Chapter 6:
Multilayer Neural Networks

- Introduction
- Perceptron
- 3-Layer ANN

INTRODUCTION

- **Computing paradigms:**
  - **Symbolic Reasoning Systems:** von Neumann machines based on the processing/memory abstraction of human information processing. These systems manipulate symbols through the mathematical logic of formal reasoning or process information through statistical methods.
  - **Sub-symbolic Connectionist Systems:** highly-parallel architectures that process only numeric symbols and appear as networks of cells modeling the neurons of the brain.

- **Artificial Neural Networks (ANN)** are connectionist systems representing a different paradigm for computing based on the parallel architecture of animal brains. A neural network is a form of a multiprocessor computer system with:
  - simple and independent processing elements,
  - a high degree of interconnection,
  - a simple scalar information flow, and
  - an adaptive interaction between elements.

TERMINOLOGY

- **Definition:** an **Artificial Neural Network (ANN)** is an artificial information processing system that is non-algorithmic, non-digital and intensively parallel (i.e. it is neither a computer nor programmed like a computer).

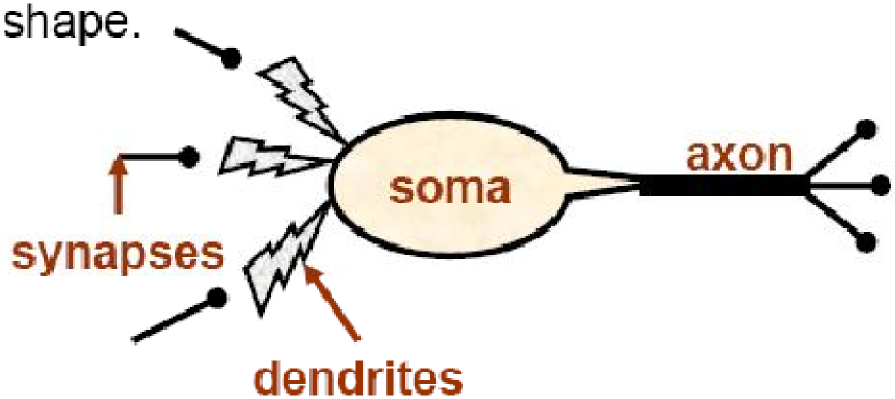
- **Structure:** an ANN consists of a number of a very simple and highly interconnected processors called **neurodes**, which are the analogs of the biological neural cells (called **neurons**) in the brain.

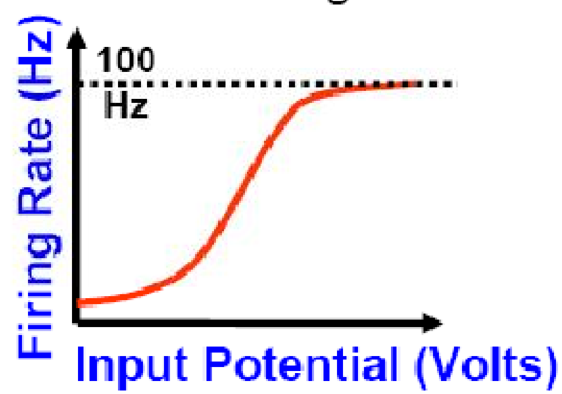
- **Connections: neurodes** are connected by a large number of weighted links, over which signal can pass. Each neurode receives many signals over its incoming connections, but it never produces more than one single outgoing signal.

- **Biological Analogy:** neurodes are only a crude approximation of biological neurons and artificial neural networks cannot be assumed to operate exactly the way a biological neural network does.

NEURON MODEL

- A **biological neuron**, or nerve cell, has **four basic parts:**
  - **Dendrites:** branch-like structures providing sensory input to the cell body.
  - **Cell body** or **Soma:** contains the cell nucleus; it sums the membrane potential provided by the connected synapses and it "fires" an electrical spike through the axon.
  - **Axon:** carries out the electrical signal to subsequent synapses.
  - **Synapses** (or synaptic buttons): serve as output devices.
- The cell body sums the membrane potential between the synapses and the dendrites, and "fires" voltage spikes down its output axon at a rate which depends on the sum of the input voltage.
  - Max. firing rate may be several hundred spikes/second.
  - Membrane potential varies from –0.1 to 0 Volts (for 0 to max. firing rate).
- The neuron **activation function** has been found to be sigmoid in shape.





HUMAN BRAIN NETWORK

- **Brain metrics:**
  - Number of neurons: ≈10-100 billion (10^{10} – 10^{11}).
  - Connections (number of inputs) per neuron: ≈10-100 thousand (10^4 – 10^5).
  - Neuron switching time: ≈1 millisecond (10^{-3} second)
  - Scene recognition time: ≈ 0.1second.

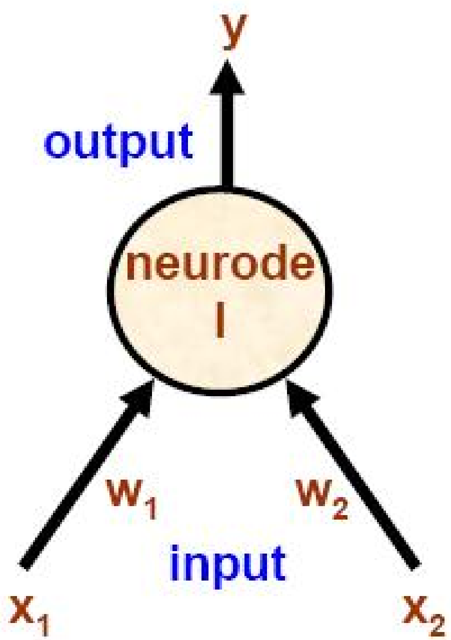
- **Human Brain versus ANN:**
  - **Speed:** neurons in the human brain transmit signals significantly slower (equivalent to a 12 KHz clock: up to 5 x 10^5 slower) than neurodes in an artificial neural network.
  - **Connectivity:** relative lack of speed in human neurons is compensated by the "massive parallelism" of the brain (yet to be achieved by an artificial neural network or any type of parallel computer).
  - **Rate of Efficiency** (on power consumption): 6 to 10 times greater in favour of the brain.
  - **Performance:** to double every 12 to 18 months for neural nets, while it takes 100,000 years for mammals to expand their brain capacity by 1 cubic inch.

BRIEF HISTORY: EARLY YEARS

- **Promise:** simple way to learn complex concepts.
- 1943: *Warren McCulloch* (psychiatrist) and *Walter Pitts* (mathematician) explored computational capabilities of network made of simple neurons.
  - Simulate logical functions through networks of neurons that fire if sum of its excitatory inputs exceed its threshold (and there are no inhibitory inputs).
- 1949: *Donald Hebb* introduced the concept of learning (now known as **Hebbian learning**) in his book "The Organization of Behaviour".
  - Hebbian rule: effectiveness of a variable synapse between two neurons is increased by the repeated activation of one neuron (similar to the biological process in which a neural pathway is strengthened each time it is used).
- 1958: *Frank Rosenblatt* defined the **perceptron** model as the simplest form of a neural network.
  - The perceptron consists of three layered parts (sensory, association and response), but only one layer has variable weights.
  - The perceptron is actually a single neuron with adjustable synaptic weights and a threshold activation function.

PERCEPTRON

- **Conceptual model:** Warren McCulloch and Walter Pitts (1943).
- **Training concept:** Donald Hebb (1949).
- **Training algorithm:** Frank Rosenblatt (1958).
- **Problem addressed:** linearly separating patterns into two categories.
- **Perceptron Convergence Theorem:** learning algorithm will always converge to a solution if the (two) classes are linearly separable.



- **Transfer Function:** polynomial

$$I = \Sigma w_i x_i$$

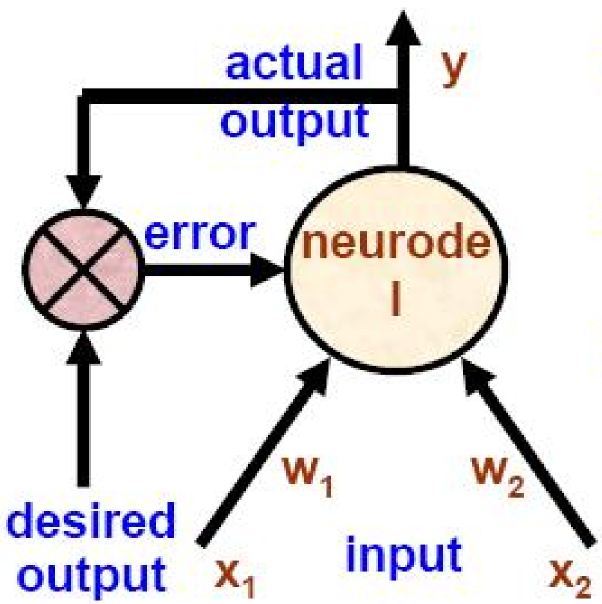
- **Activation (Output) Function:** bipolar. For a given threshold T, the output is:
  - $y = -1$ if $I < T$
  - $y = +1$ if $I \geq T$
- **Rosenblatt's Training Law:**

$$w_{new} = w_{old} + \beta y x$$

  - $\beta = -1$ if the perceptron answer is wrong.
  - $\beta = +1$ if the perceptron answer is right.

ADALINE

- 1960: *Bernard Widrow* and *Ted Hoff* proposed the Least-Mean-Square (**LMS**) algorithm (also known as the **delta rule** or Widrow-Hoff rule) for training the **Adaline** (**Ada**ptive **line**ar element).
  - **Madaline:** many Adalines.
  - **Activation function: linear** (instead the threshold used by perceptron).
  - **Training Algorithm:**: **delta rule** (minimizing average squared error).



- **Transfer Function:** polynomial.

$$I = \Sigma w_i x_i$$

- **Output Function:** linear.

$$y = a \cdot I + b$$

- **Delta Rule:**

$$w_{new} = w_{old} + \beta E x$$

where:

E = <desired output> - <actual output>

β is a **learning constant** ($0 < \beta \leq 1$)

EXAMPLE: AND-GATE MODEL

- **AND Gate Neurode:**
  - **Transfer Function:** $I = w_1 x_1 + w_2 x_2$ with $w_1 = 0.5$, $w_2 = 0.5$.
  - **Activation Function:** $y = 0$ if $I < 0.6$ and $y = 1$ if $I \geq 0.6$.

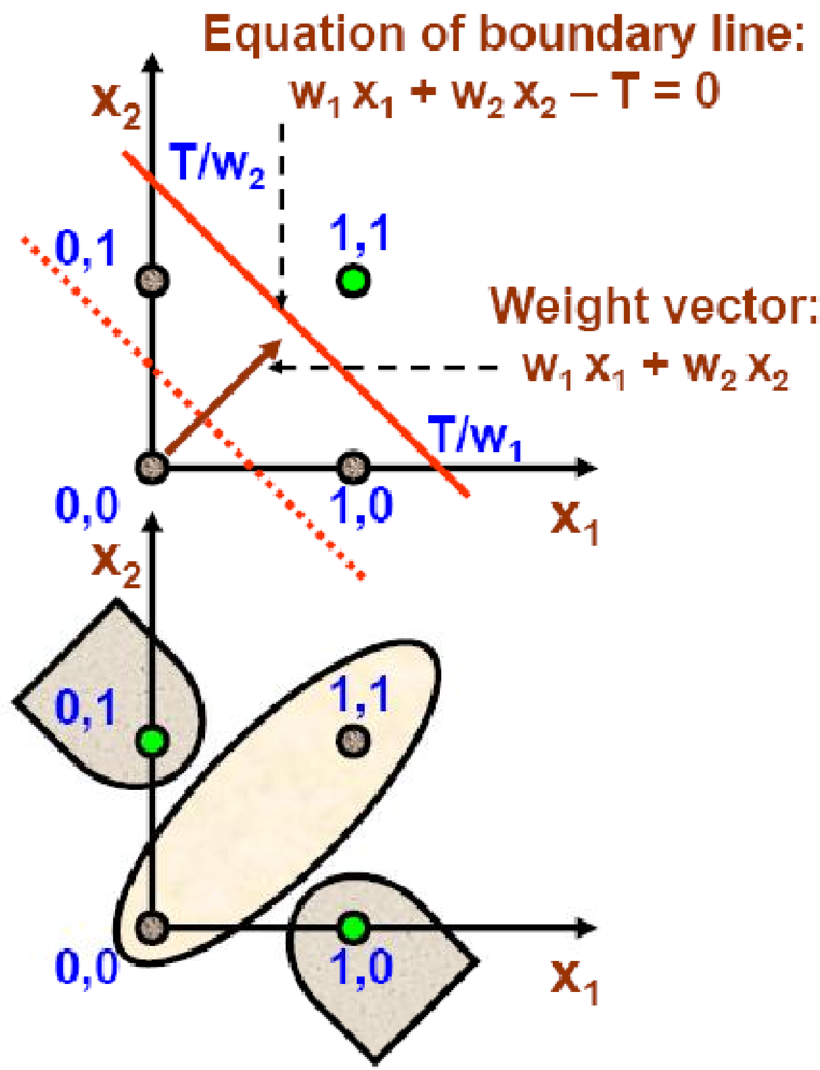
x_1	x_2	I	y
0	0	0.0	0
1	0	0.5	0
0	1	0.5	0
1	1	1.0	1

- The neurode is said to **recognize** (classify correctly), **know** or **understand** the input patterns, such as (1,1).
- If wrongly classifying a pattern, the neurode is said **not to know**, **misunderstand** or **miss** the pattern.

PERCEPTRON LIMITATIONS

- **Linearly separable** patterns can be separated by a single line and can be classified by a perceptron. Examples: AND and OR gates.

- **Non-linearly separable** patterns cannot be separated by a single line and cannot be classified by a perceptron. Example: XOR gate.



DARK AGES: DISILLUSIONMENT

- 1969: *Marvin Minsky* and *Seymour Papert* published “Perceptrons”, a book proving mathematically the limitations of Rosenblatt’s perceptrons:
  - “Appalled by the persistent influence of perceptrons (and similar ways of thinking) on practical pattern recognition, we determined to set out our work as a book.”
  - “Perceptrons have been widely publicized as “pattern recognition” or “learning machines” and as such have been discussed in a large number of books, journal articles, and voluminous reports. Most of this writing is without scientific value and we will not refer by name to the works we criticize.”
  - “The real-world problem are not always linearly separable, and therefore neural networks cannot be used to solve such problems.”

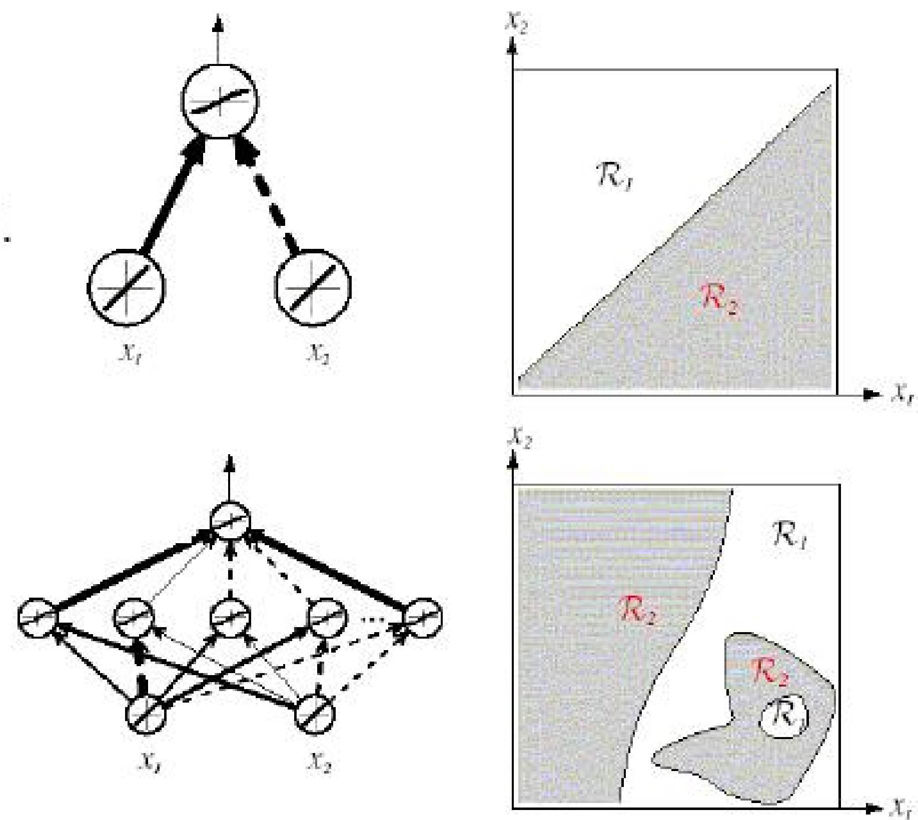
- As a result of this monograph, research in neural nets was almost abandoned in the 1970s. The handful of researchers that continued working on ANN were mostly outside US.
  - 1972, Teuvo Kohonen: **associative memory**.
  - 1973: vad der Malsburg: **self-organizing maps**.
  - 1975, Kuniko Fukushima: **neocognitron** (multi-layer perceptron).
  - 1976, Stephen Grossberg: **associative learning.**

MEET THE HIDDEN LAYER

Whereas a two-layer network classifier can only implement a linear decision boundary, given an adequate number of hidden units, three-, four- and higher-layer networks can implement arbitrary decision boundaries.

The decision regions need not be convex or simply connected.

It is considered that most practical problems require no more than two hidden layers.



REBIRTH OF CONNECTIONISM

- 1982, *John Hopfield* (a physicist) presented a class of recurrent networks (known as **Hopfield nets**) as associative memories with elementary physical properties and explained their behavior through statistical mechanism and a thermodynamic theory of computational energy.
- **Birth of non-linear networks:**
  - **Idea:** expand the two-layer network model to include provisions for additional middle ("**hidden**") layers.
  - **Needed:** an algorithm to train neural nets with more than two layers (preferably one that uses continuous and non-linear activation rules).
- **Algorithms** were developed by:
  - 1974, *Paul Werbos* proposed the back-propagation paradigm and algorithm.
  - Similar algorithms: *David Parker* (1982, 1985) and *Yann LeCun* (1986).
- 1986, *David Rumelhart*, *Geoffrey Hinton* and *Ronald Williams* published a method ("**backward propagation of errors**") allowing a network to learn how to discriminate non-linearly separable patterns. The method, a generalization of LMS, is now known as the **generalized delta rule**.