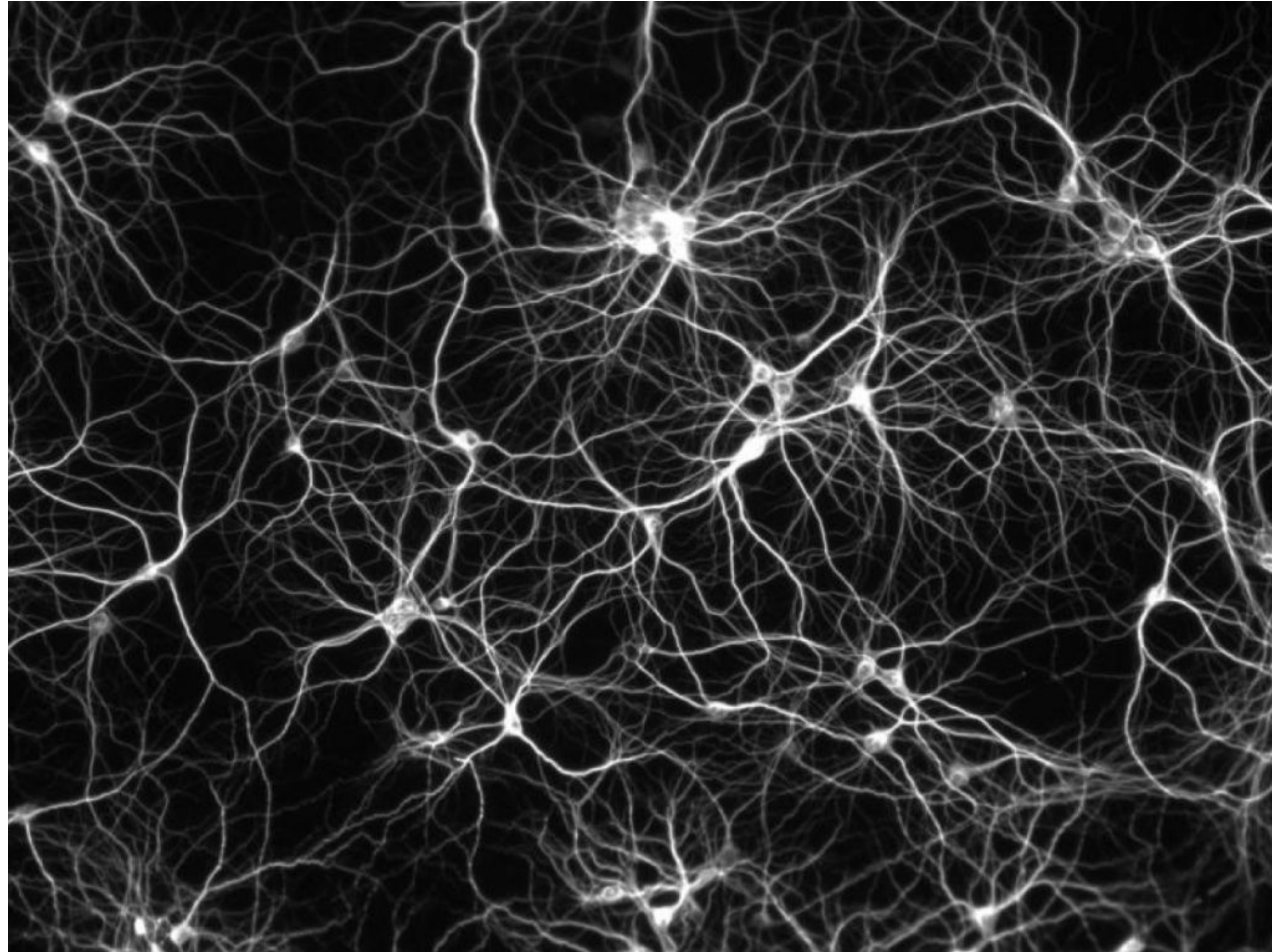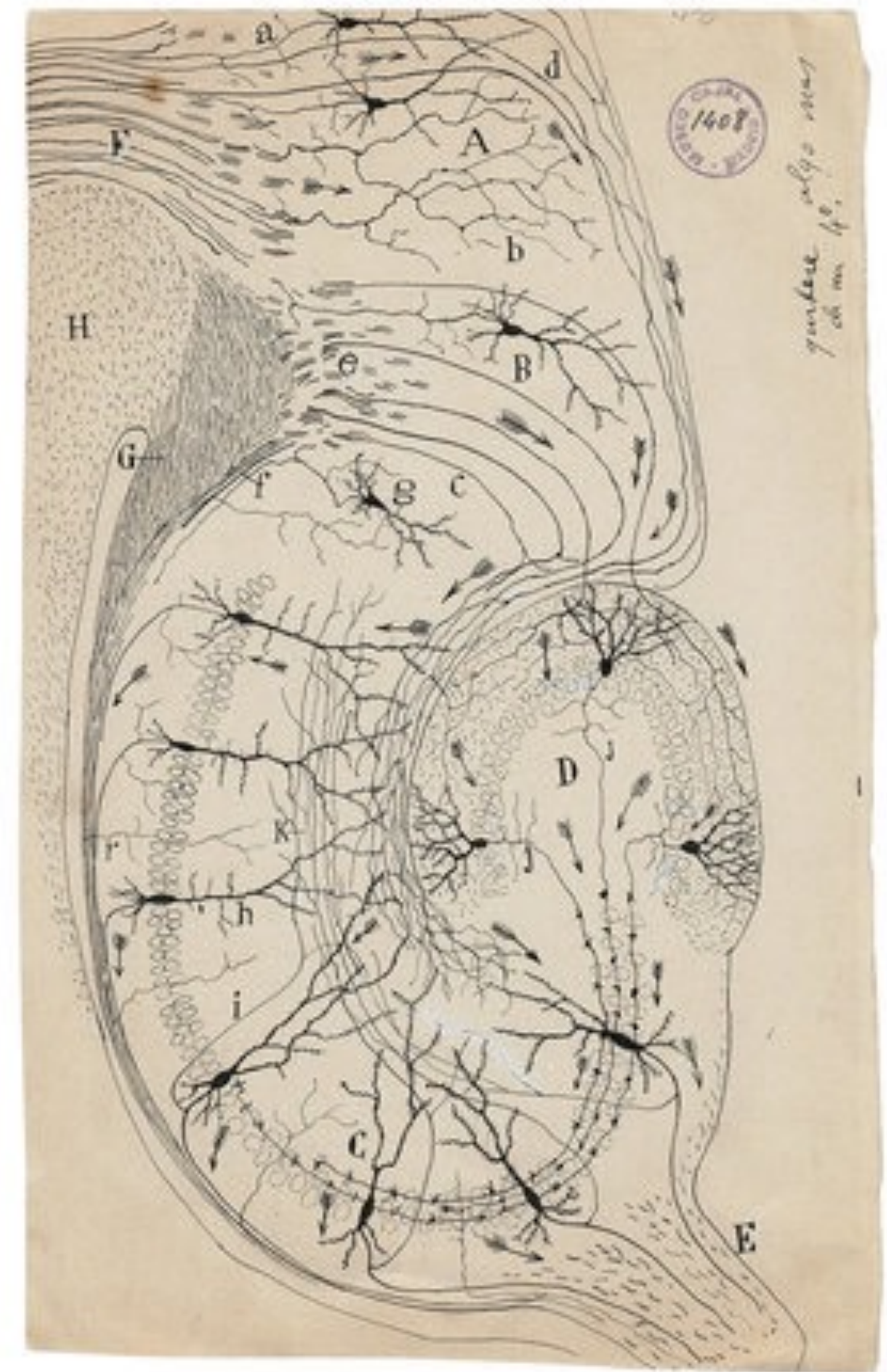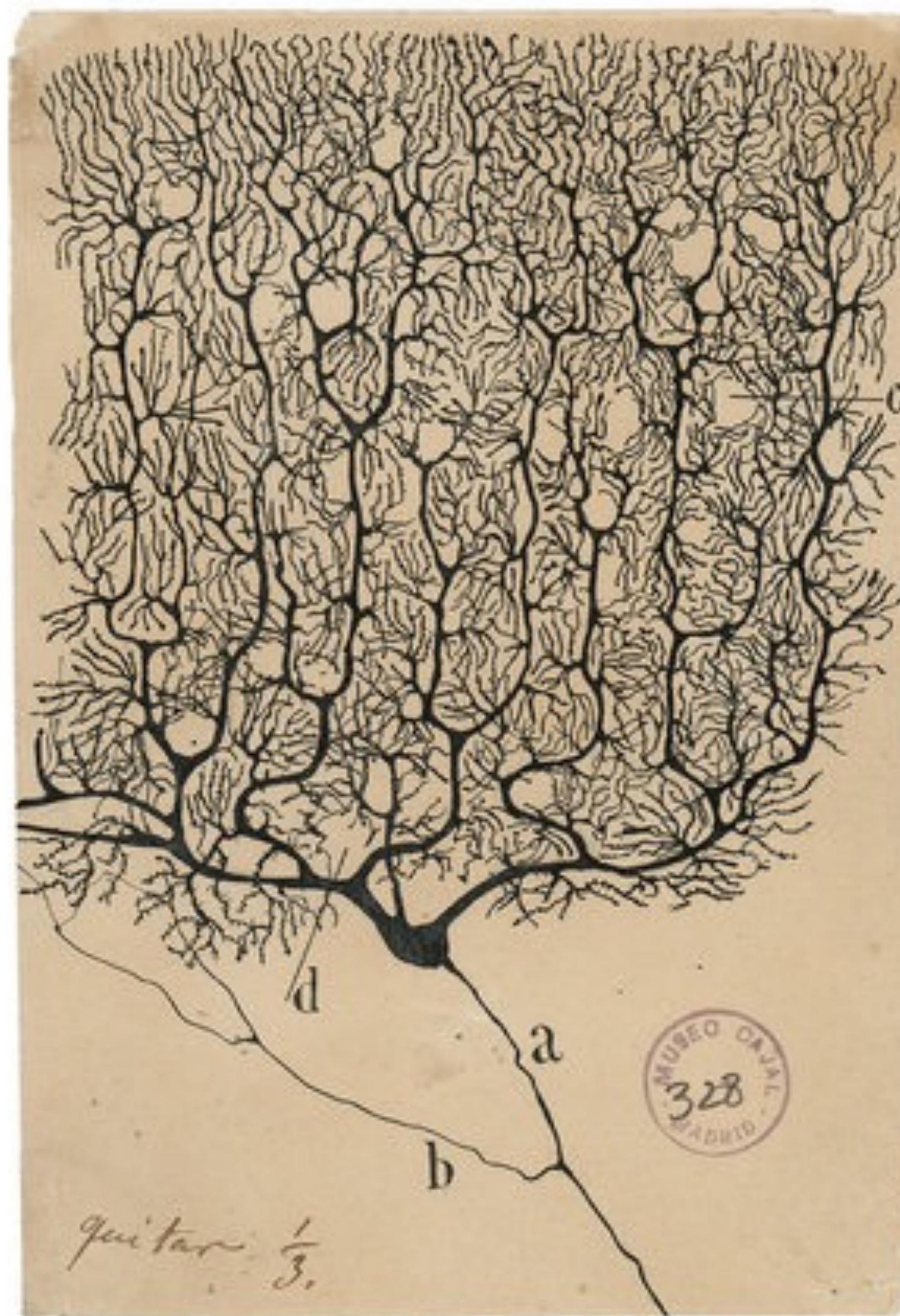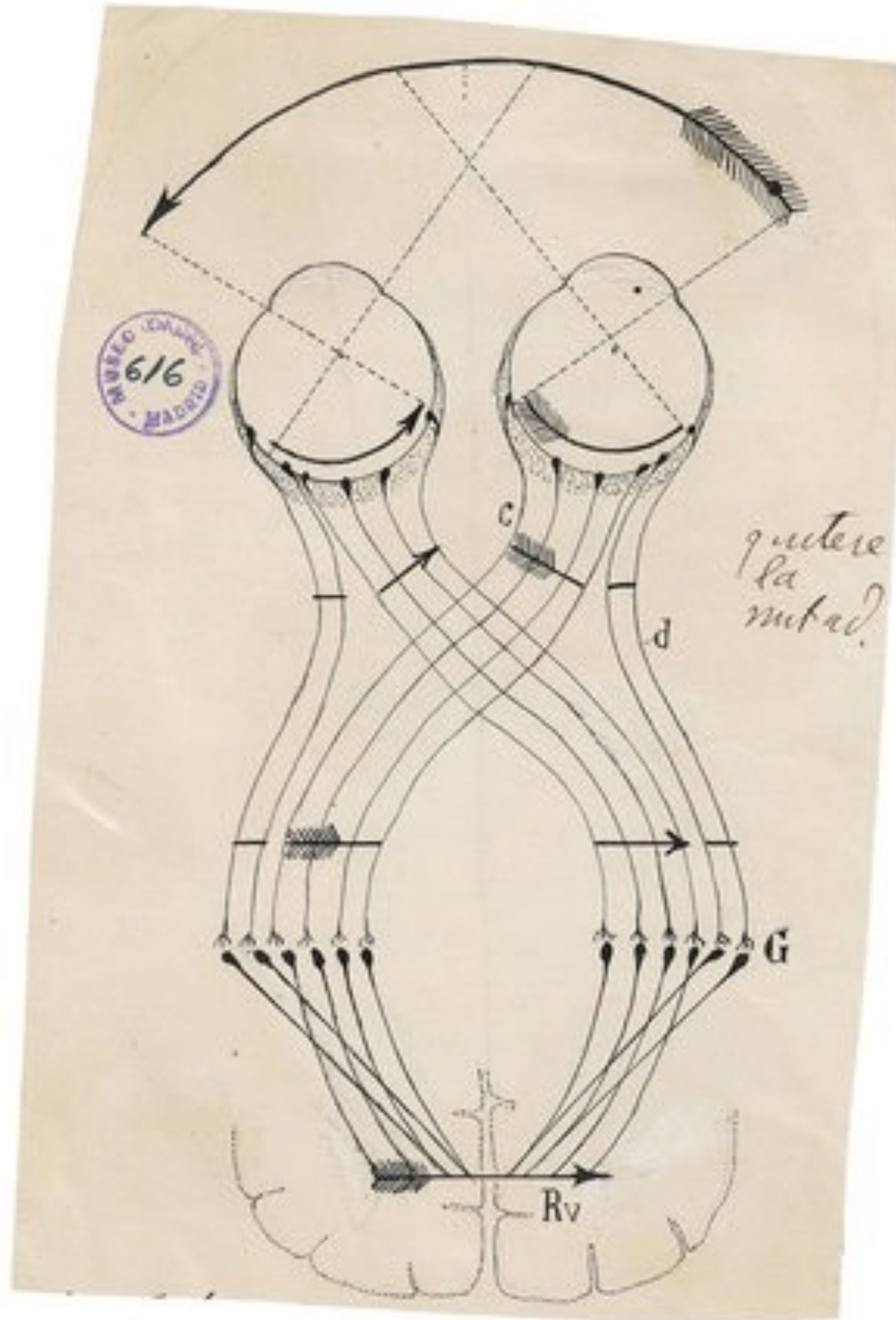# Course 8
Neural Networks
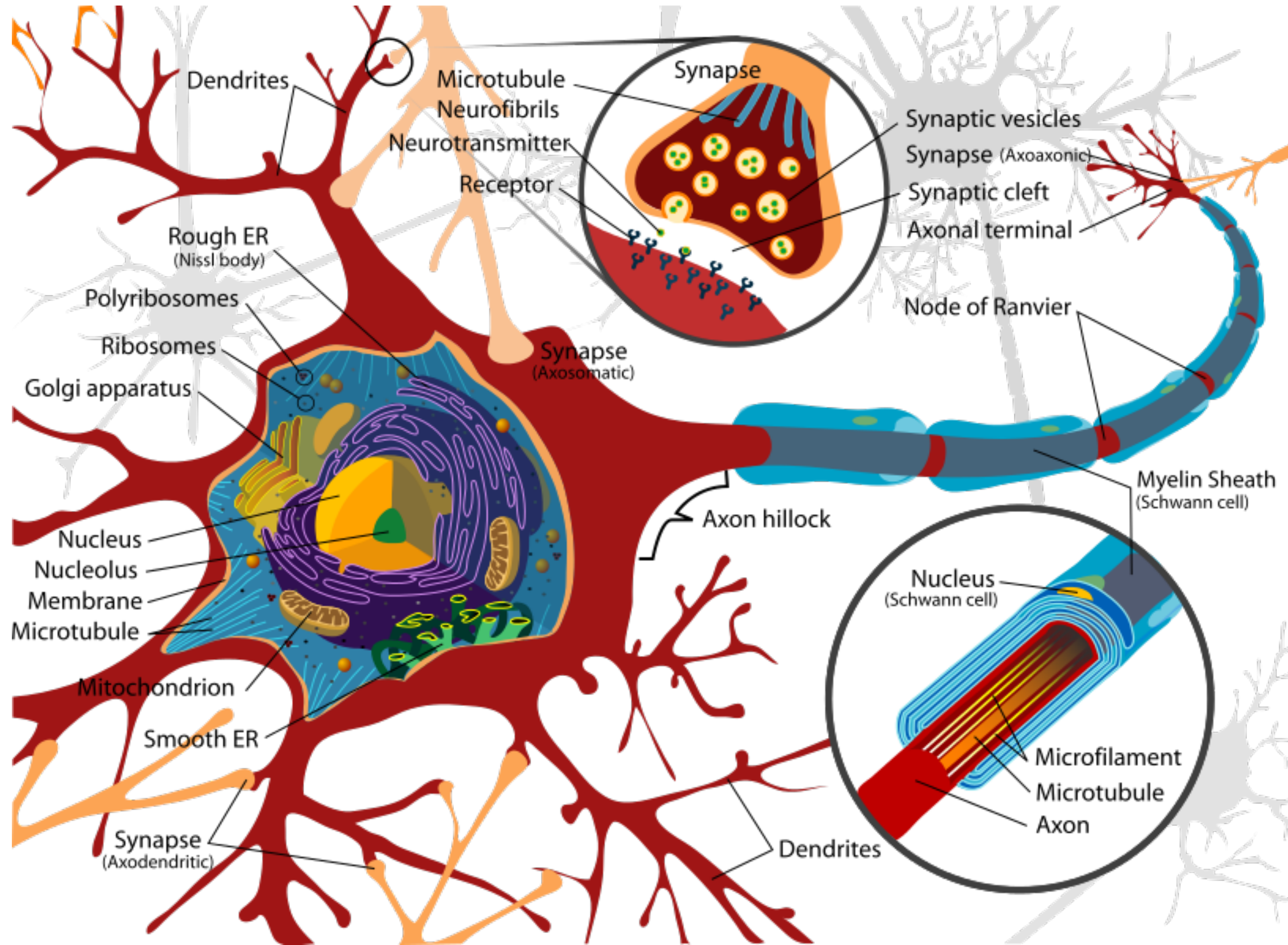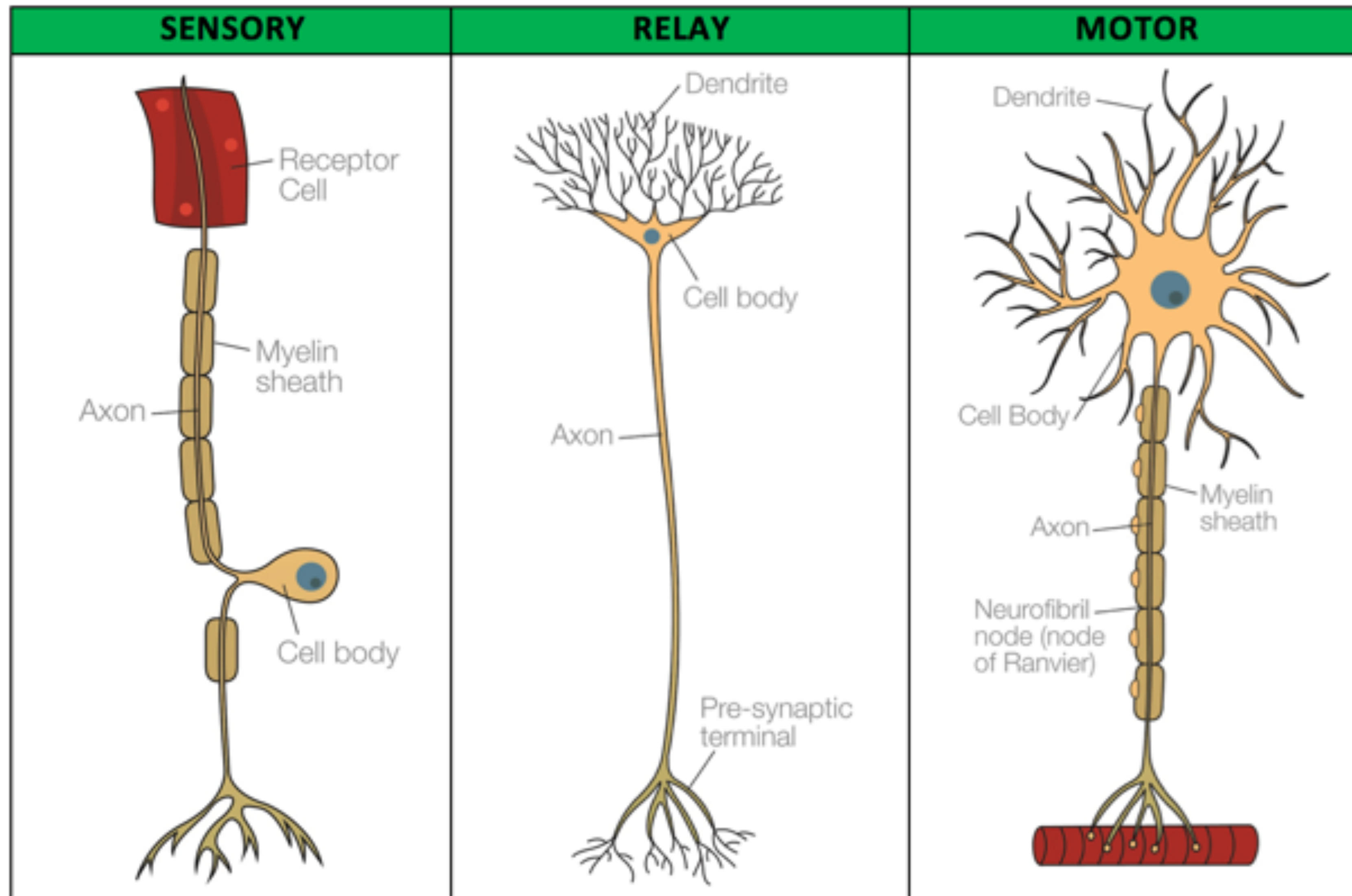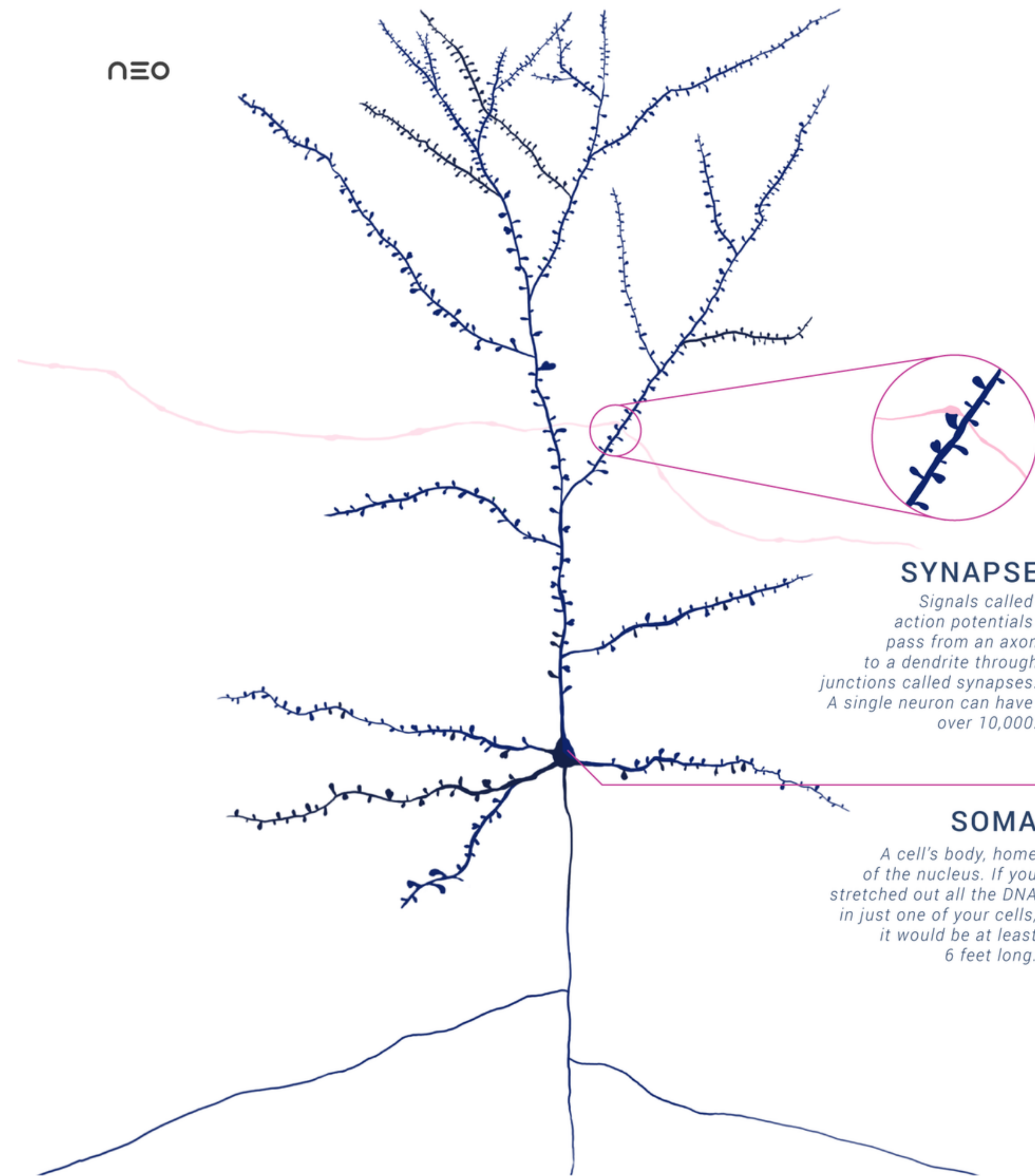
Christophe Eloy

Neural networks

Illustrations by Santiago Ramón y Cajal, the Spanish neuroscientist, from the book "The Beautiful Brain." From left: A diagram suggesting how the eyes might transmit a unified picture of the world to the brain; a purkinje neuron from the human cerebellum; and a diagram showing the flow of information through the hippocampus in the brain.

Neuron

Neurons

NEURON ANATOMY

SYNAPSE

Signals called action potentials pass from an axon to a dendrite through junctions called synapses. A single neuron can have over 10,000.

DENDRITES

Signals come in through dendrites. These vast, tree-like branches grow up and out from the soma. Dendrites are thicker than axons and covered in synapses.

SOMA

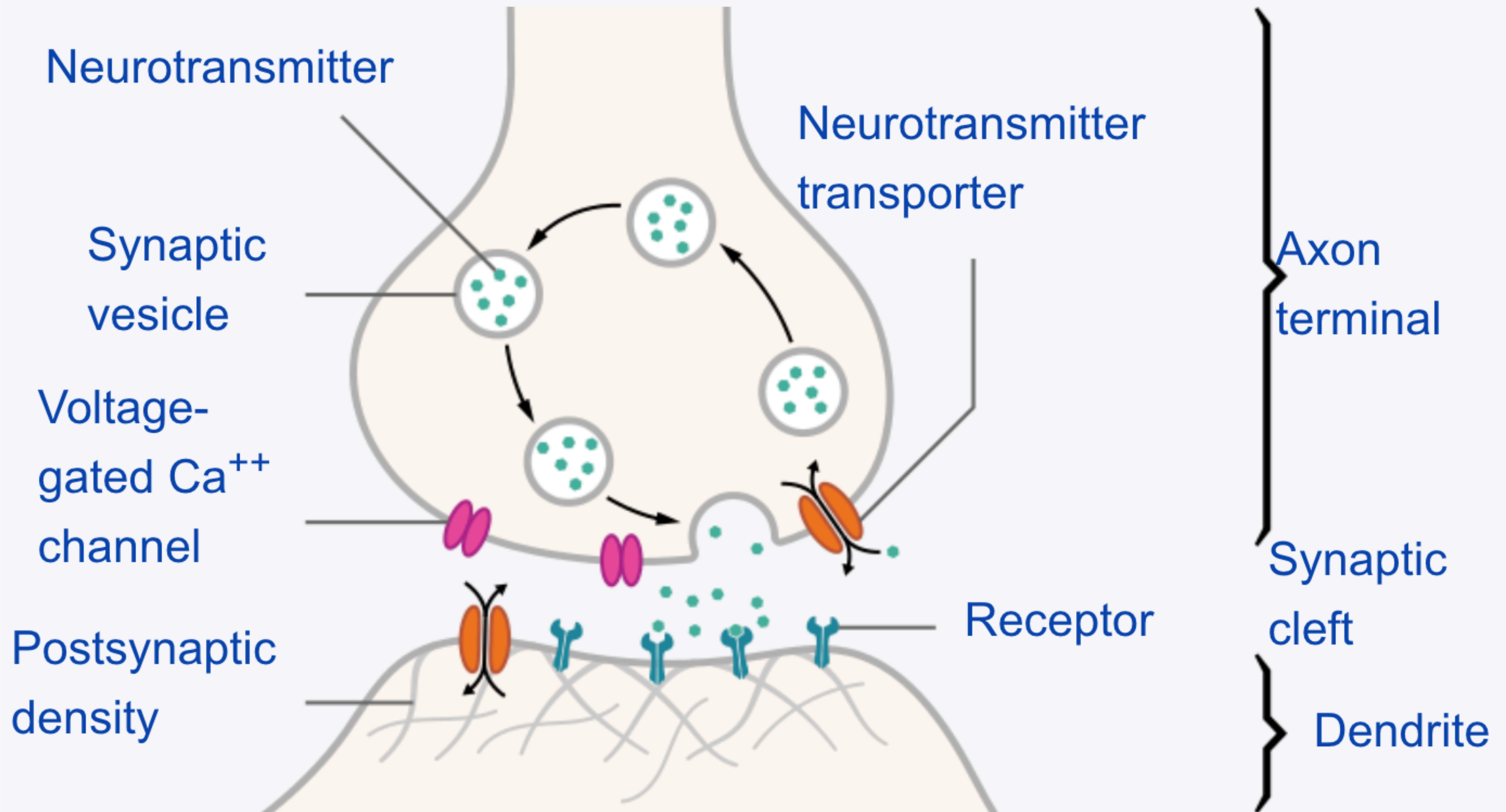A cell's body, home of the nucleus. If you stretched out all the DNA in just one of your cells, it would be at least 6 feet long.

AXON

Signals go out through axons, which branch many times and stretch vast distances. Neurons send action potentials down their axons and through synapses they've formed to communicate with other cells. The longest axons in your body reach from your toes to your spine.

Structure of a typical chemical synapse

Neurotransmitter

Synaptic vesicle

Voltage-gated Ca^{++} channel

Postsynaptic density

Neurotransmitter transporter

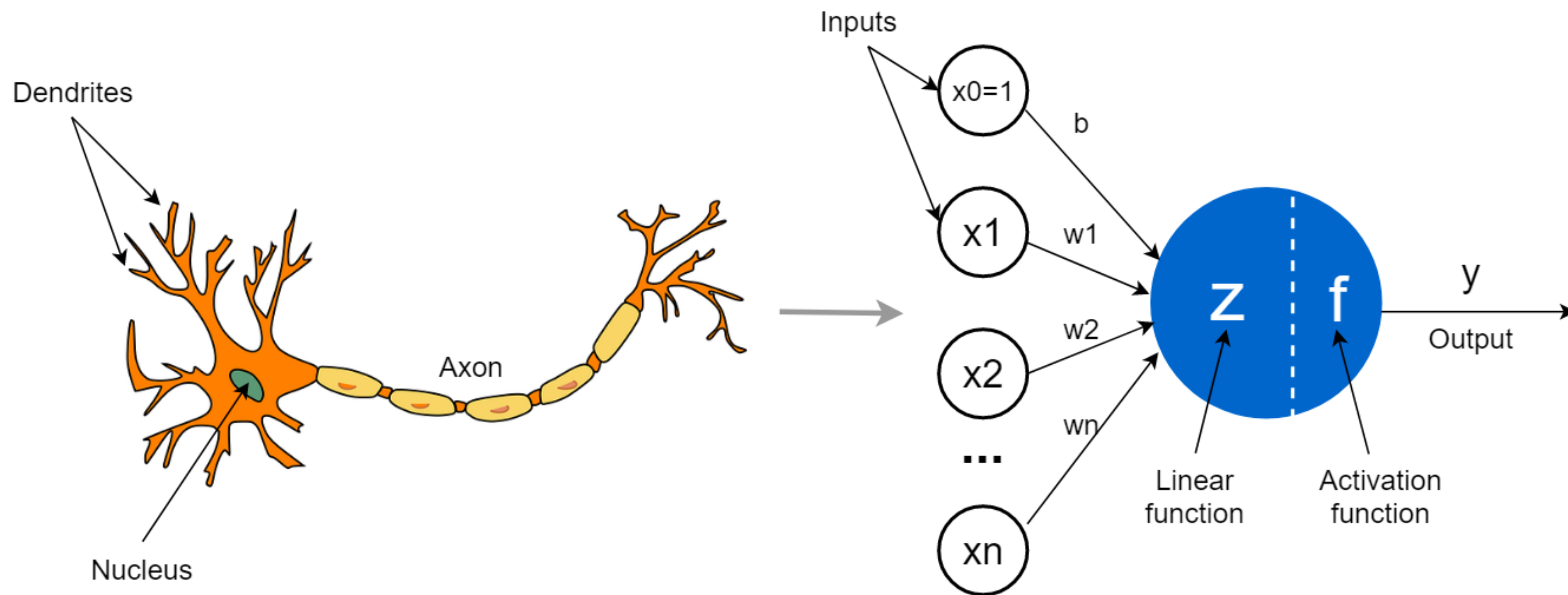Axon terminal

Synaptic cleft

Receptor

Dendrite

Human brain

- There are ~ 100 billion neurons

- Each neurons has an average of 7000 synaptic connections

- 10^{15} synapses

- 1.3 – 1.4 kg

Gigandet X, Hagmann P, Kurant M, Cammoun L, Meuli R, et al. *PLoS ONE* (2008)

Artificial neuron

Dendrites

Nucleus

Axon

Inputs

x0=1

x1

x2

...

xn

b

w1

w2

wn

z   f

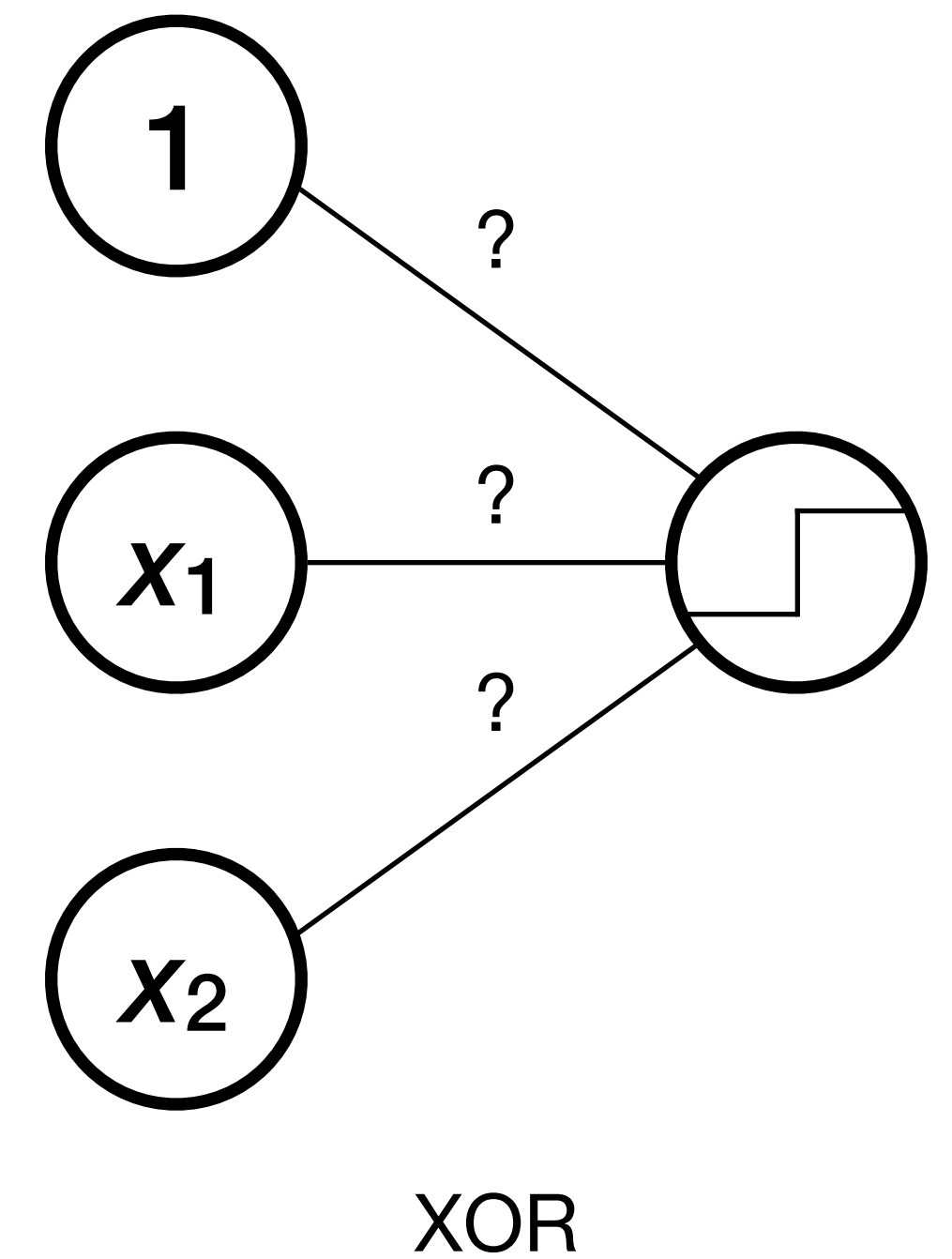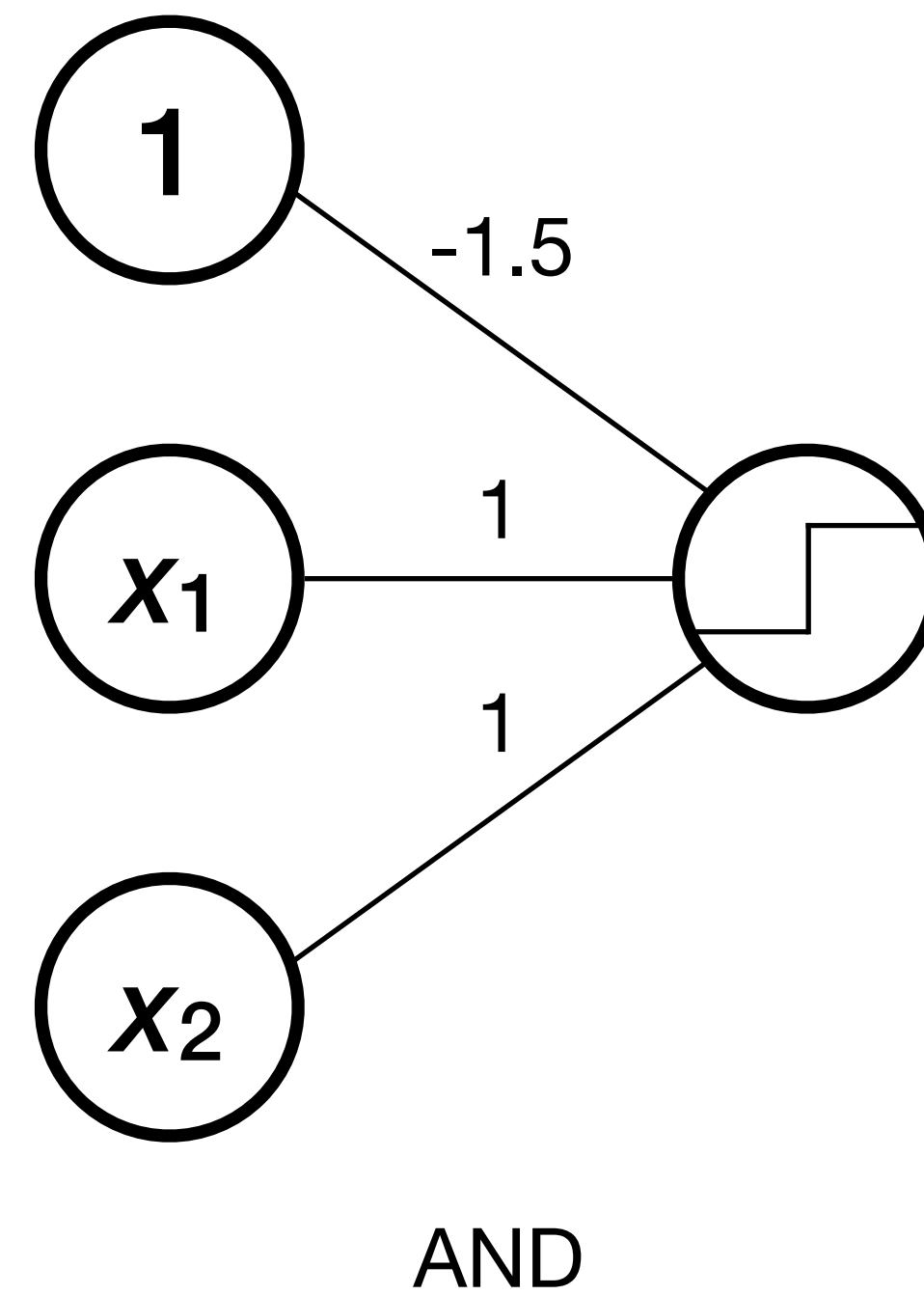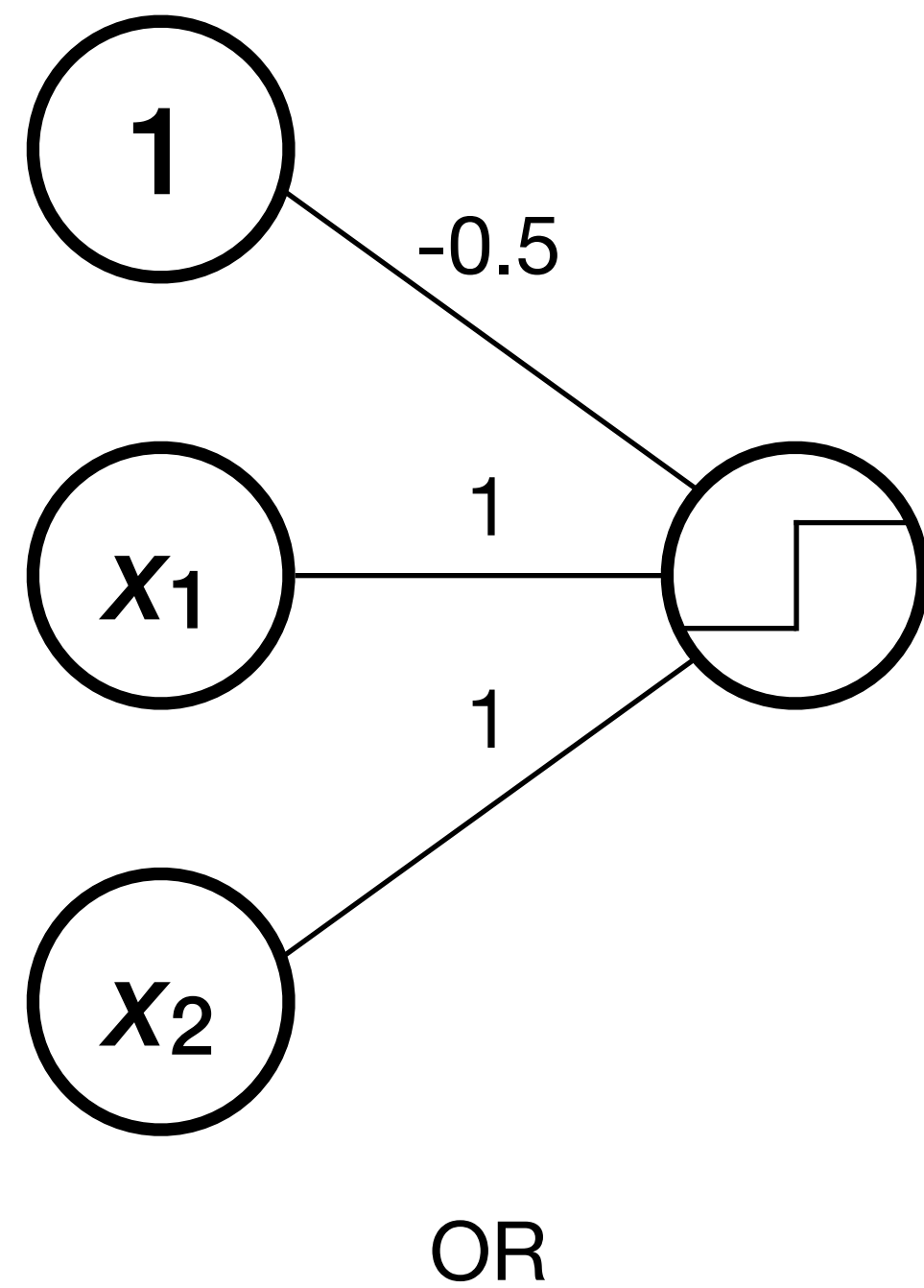Linear function

Activation function

y

Output

$$z = b + \sum_{j=1}^{N} w_j x_j$$

$$y = f(z)$$

Logical gates

OR   AND   XOR

Universal approximation theorem

- A feed-forward network with a single hidden layer containing a finite number of neurons can approximate **continuous** functions on compact subsets of $\mathbf{R}^n$, under mild assumptions on the activation function.

- Proved by George Cybenko in 1989 for logistic activation function

Nonlinearities

- Heaviside (step) function

- tanh

- logistic $\dfrac{1}{1 + e^{-x}}$

- ReLU (Rectified Linear Units) : max(0, x)

Feed forward

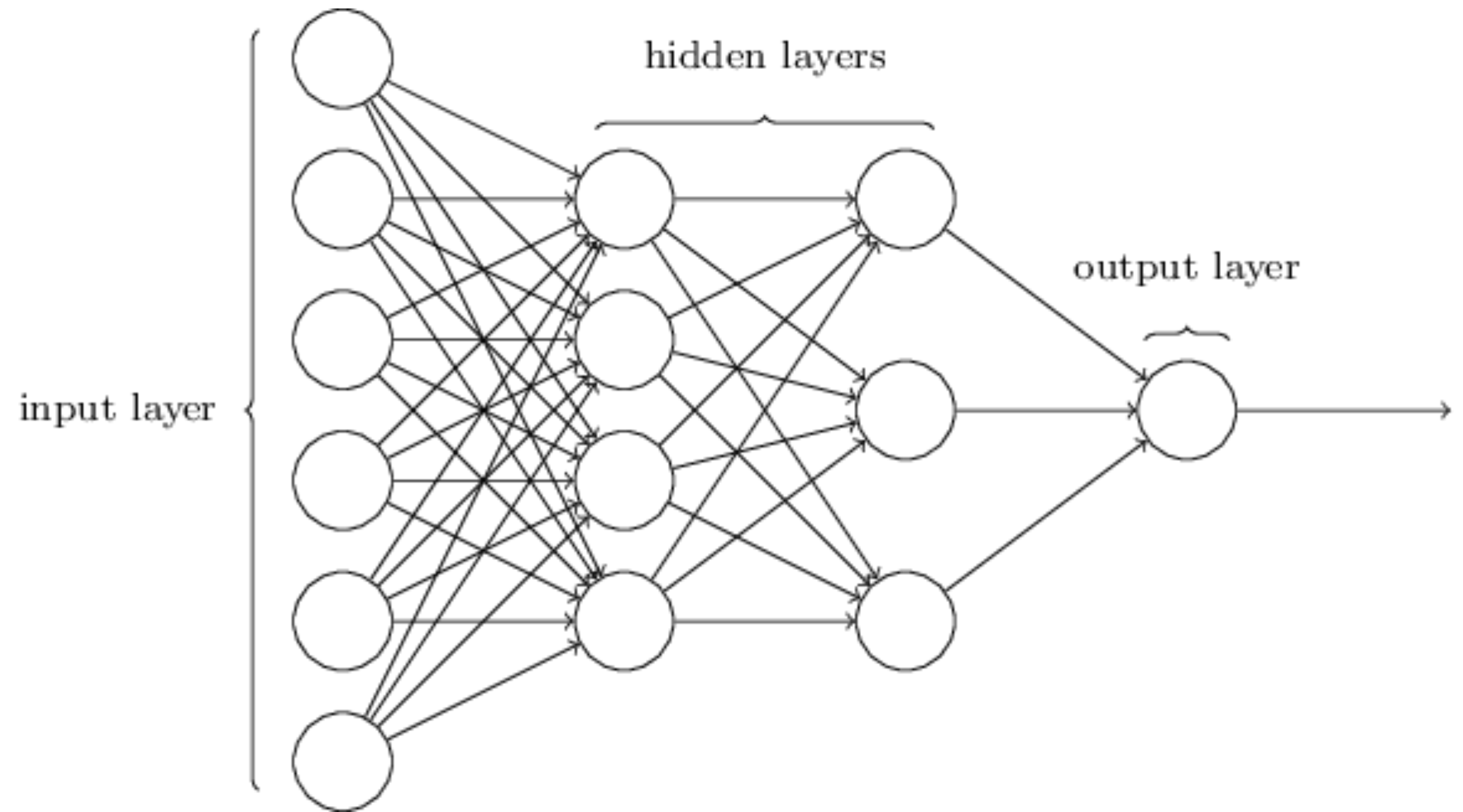$$o^{(1)} = x$$

$$p^{(2)} = \Theta^{(1)} o^{(1)}$$

$$o^{(2)} = g\left(p^{(2)}\right)$$

$$p^{(3)} = \Theta^{(2)} o^{(2)}$$

$$o^{(3)} = g\left(p^{(3)}\right)$$

$$p^{(4)} = \Theta^{(3)} o^{(3)}$$

$$o^{(4)} = g\left(p^{(4)}\right) = h_\theta(x)$$

hidden layers

input layer

output layer

Backpropagation

- Error: $E = \frac{1}{2} \left(h_\theta(x) - y \right)^2$

- Vectors δ's

$$\delta^{(L)} = \left(h_\theta - y \right) g' \left(p^{(L)} \right)$$

$$\delta^{(l)} = \left(\Theta^{(l)} \right)^T \delta^{(l+1)}.*g' \left(p^{(l)} \right)$$

- Gradient of error: $\dfrac{\partial E}{\partial \Theta_{ij}^{(l)}} = \delta_i^{(l+1)} o_j^{(l)}$