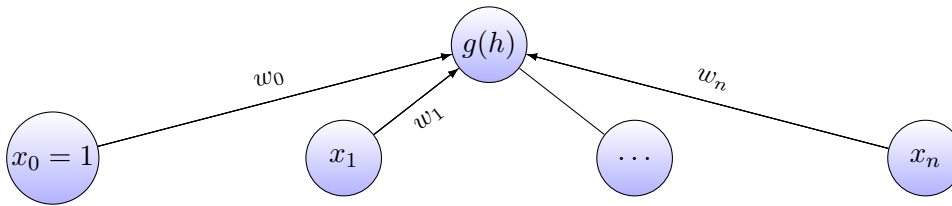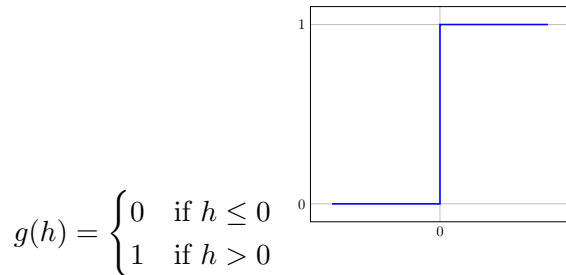# The evolution of the artificial (computational) neuron

$$h = w_0 x_0 + w_1 x_1 + \ldots, + w_n x_n, \quad O = g(h)$$
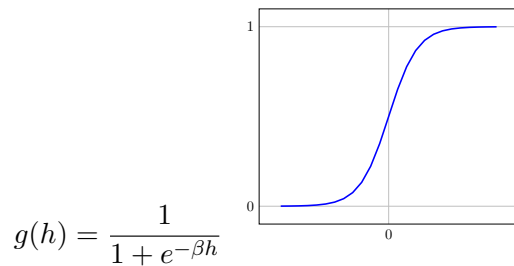

1943: McCulloch and Pitts. Pure threshold unit.

$$g(h) = \begin{cases} 0 & \text{if } h \leq 0 \\ 1 & \text{if } h > 0 \end{cases}$$

1962: Rosenblatt. Essentially the same as above.

1969: Minsky and Pappert. Essentially same as above.

1986: Rumelhart, Hinton, and Williams. Sigmoid. Multiple layers.

$$g(h) = \frac{1}{1 + e^{-\beta h}}$$

2010: Nair and Hinton. ReLU (Rectified Linear Unit). Deep learning.

$$g(h) = \max(0, h) = \begin{cases} 0 & \text{if } h \leq 0 \\ h & \text{if } h > 0 \end{cases}$$