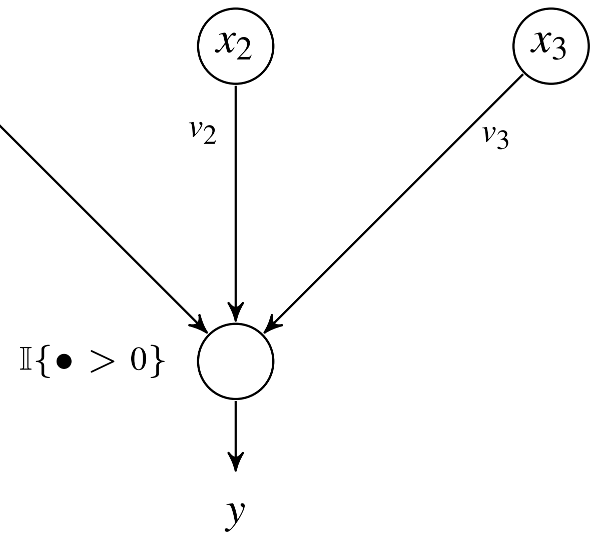
Roughly, a neuron is a brain cell that:

- Collects electrical signals (typically from other neurons)
- Processes them
- Generates an output signal

What happens inside a neuron is an intensely studied problem in neuroscience.

Source: easyscienceforkids.com

A neuron is modeled as a "thresholding device" that combines input signals:



McCulloch-Pitts neuron model (1943)

- Collect the input signals x_1, x_2, x_3 into a vector $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$
- Choose fixed vector $\mathbf{v} \in \mathbb{R}^3$ and constant $c \in \mathbb{R}$.
- Compute:
$$y = \mathbb{I}\{\langle \mathbf{v}, \mathbf{x} \rangle > 0\} \qquad \text{for some } c \in \mathbb{R} \; .$$
- In hindsight, this is a neural network with two layers, and function $\phi(\bullet) = \mathbb{I}\{\langle \mathbf{v}, \mathbf{x} \rangle > 0\}$ at the bottom unit.