UNIVERSIDADE CATÓLICA PORTUGUESA

BRAGA

Machine Learning

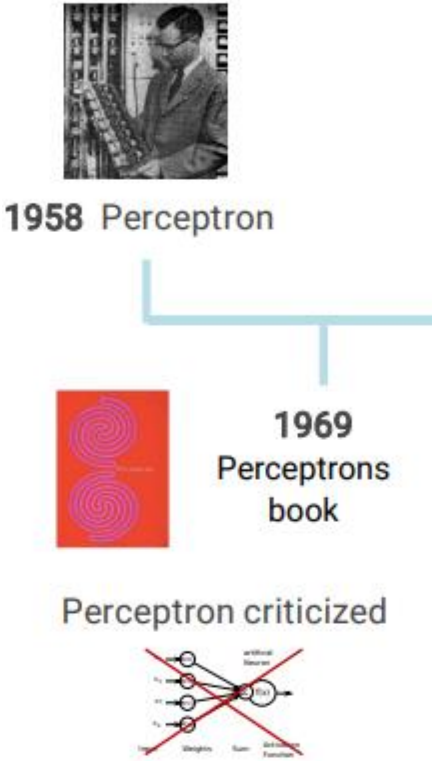
Session 20 - T

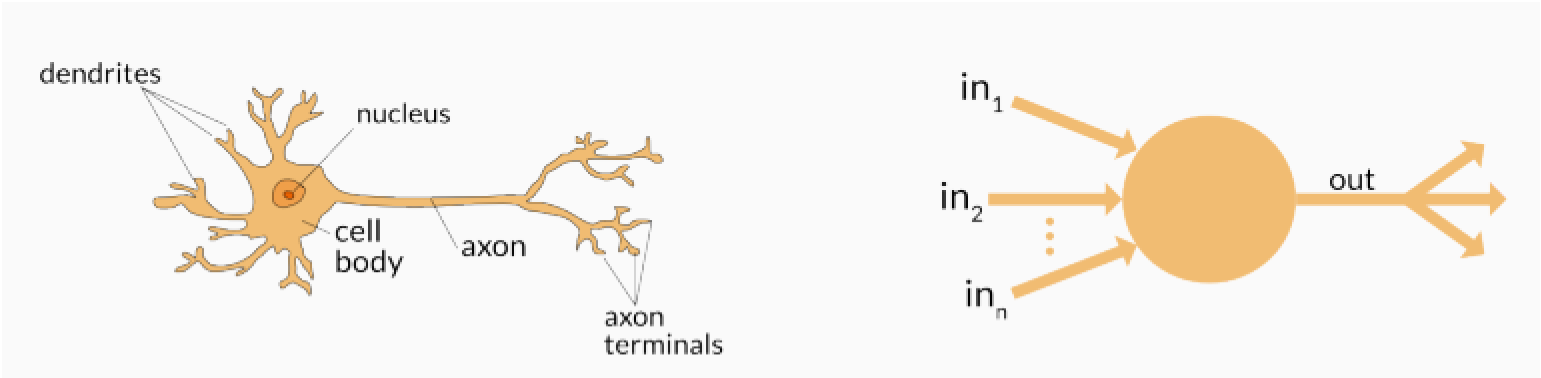
Neural Networks

Ciência de Dados Aplicada

2023/2024

Perceptron


1958 Perceptron
1969
Perceptrons
book
Perceptron criticized

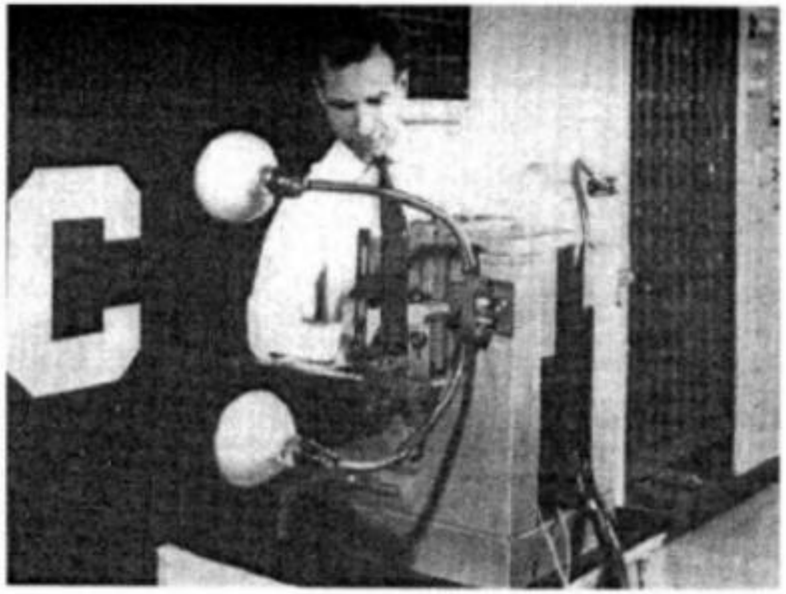

dendrites
nucleus
cell
body
axon
axon
terminals
in1
in2
inn
out

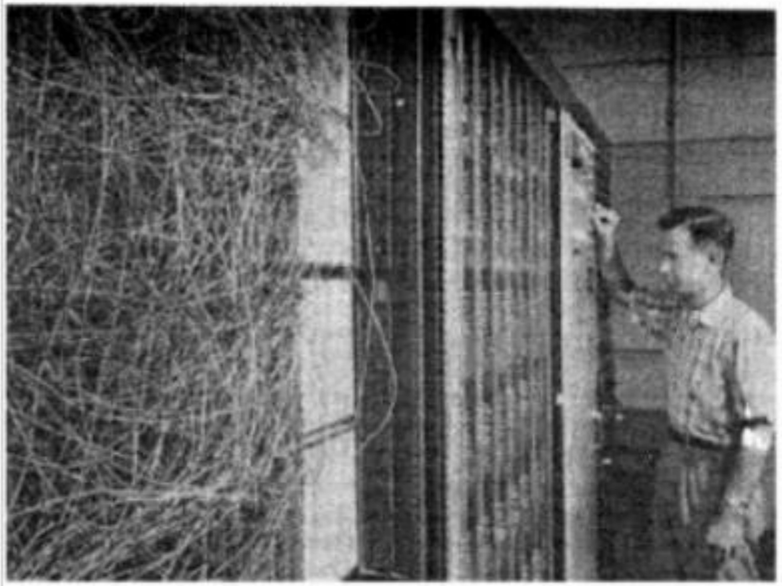
Perceptron

- First binary classifier based on supervised learning;

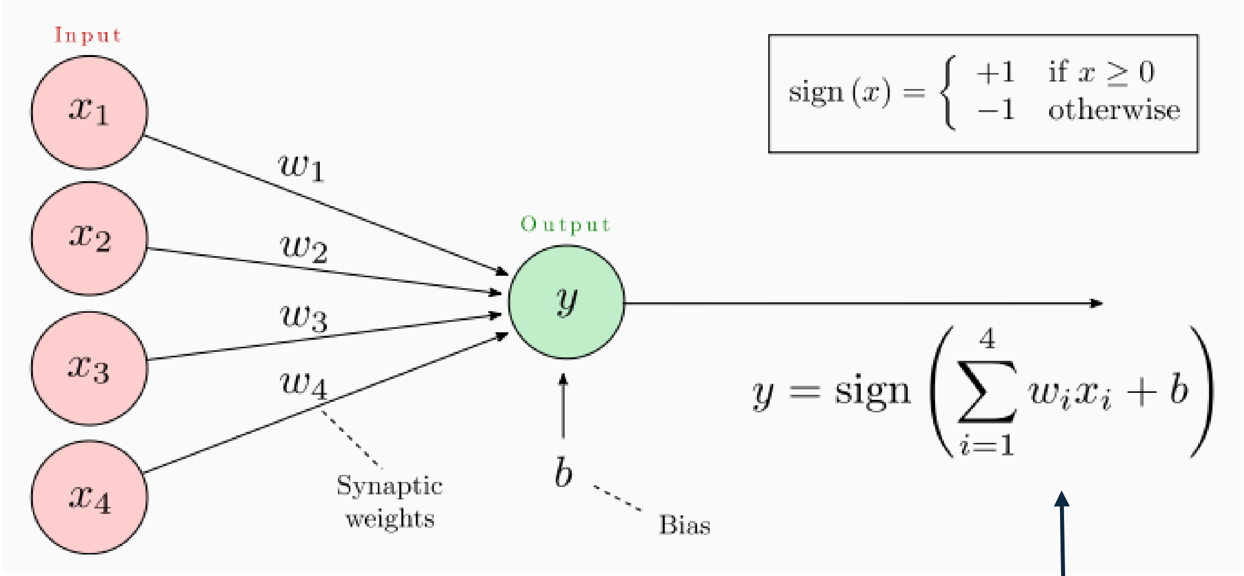
- Foundation of modern artificial neural networks;

Perceptron (Frank Rosenblatt, 1958)





Representation of the Perceptron

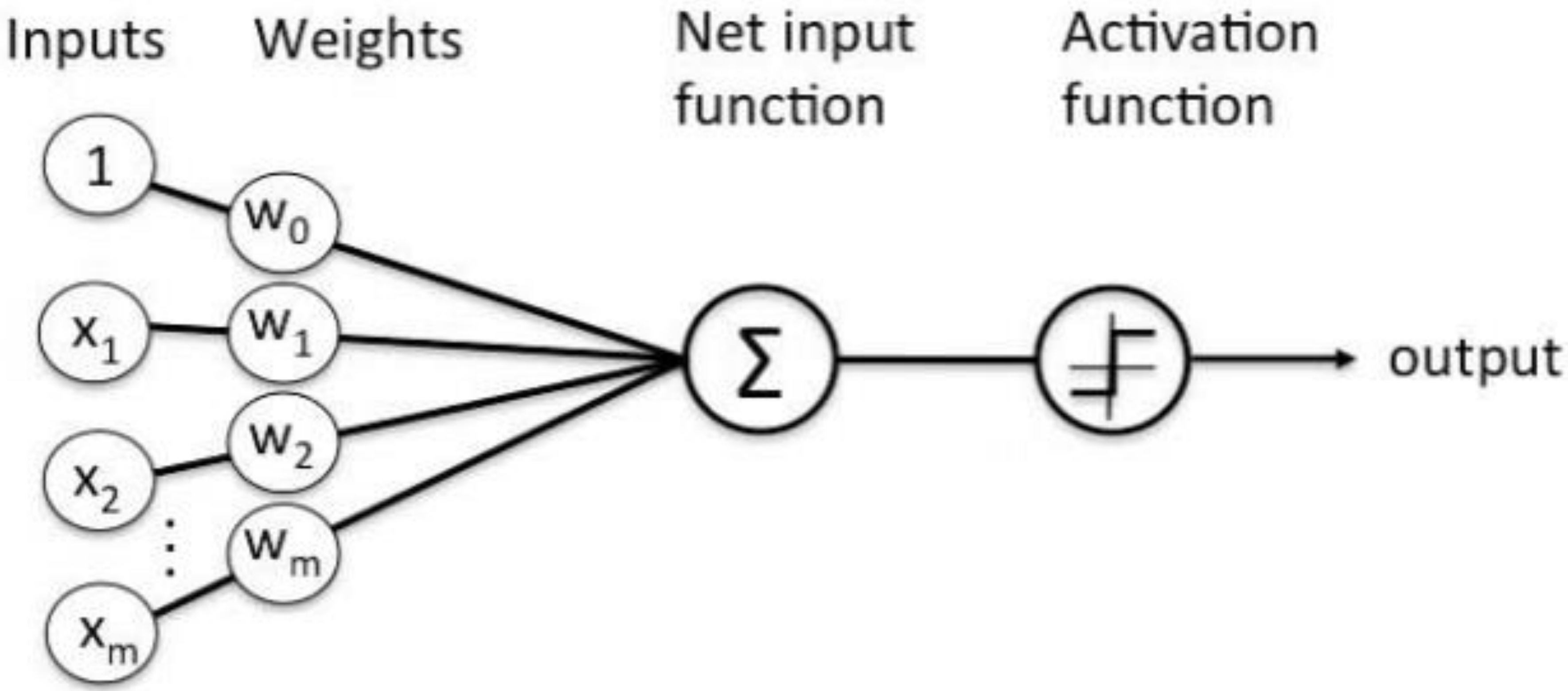


Activation function

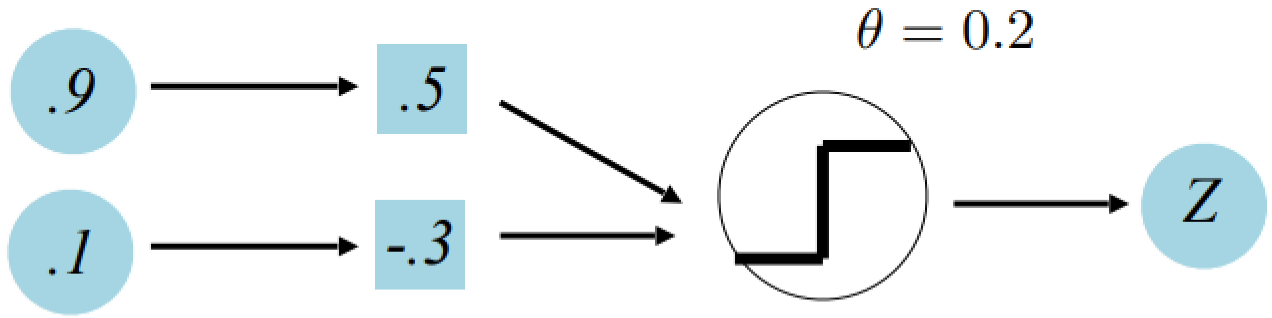
- Parameters of the perceptron:
  - w_k: weights
  - b: bias

- Training → adjusting the weights and bias.

Alternative Representation of the Perceptron

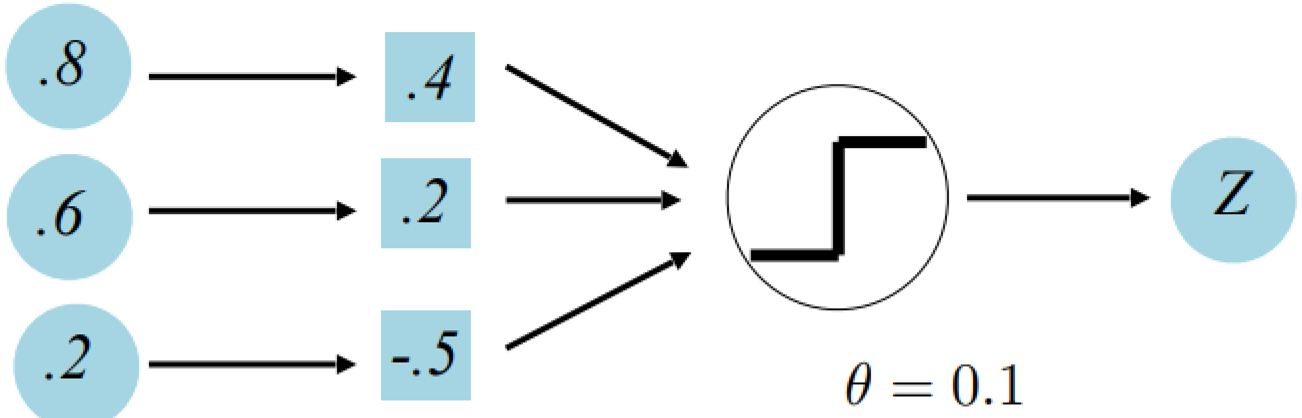


Perceptron Examples



$z = ?$

$$z = \begin{cases} 1, if \ w \cdot x > \theta \\ 0, \ w \cdot x <= \theta \end{cases}$$



$z = ?$

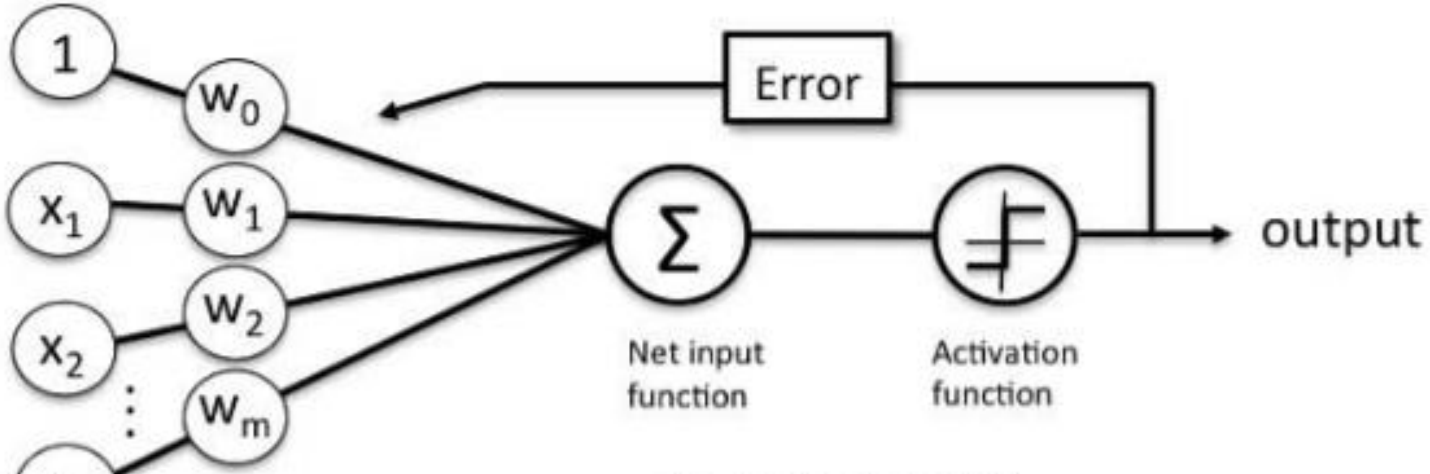
Perceptron Learning Rule

- Suppose that x is a feature vector, y is the correct class label, and y' is the predicted class label computed using the current weights.
  - If y' = y, do nothing
  - Otherwise $w_i^{new} = w_i^{old} + \Delta w_i^{old}$, $\Delta w_i = \alpha\,(t - o)x_i$

perceptron prediction

learning rate

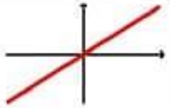
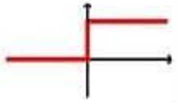
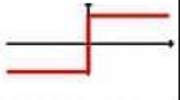
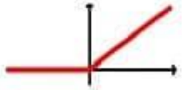
real target value



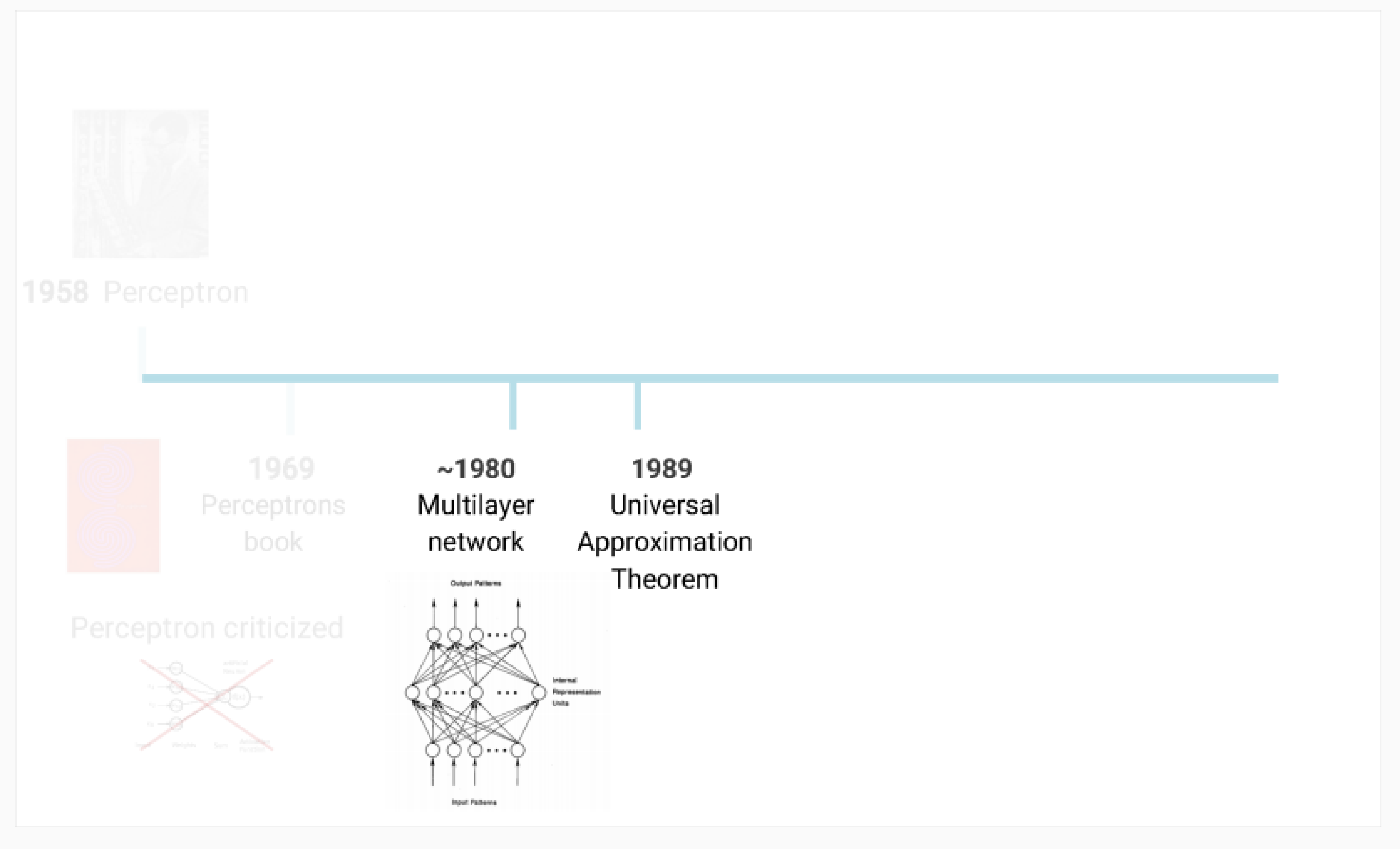
Perceptron rule.

Activation Functions

- Outputs the label given an input or a set of inputs

Activation Function	Equation	Example	1D Graph
Linear	$\phi(z) = z$	Adaline, linear regression	
Unit Step (Heaviside Function)	$\phi(z) = \begin{cases} 0 & z < 0 \\ 0.5 & z = 0 \\ 1 & z > 0 \end{cases}$	Perceptron variant	
Sign (signum)	$\phi(z) = \begin{cases} -1 & z < 0 \\ 0 & z = 0 \\ 1 & z > 0 \end{cases}$	Perceptron variant	
Piece-wise Linear	$\phi(z) = \begin{cases} 0 & z \le -\frac{1}{2} \\ z + \frac{1}{2} & -\frac{1}{2} \le z \le \frac{1}{2} \\ 1 & z \ge \frac{1}{2} \end{cases}$	Support vector machine	
Logistic (sigmoid)	$\phi(z) = \frac{1}{1 + e^{-z}}$	Logistic regression, Multilayer NN	
Hyperbolic Tangent (tanh)	$\phi(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$	Multilayer NN, RNNs	
ReLU	$\phi(z) = \begin{cases} 0 & z < 0 \\ z & z > 0 \end{cases}$	Multilayer NN, CNNs	

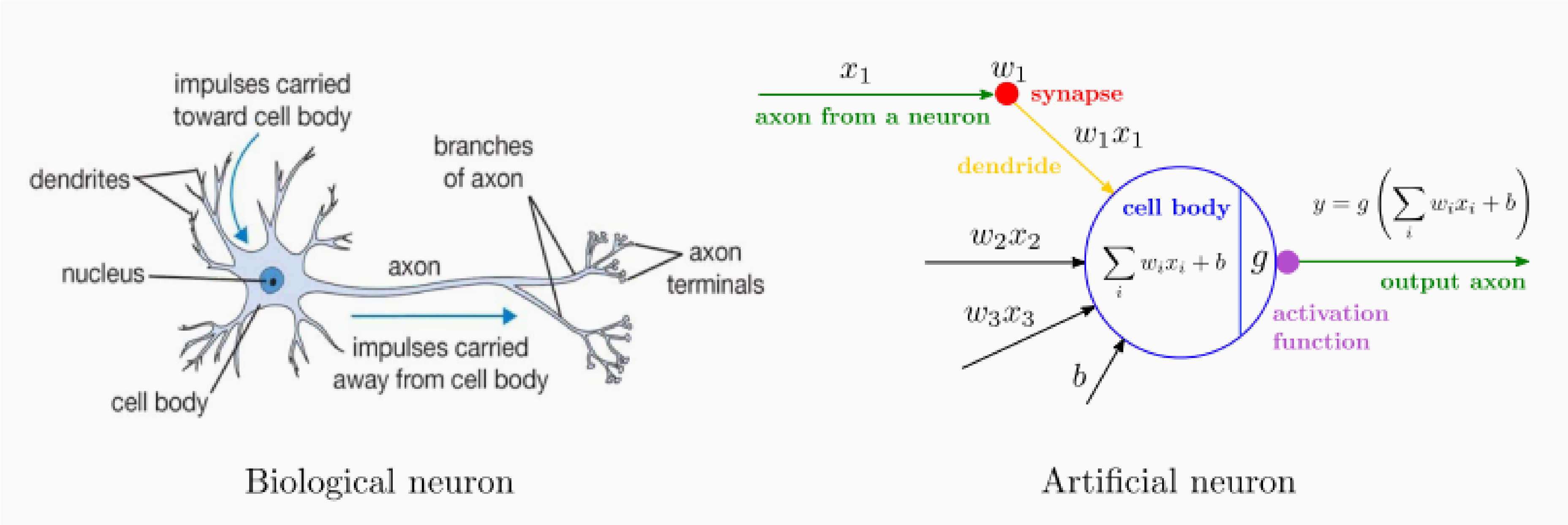
Multilayer Perceptron



Mulilayer Perceptron

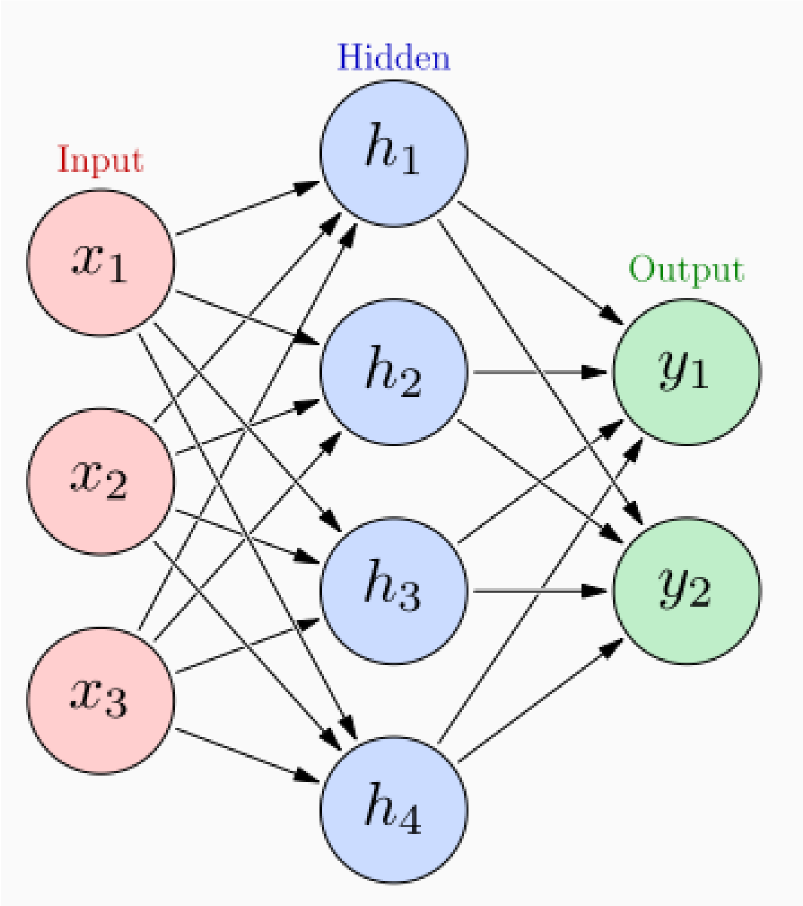
- AKA Artificial Neural Networks

- Artificial Neuron



Artificial Neural Networks

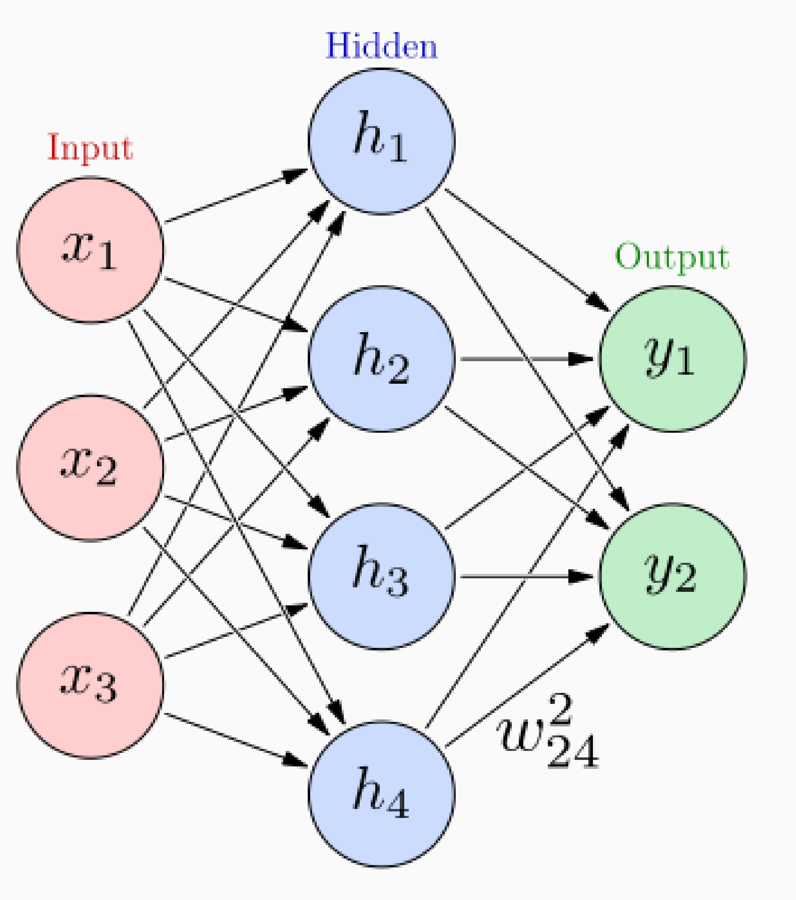
- Inter-connection of several artificial neurons (also called nodes or units);
- Each "level" in the graph is called a lauer:
  - Input layer;
  - Hidden layer(s);
  - Output layer.
- Each neuron in the hidden layers acts as a classifier / feature detector;
- Fedforward neural network (no cycles):
  - First na simplest type of neural network;
  - Information moves in one direction.



Artificial Neural Networks

$$h_1 = g_1\left(w^1_{11}x_1 + w^1_{12}x_2 + w^1_{13}x_3 + b^1_1\right)$$
$$h_2 = g_1\left(w^1_{21}x_1 + w^1_{22}x_2 + w^1_{23}x_3 + b^1_2\right)$$
$$h_3 = g_1\left(w^1_{31}x_1 + w^1_{32}x_2 + w^1_{33}x_3 + b^1_3\right)$$
$$h_4 = g_1\left(w^1_{41}x_1 + w^1_{42}x_2 + w^1_{43}x_3 + b^1_4\right)$$

$$y_1 = g_2\left(w^2_{11}h_1 + w^2_{12}h_2 + w^2_{13}h_3 + w^2_{14}h_4 + b^2_1\right)$$
$$y_2 = g_2\left(w^2_{21}h_1 + w^2_{22}h_2 + w^2_{23}h_3 + w^2_{24}h_4 + b^2_2\right)$$



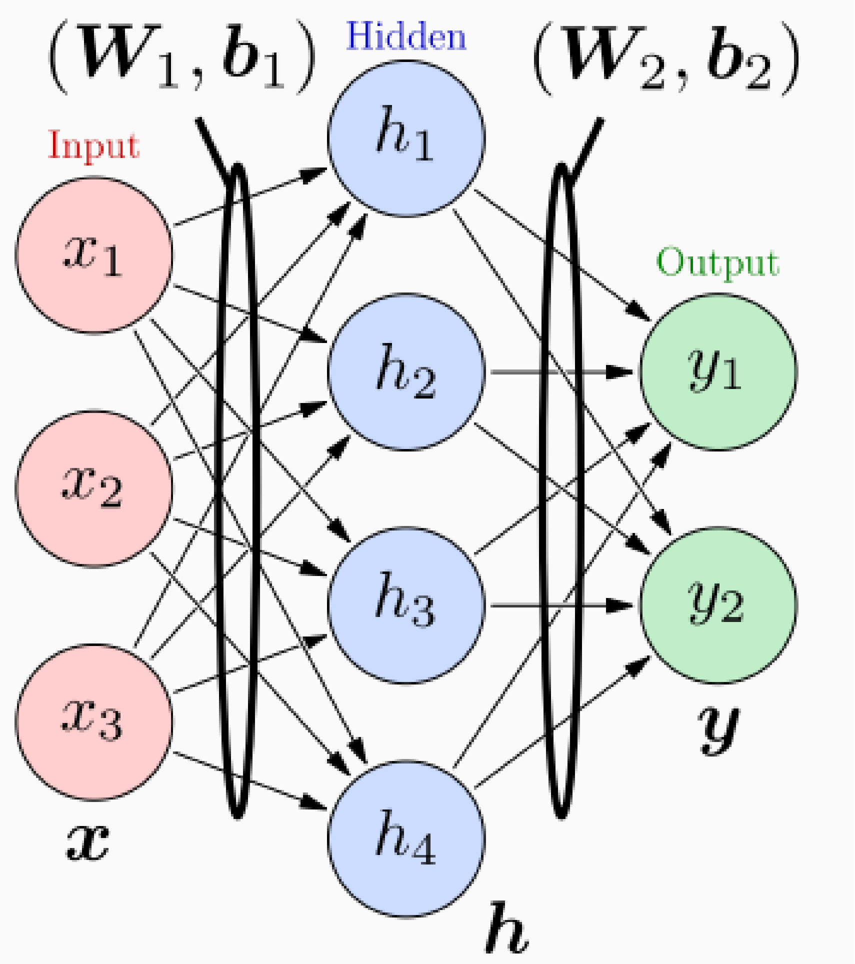
- w^k_{ij} weight between previous node j and next node i at layer k;
- g_k is any activation function applied to each its input vector

Artificial Neural Networks

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

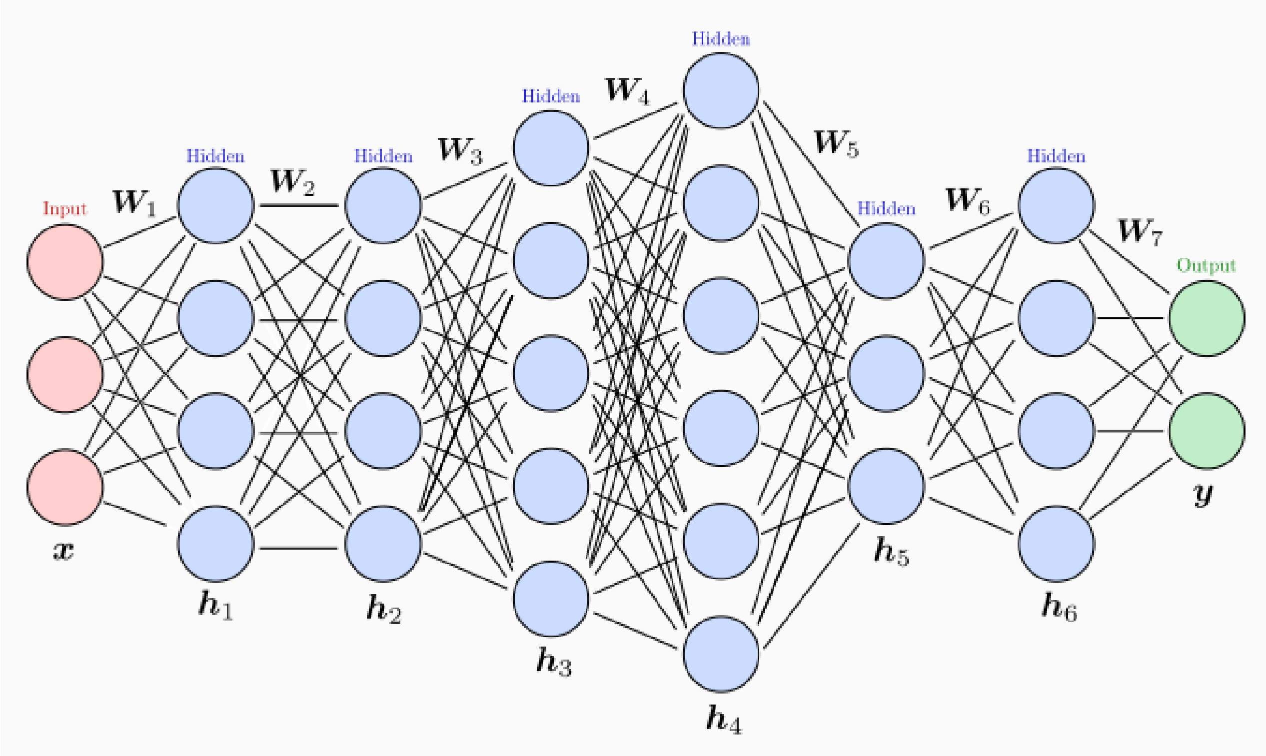
$$h_1 = g_1\left(w^1_{11}x_1 + w^1_{12}x_2 + w^1_{13}x_3 + b^1_1\right)$$
$$h_2 = g_1\left(w^1_{21}x_1 + w^1_{22}x_2 + w^1_{23}x_3 + b^1_2\right)$$
$$h_3 = g_1\left(w^1_{31}x_1 + w^1_{32}x_2 + w^1_{33}x_3 + b^1_3\right)$$
$$h_4 = g_1\left(w^1_{41}x_1 + w^1_{42}x_2 + w^1_{43}x_3 + b^1_4\right)$$
$$\boldsymbol{h} = g_1\left(\boldsymbol{W}_1\boldsymbol{x} + \boldsymbol{b}_1\right)$$

$$y_1 = g_2\left(w^2_{11}h_1 + w^2_{12}h_2 + w^2_{13}h_3 + w^2_{14}h_4 + b^2_1\right)$$
$$y_2 = g_2\left(w^2_{21}h_1 + w^2_{22}h_2 + w^2_{23}h_3 + w^2_{24}h_4 + b^2_2\right)$$
$$\boldsymbol{y} = g_2\left(\boldsymbol{W}_2\boldsymbol{h} + \boldsymbol{b}_2\right)$$



- The matrices W_k and biases b_k are learned from labeled training data.

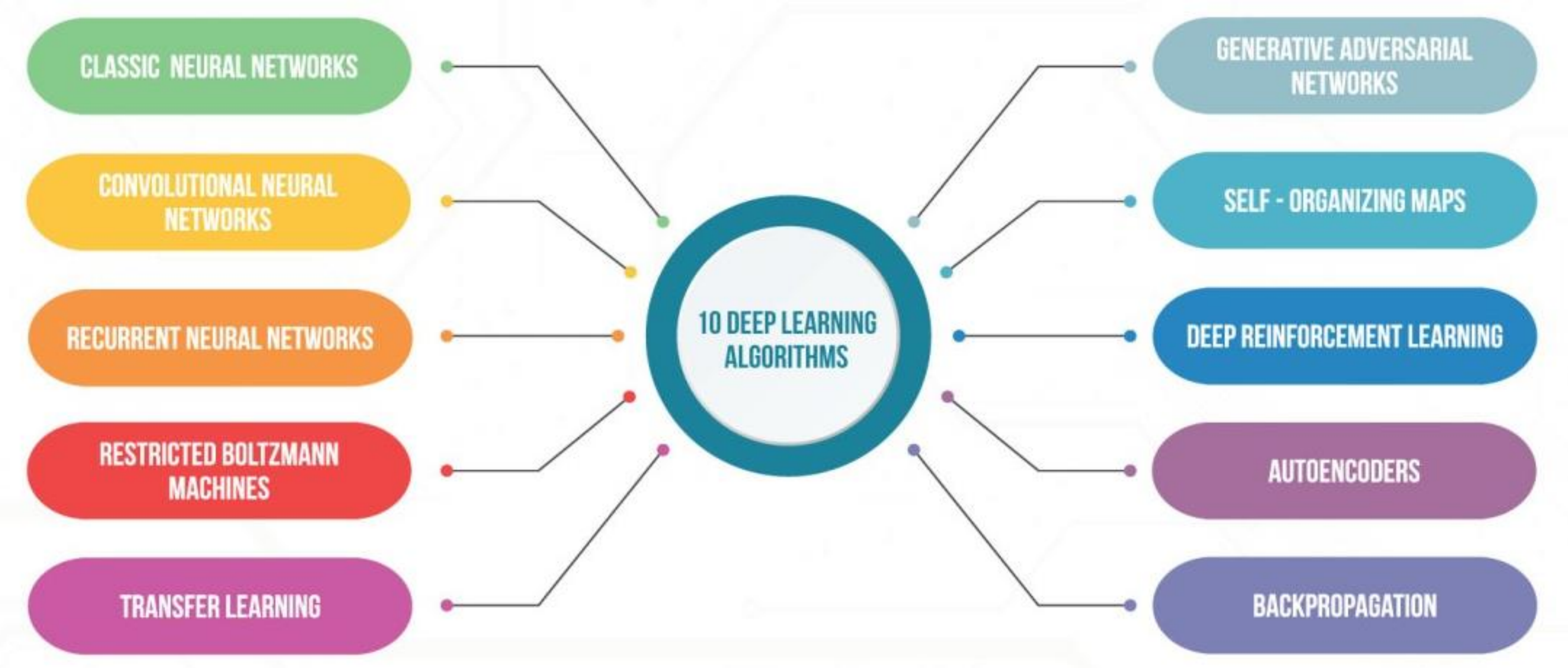
Artificial Neural Networks



- It can have 1 hidden layer only (shallow network);
- It can have more than 1 hidden layer (deep network);
- Each layer can have a different size, and hidden and output layers often have different activation functions.

Deep Learning

- Not covered in this curricular unit!



Resources

- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. In Psychological Review (Vol. 65, Issue 6, pp. 386–408). American Psychological Association (APA). https://doi.org/10.1037/h0042519

- https://simplilearn.com/tutorials/deep-learning-tutorial/perceptron