EE5904/ME5404: Neural Networks

Xiang Cheng

Associate Professor

Department of Electrical & Computer Engineering
The National University of Singapore

Phone: 65166210 Office: Block E4-08-07
Email: elexc@nus.edu.sg

Assessment:

•Continuous Assessment (CA): 40%

•20% Three assignments from part I.

•20% Two mini-projects from part II.

•Final Exam: 60%

Simulation Tools: MATLAB with Deep Learning toolbox

NUS
National University of Singapore

The main reference book for part I

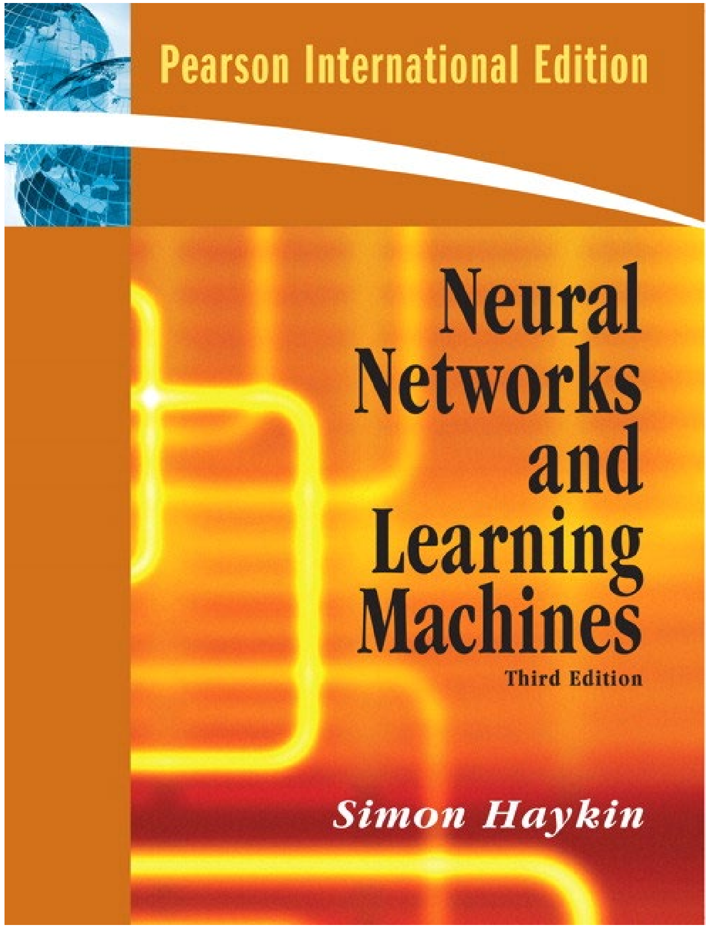
Neural Networks and Learning Machines: International Edition, 3/e

Author : HAYKIN

Publisher : Pearson

ISBN : 9780131293762

Available in NUS Library, Amazon



What should you expect from this module?

Learning Objectives

Introduce the students to fundamental concepts and applications of artificial neural networks:

1. To understand the *structures* and *learning processes* of artificial neural networks; *Human brain computes in a very different way from PC…*
2. To learn the significance of neural computing technology and its application to real world pattern classification and regression problems;
3. To experience the use of simulation tool like the deep learning toolbox of **MATLAB** to implement the artificial neural networks.

The best way to learn is to do it by yourself!

After learning this course, you will be confident that you can use "AI" to solve real world application problems.

What do I expect from you?

1. Be prepared. Roughly go through the material in the reference before the class.

2. I am going to spoon-feed you **with lots of questions !**

These questions are designed to arouse your interest and to help you to figure out most of the stuff by **your own thinking**!

I want you to have an exciting experience of discovering new knowledge inside the classroom! I will play the role of facilitator rather than a lecturer.

You will have fun by actively participating in thinking and discussing these questions.

It will be a waste of your time if you just want to know the answers without any thinking.

3. Do the projects by yourself.

You can discuss the questions with your classmates.

But do not copy and paste!

Introduction:
What is neural network? And Why?

What is the most important technology invented in 20th century?

The digital computer.

How does the digital computer process information?

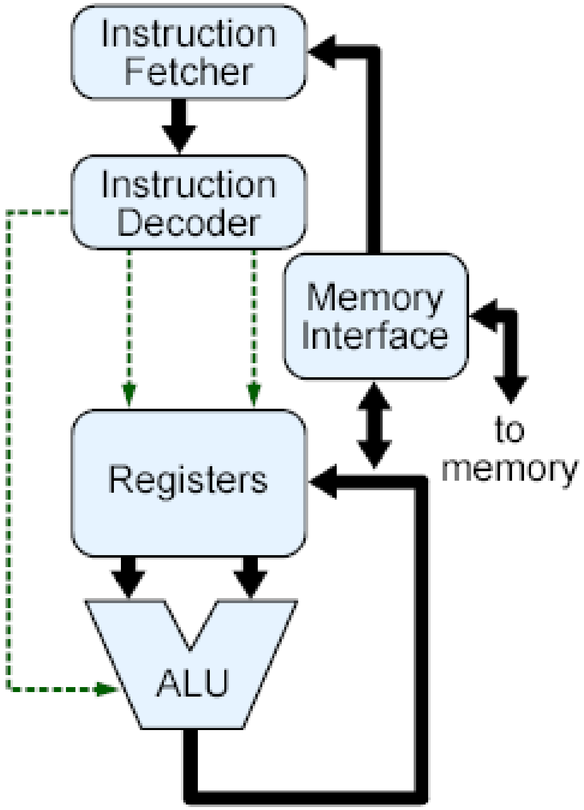
The computer performs binary operations according to a list of instructions (program).

How many operations can your laptop execute in one second?

CPU speed of 2 GHz --→ 2 billion (10^9)

How many operations can a computer (with one CPU) execute at any given instant?

•**Only ONE! The operations are serial: one after another!**



The modern computers are so fast that it may appear that many programs are running at the same time even though only one is ever executing at any given instant.

Can computer beat the human brain now?

Yes and No.

Can you list a few tasks that the computer can beat the human brain?

- Playing chess---the Deep Blue defeated the world champion Garry Kasparov in 1997.
- Solving equations! $x^5 + x + 1 = 0$

What was the most recent famous victory of computer beating the human brain?

The most recent victory of Machine v.s. Human:

Alpha Go (from DeepMind) beat the legendary go player Lee Se-dol in 2016!

Are we doomed to lose to the machine?

Don't worry. There are certain things that you can do much better than the computer

at least for now!

<u>Can anyone of you give me one example?</u>

Pattern recognition such as recognizing one familiar face among a crowd!

Half a century ago, artificial-intelligence pioneer, Marvin Minsky of MIT predicted that computers would exceed human intelligence within a generation.

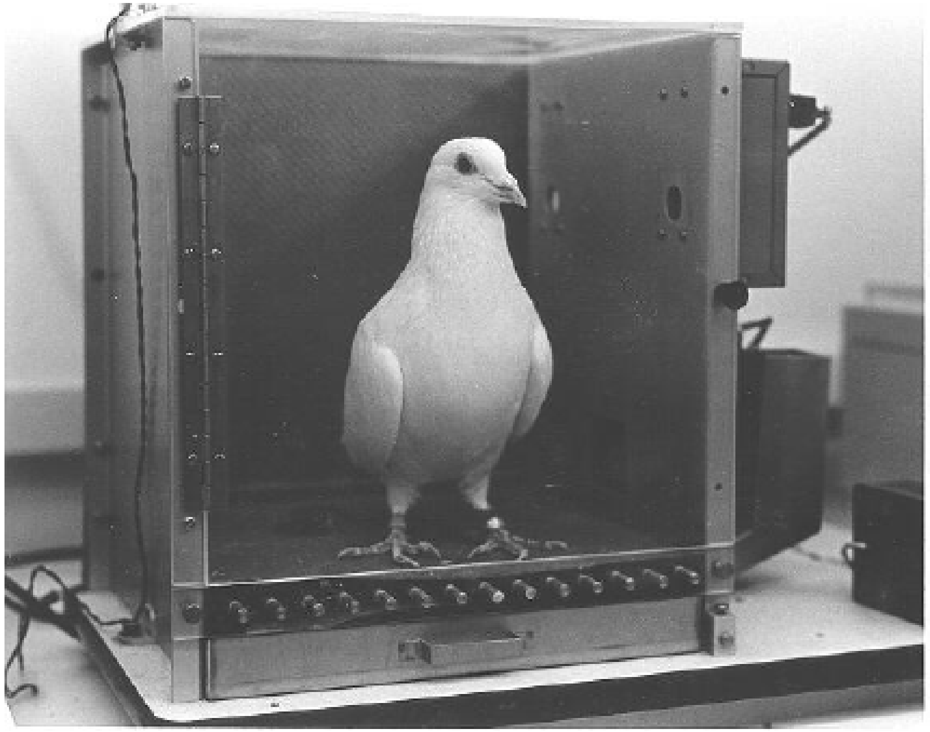
Later, he admitted:
"The world's most powerful computers lack the common sense of a toddler; they cannot even distinguish cats from dogs unless they are explicitly and painstakingly programmed to do so."

How about the brains of other animals?
Are they also good at pattern recognition?

Pigeons as art experts (Watanabe. et al. 1995)

Experiment:

Pigeon in Skinner box



Present paintings of two different artists (e.g., Chagall/Van Gogh)

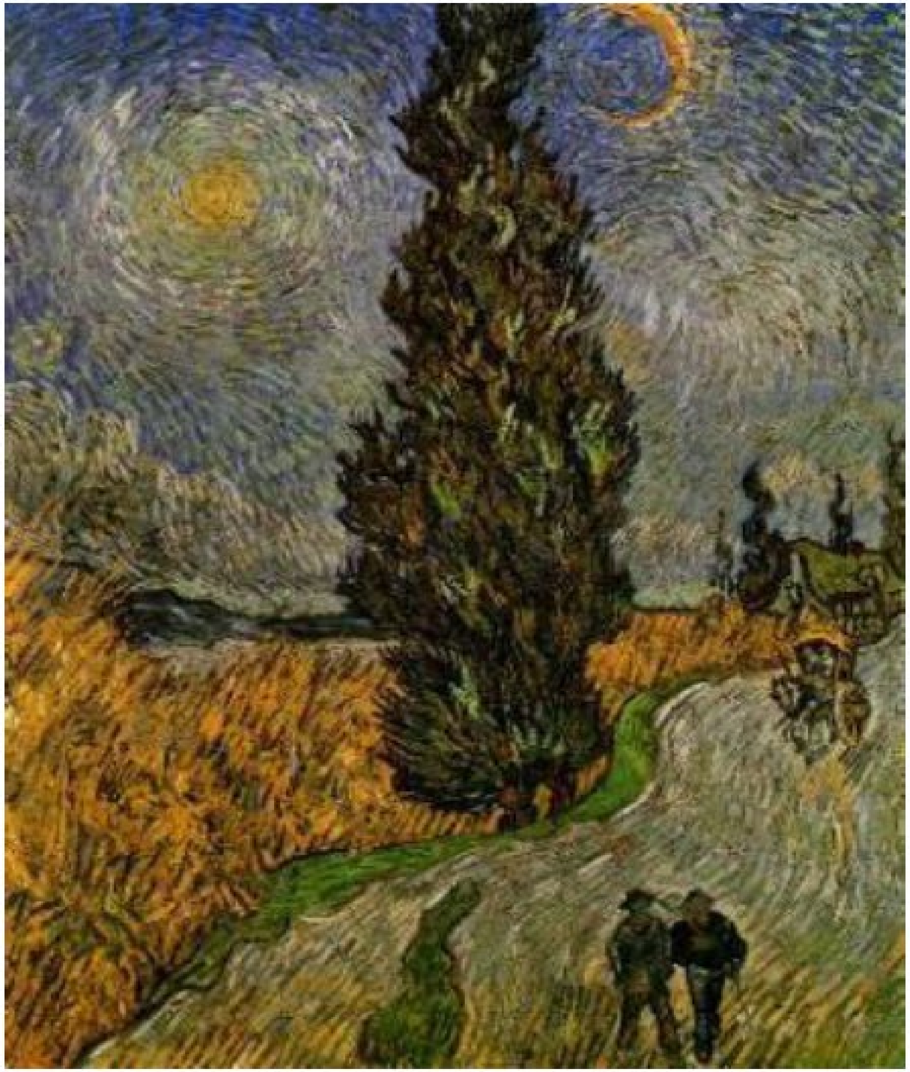
Reward the pigeon for pecking when it is presented a particular artist (e.g., Van Gogh)

Marc Chagall (1887-1985)
Russian Jewish modernism artist

Marc Chagall

Vincent Willem van Gogh (1853-1890)
Dutch Post-Impressionist artist

Vincent Van Gogh



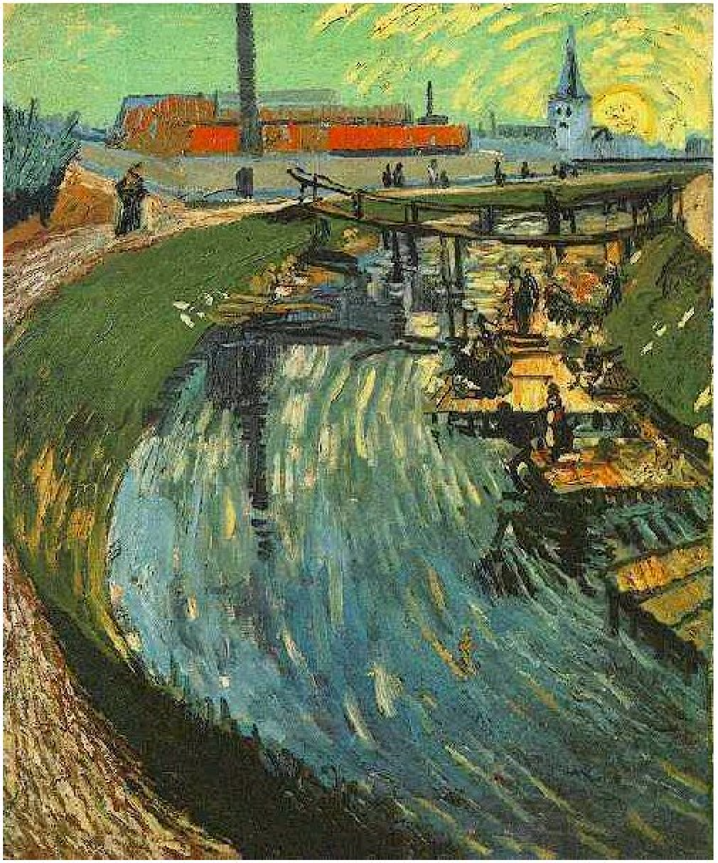
<u>What is the accuracy you can imagine for the training set?</u>

Pigeons could tell the difference between Van Gogh and Chagall with 95% accuracy (when presented with pictures they had been trained on).

<u>Can a computer achieve the same or even higher accuracy on the training set?</u>

A computer can also easily memorize all the paintings in the training set!

<u>What would happen if the pigeons are presented with something they never saw before?</u>



Can pigeon recognize new paintings?

The most remarkable thing is that it is still 85% successful!

Can a computer achieve the same feat on the new paintings?

Very unlikely at this moment!

Even the Pigeons can beat the computer!

So the pigeons do not simply memorise the pictures.

They can extract and recognize patterns (the styles of the two artists);

They learn from their training process and make predictions on the new ones.

So the human or the pigeon brains must do something very different from the computer!

National University of Singapore

Computer v.s. Human Brain

Which one is faster?

The Laptop can run 2 billion operations per second

A typical firing rate for a neuron is around 100 spikes per second.

The PC is millions times faster.

How can the brain make up for the slow rate of operation?

The secret lies in the way the operations are executed.
The operations in a computer is serial: one after another.

How about the operations in brain? Serial or Parallel?

Massively parallel! For example, when you look at me, about a million axons go from each eye to the brain, all working *simultaneously*!

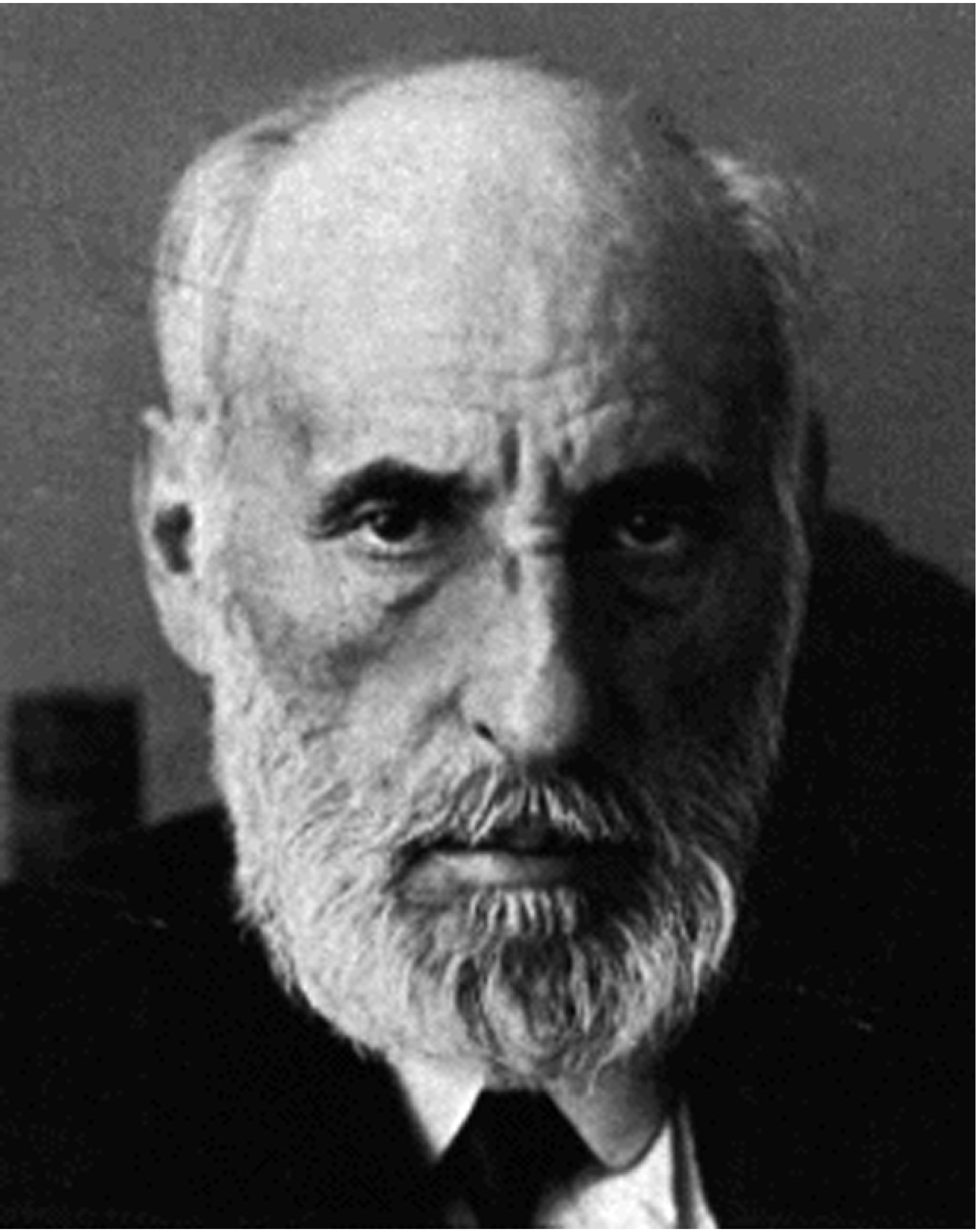
Human Brain

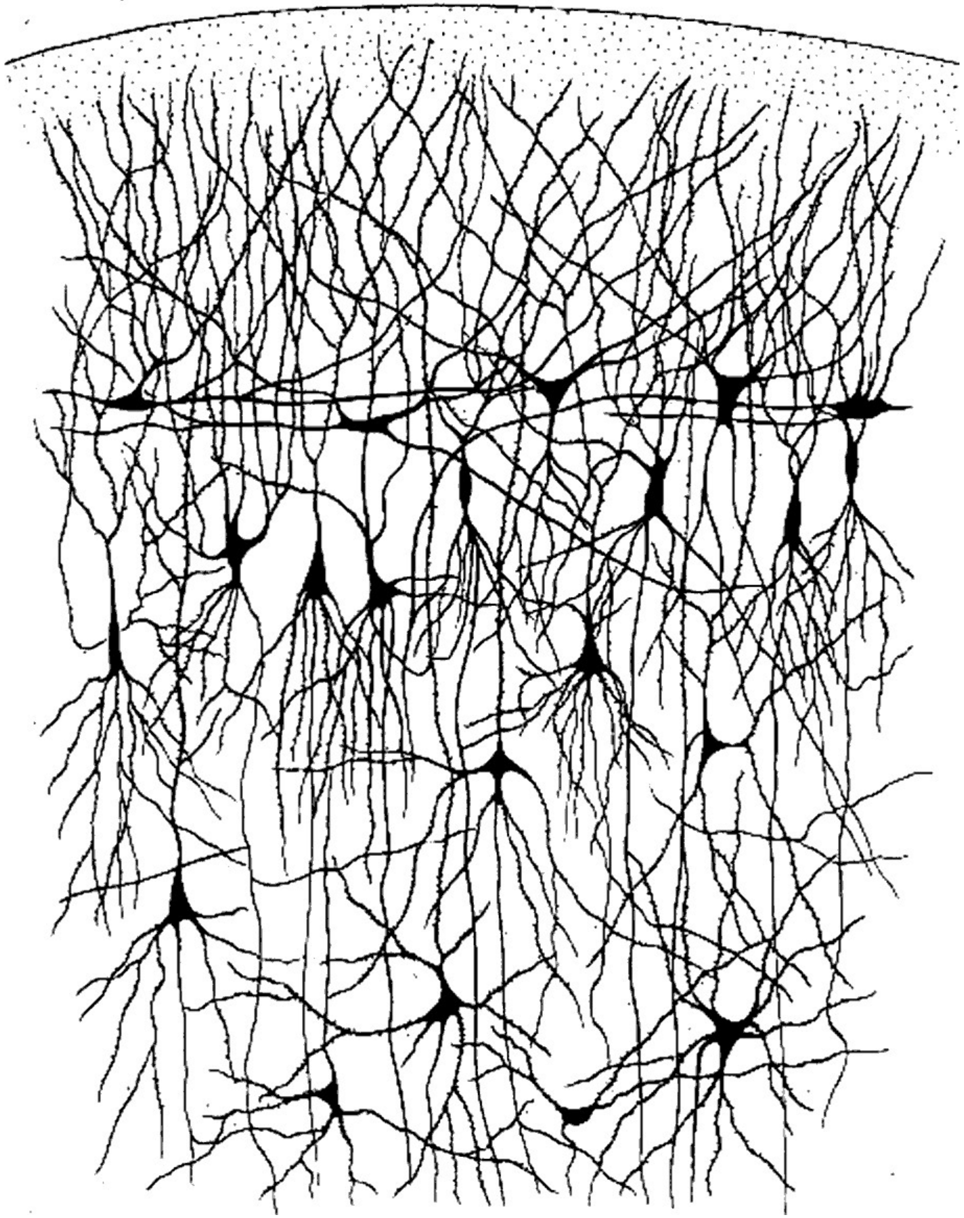
1. A huge number of nerve cells (neurons) and interconnections among them. The number of neurons in the human brain is estimated to be in the range of 100 billion (10^{11}) with quadrillion (10^{15}) synapses (interconnections).

2. The function of a biological neuron seems to be much more complex than that of a logic gate.

Summary: The brain is a *highly complex, non-linear, parallel* information processing system. It performs tasks such as pattern recognition and perception many times faster than the fastest digital computers.

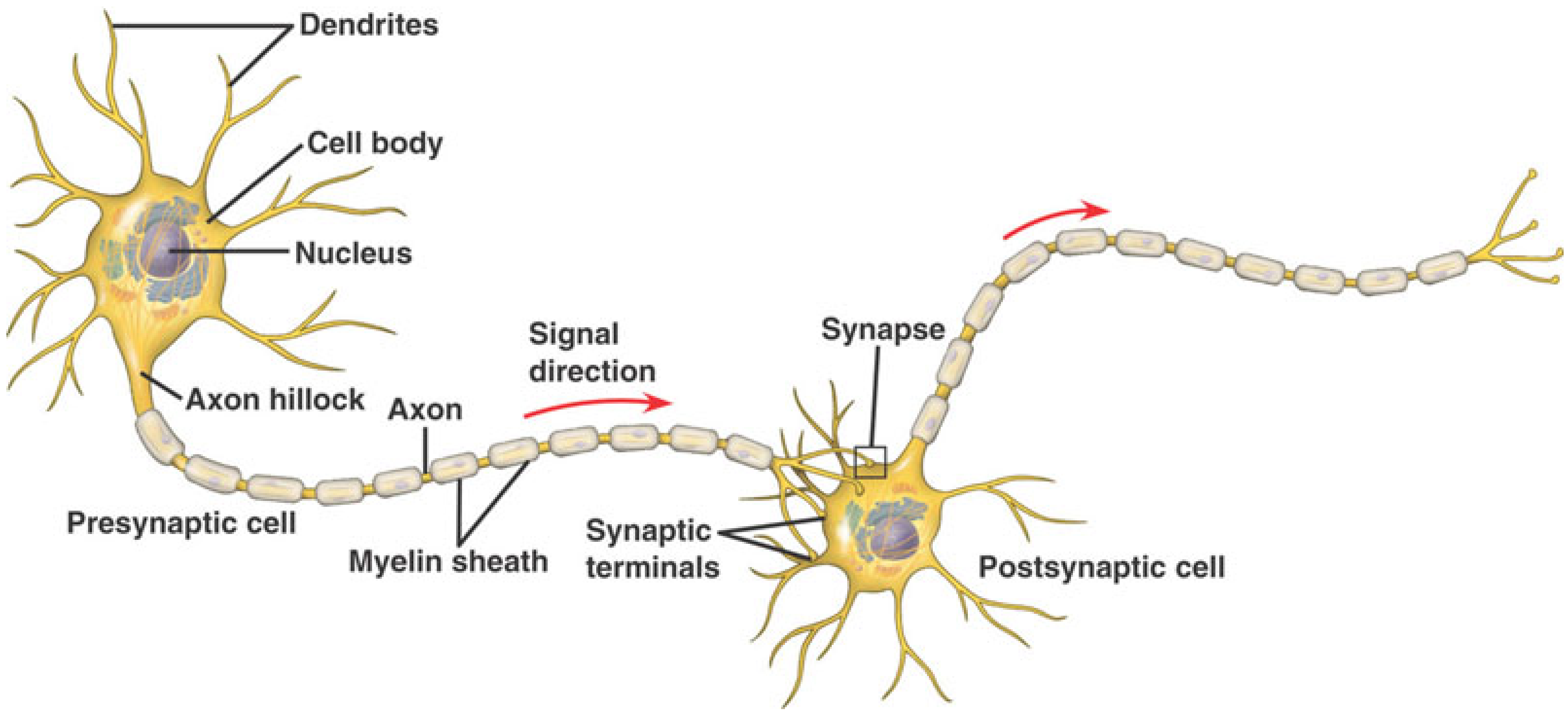
The understanding of biological neurons started more than 100 years ago:





Santiago Ramon y Cajal 1852-1934

Biological Neurons: What do they look like?



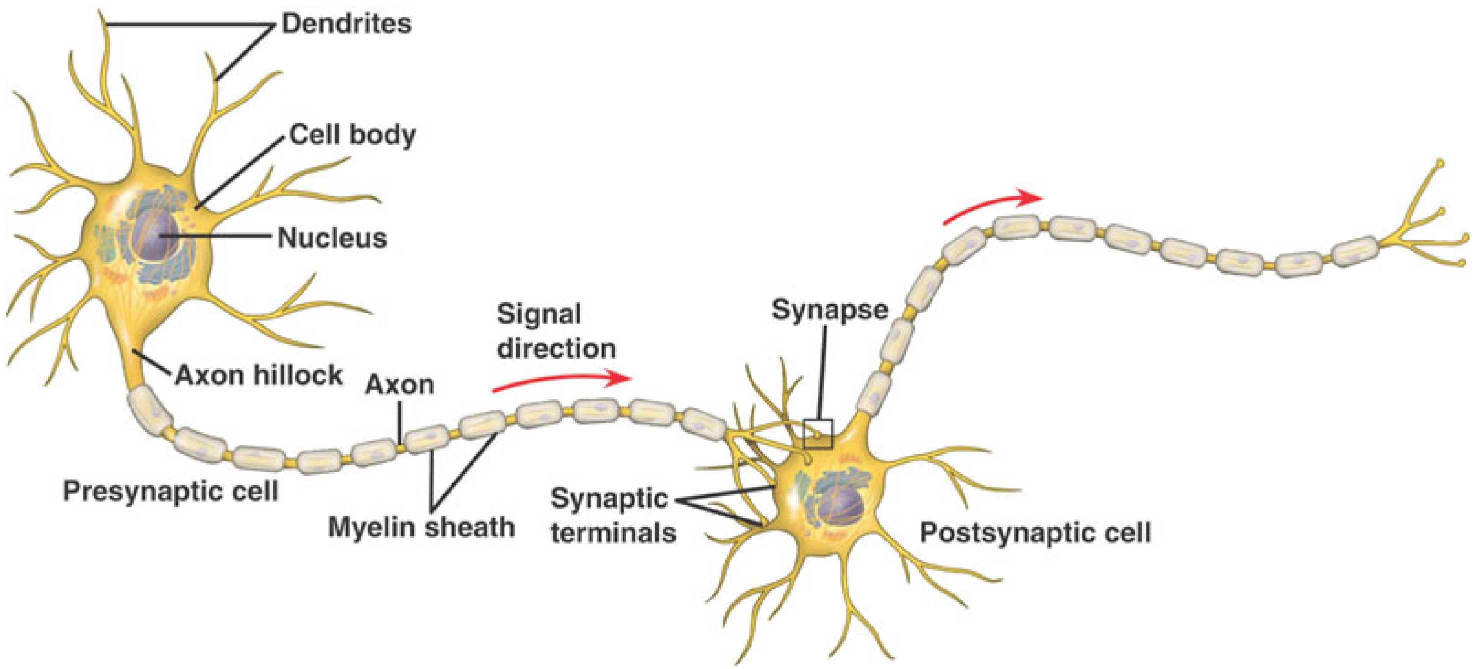
A typical biological neuron is composed of:

1. A cell body;

2. Hair-like dendrites: input channels

3. Axon: output cable; it usually branches.

Biological Neurons: What do they do?



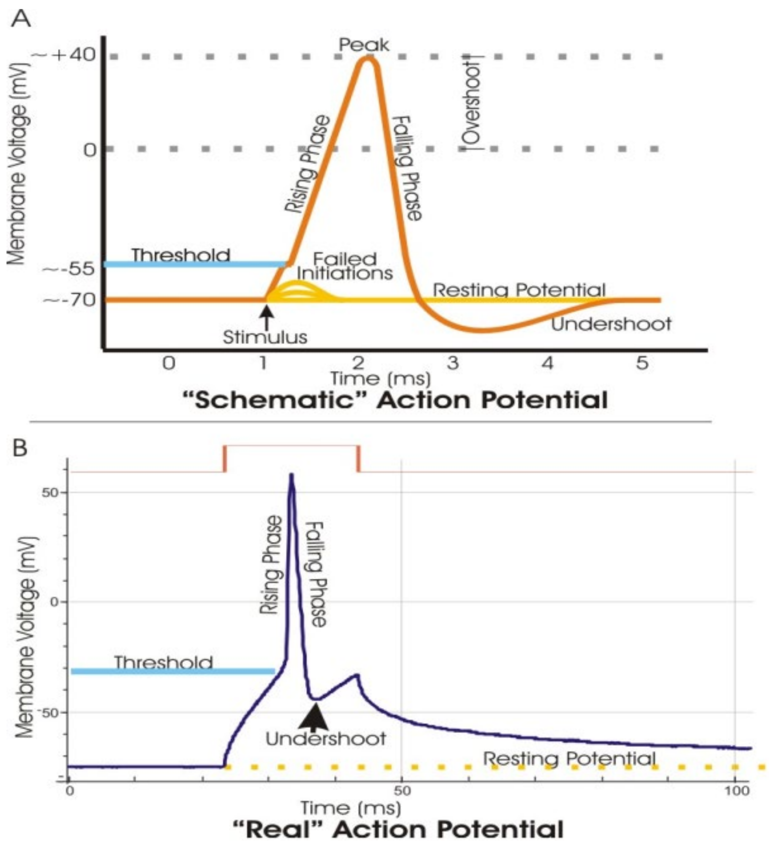
The neuron responds to many sources of electric impulses in three ways:

1. Some inputs excite the neuron;
2. Some inhibit it;
3. Some modulate its behavior.

If the neuron becomes sufficiently excited, it responds ("fires") by sending an electric pulse-(a spike)-down its output cable—(its axon). The spikes travel down each branch and sub-branches until eventually the axon contacts many other neurons and so influences their behaviors.

What are the electrical spikes fired by the neuron?

Is it caused by the moving electrons (like current in a wire) or something else?



In a neuron, the electrical effects depend upon charged atoms (ions) that move in or out of the axon through molecular gates in the membrane of the cells.

As the ions move in and out of the membrane, they make local alternations to the electrical potential (or voltage) across the cell membrane. It is this change of potential that is propagated down the axon. → Action Potential

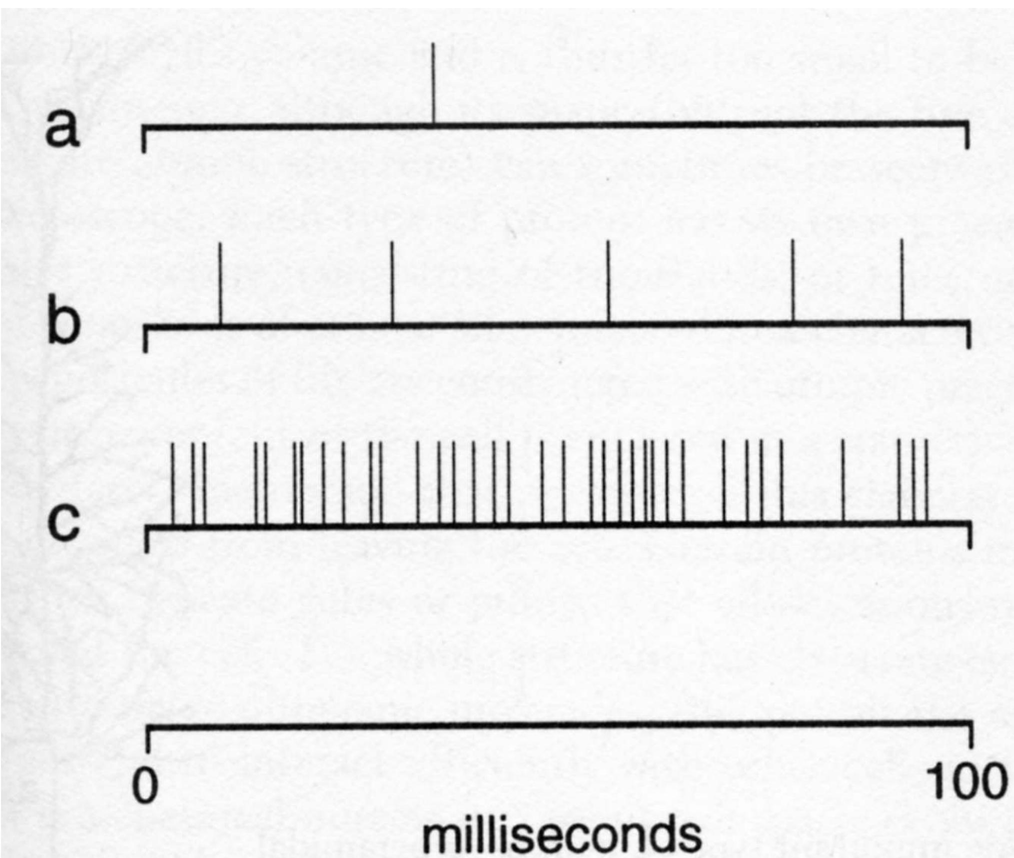
Biological Neurons: All-or-none principle and the firing rate

•The conduction of nerve impulses is an example of an <u>all-or-none</u> response.

• In other words, if a neuron responds at all, then it must respond completely. It either fires or not fires at all. There is no weak fire or strong fire.

• **<u>How do neurons respond to stimuli with different intensities?</u>**

The greater the intensity of stimulation does not produce a stronger signal but can produce *more* impulses per second, i.e. higher firing rate.



When nothing much is happening a neuron fires at relatively slow, irregular, "background" rate.

When it becomes excited, its rate of firing increases to a much higher rate.

How fast are the electrical spikes traveling?

Hermann von Helmholtz
(1821-1894)

In 1850, Helmholtz measured the propagation speed of electric stimuli in the nerve.

How did he do it more than 170 years ago?

"In a human being, a very weak electric shock is applied to a limited space of skin. When he feels the shock, he is asked to carry out a specific movement with the hand or the teeth interrupting the time measurement as soon as possible."

'*message of an impression*' propagates itself '*to the brain with a speed of around 60 Meters (180 feet) per second, or 216 km/hour, which is as fast as the high-speed train.*

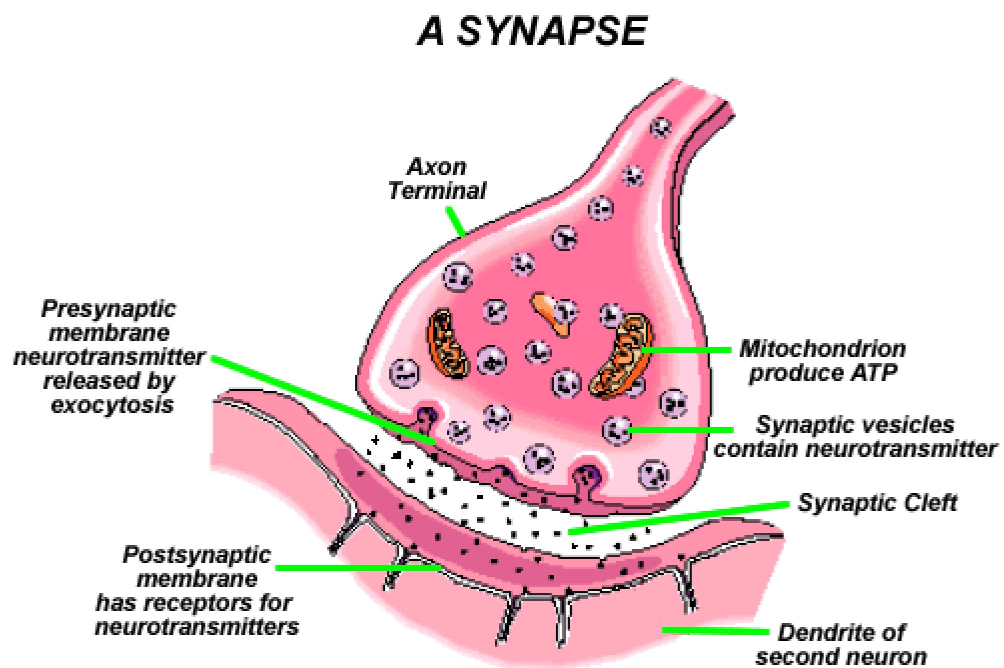
How do the spikes affect other neurons? Do they excite or inhibit the target neurons?

Synapses:

1. Small gap between the end bulb of the axon and the dendrites!

2. Basic structural and functional units that mediate the interactions between neurons.

3. Can impose excitation (active) or inhibition (inactive) on the receptive neurons.

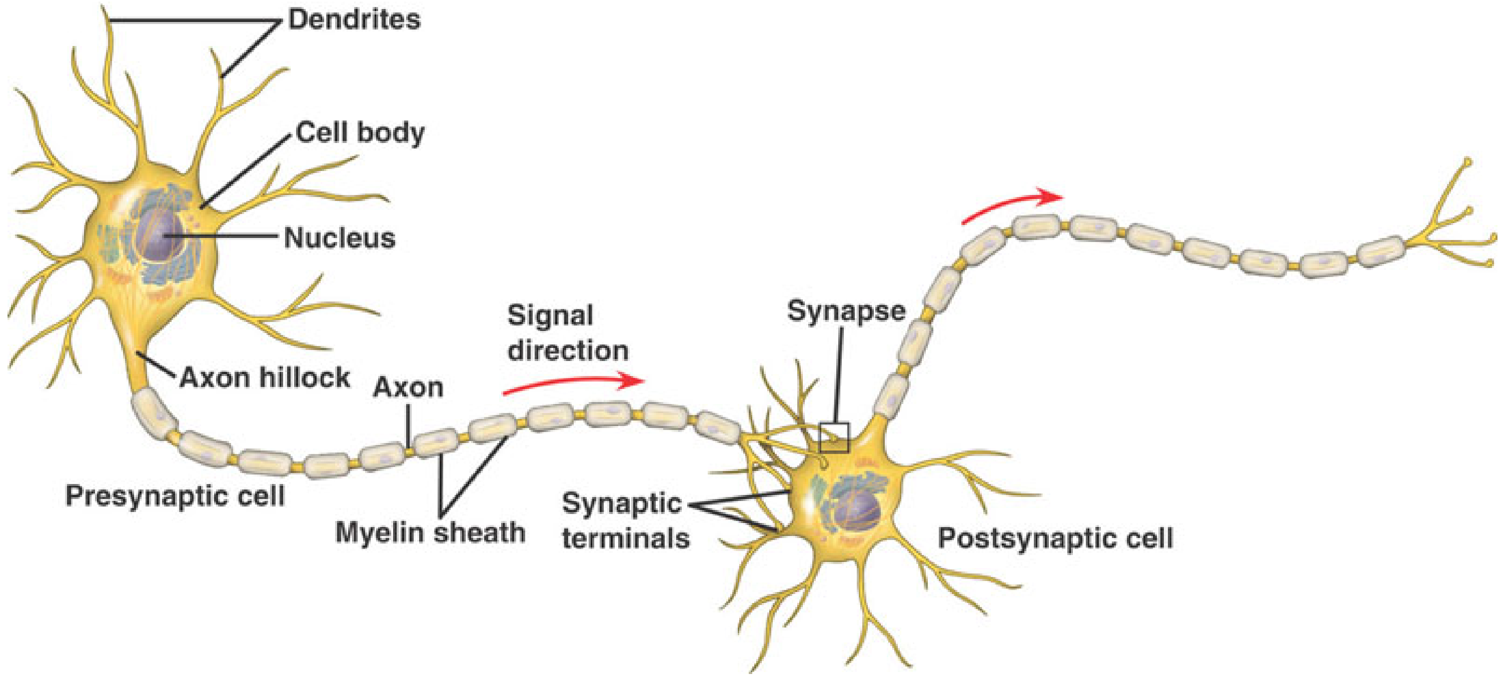


What is the mechanism?

1. When the spike arrives at the synapse, it causes little packets of chemical molecules to be released into the gap.

2. These small chemical molecules bind with the molecular gates in the membrane of the synapse of the recipient cell.

3. This causes those particular gates to open and allows charged ions to flow in or out of the membrane, so that the local potential across that membrane is changed.



[When you are learning something, what have changed in your brain? The neurons or the connections (synapses)?]

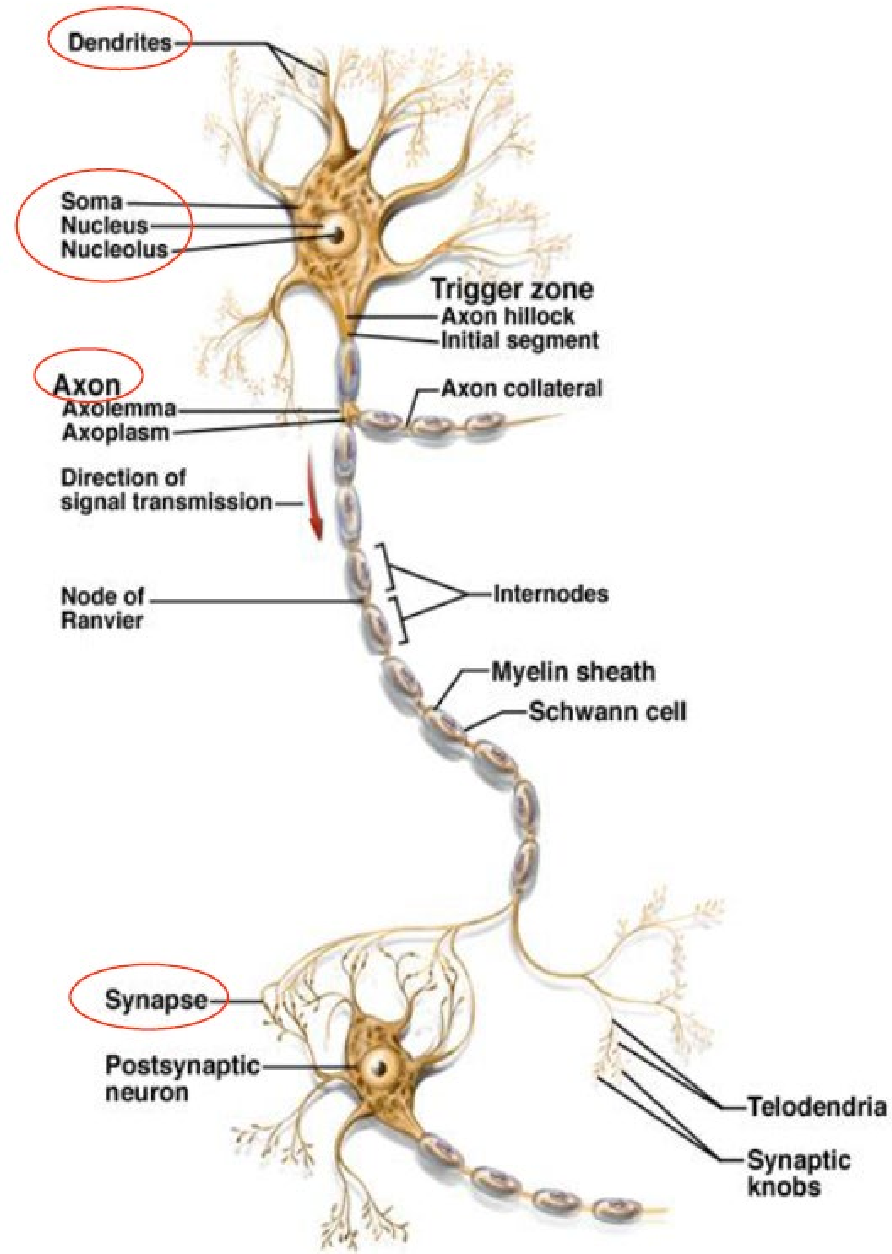
Synaptic plasticity is the ability of the connection, or synapse, between two neurons to change in strength.

Biological Neuron

The major job of neuron:

1. It receives information, usually in the form of electrical pulses, from many other neurons.
2. It does what is, in effect, a complex dynamic sum of these inputs.
3. It sends out information in the form of a stream of electrical impulses (action potential) down its axon and on to many other neurons.
4. The connections (synapses) are crucial for excitation, inhibition or modulation of the cells.
5. Learning is possible by adjusting the synapses!

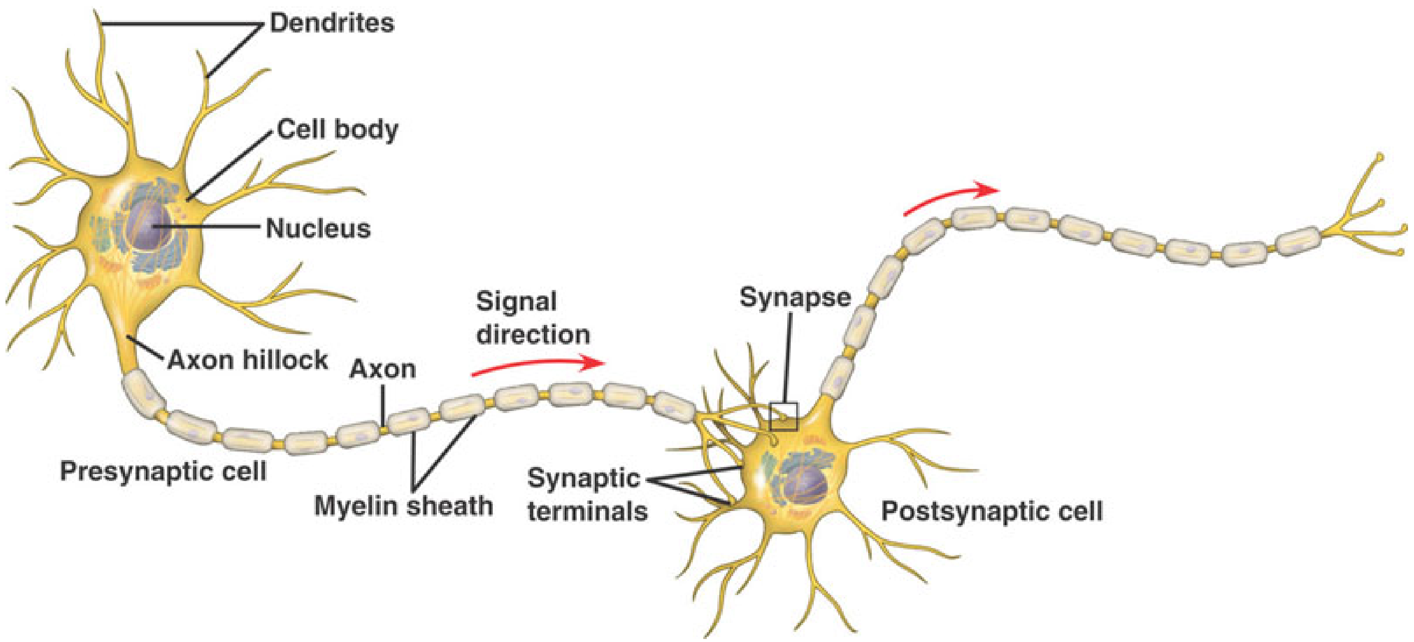
An extreme example of brain plasticity



Break

[Visions of The Future 1]

How to build the mathematical model for the neuron?



Any mathematical model is always an approximation of the reality!

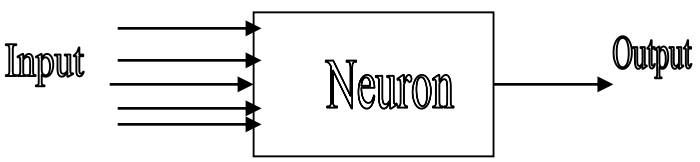
We should always start with the simplest one!

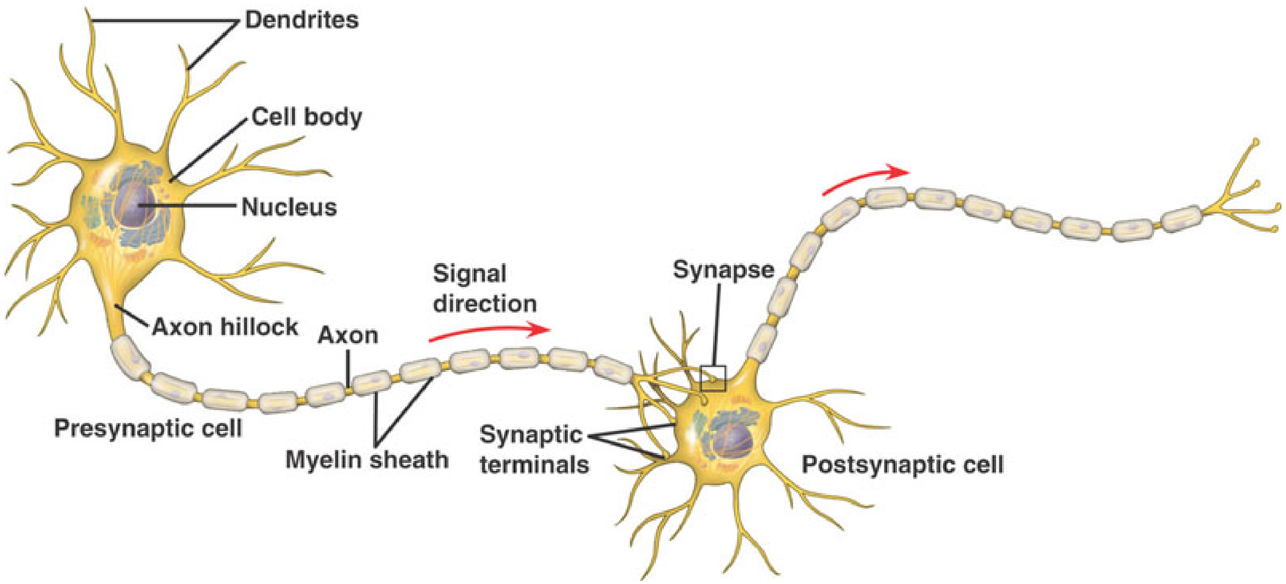
<u>Are there many inputs or only one input?</u>

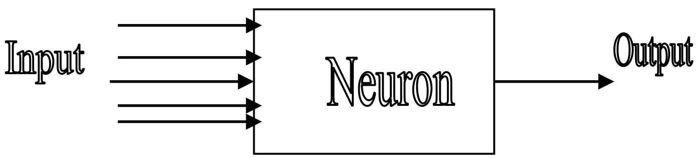
There are many dendrites!

<u>How many outputs are produced by the neuron?</u>

There is only one axon!





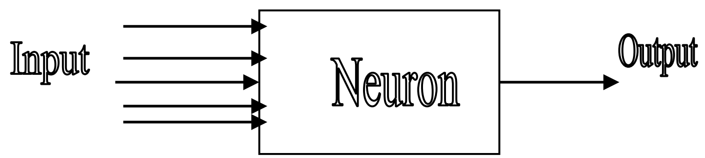


Of course, this block diagram is just a qualitative model. The next step is to build a quantitative model.

What is the simplest mathematical model of the neuron you can think of?
What is the simplest relation between the inputs and outputs?

The simplest model:

$$y = \sum_{i=1}^{m} x_i$$



The simplest model: $$y = \sum_{i=1}^{m} x_i$$

<u>If the inputs x_i are the spikes (1 or 0), is y always positive?</u>

Yes. So the neuron is almost always on fire!

<u>Does it make sense for the biological neurons?</u>

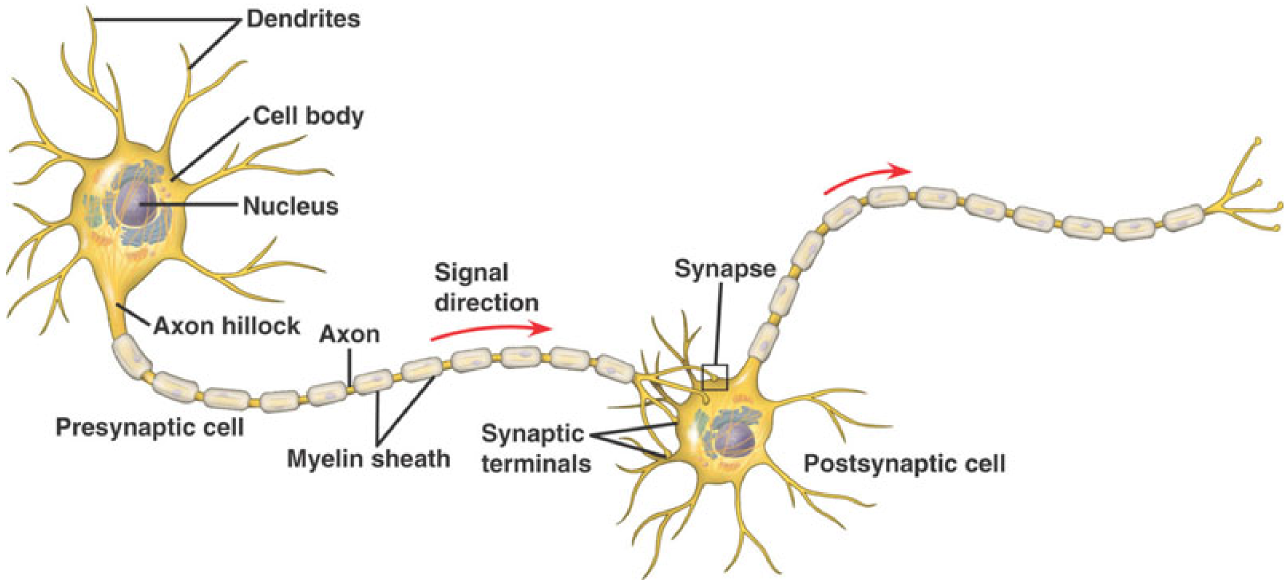
The real neurons only fire when they are sufficiently excited!

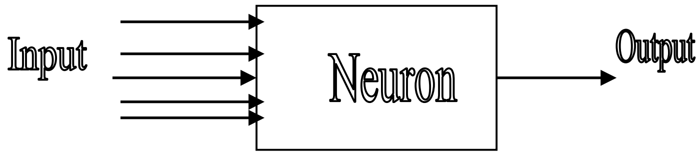
When m is large, the output can be huge!

<u>It means that there are strong fire and weak fire. Is this reasonable?</u>

The output is supposed to be fire (1) or not fire (0).

<u>How to modify the simplest model such that the output is only 1 and 0?</u>

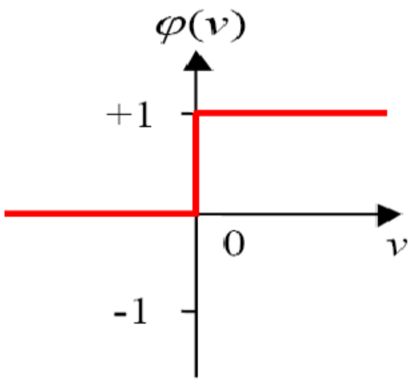




What is the simplest function which gives the output of either 1 or 0?

The Step Function

$$y = \varphi\left(\sum_{i=1}^{m} x_i - b\right)$$



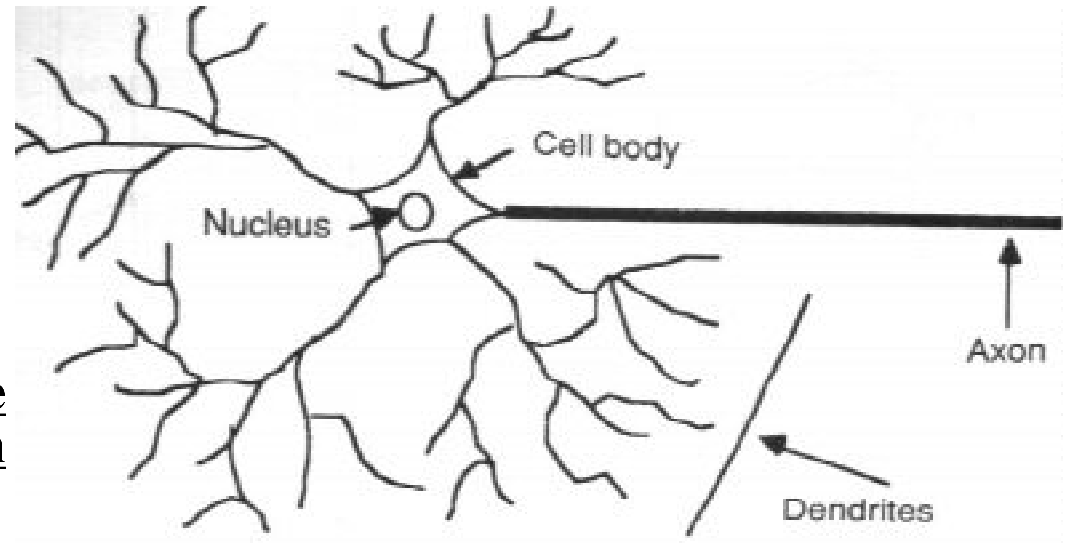
Why do we need the threshold (bias) b?

The neuron will not fire till it is "high" enough!

Is the step function linear or nonlinear? Nonlinear!

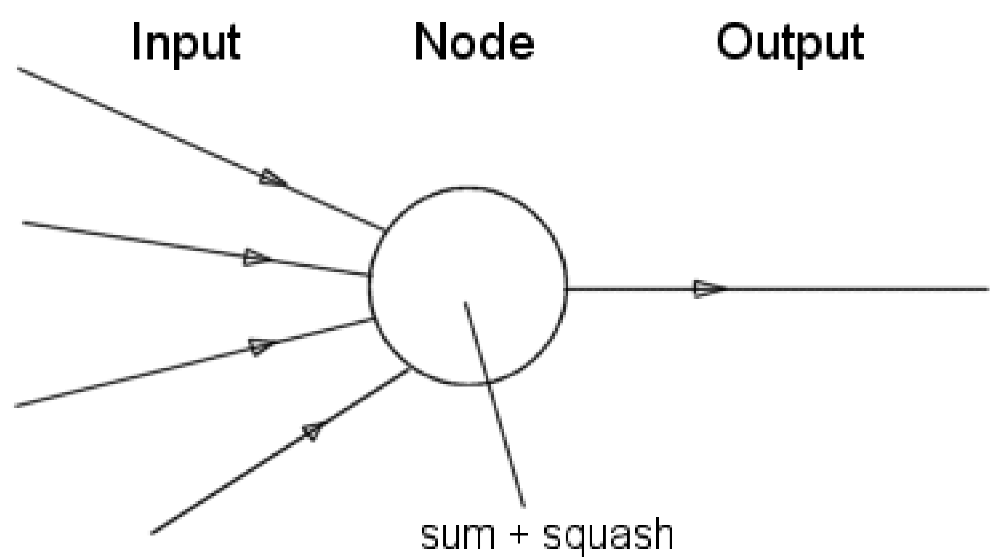
The nonlinear model of the neuron

$$y = \varphi\left(\sum_{i=1}^{m} x_i - b\right)$$



Based upon this model, is it possible for the inputs to inhibit the activation of the neuron?

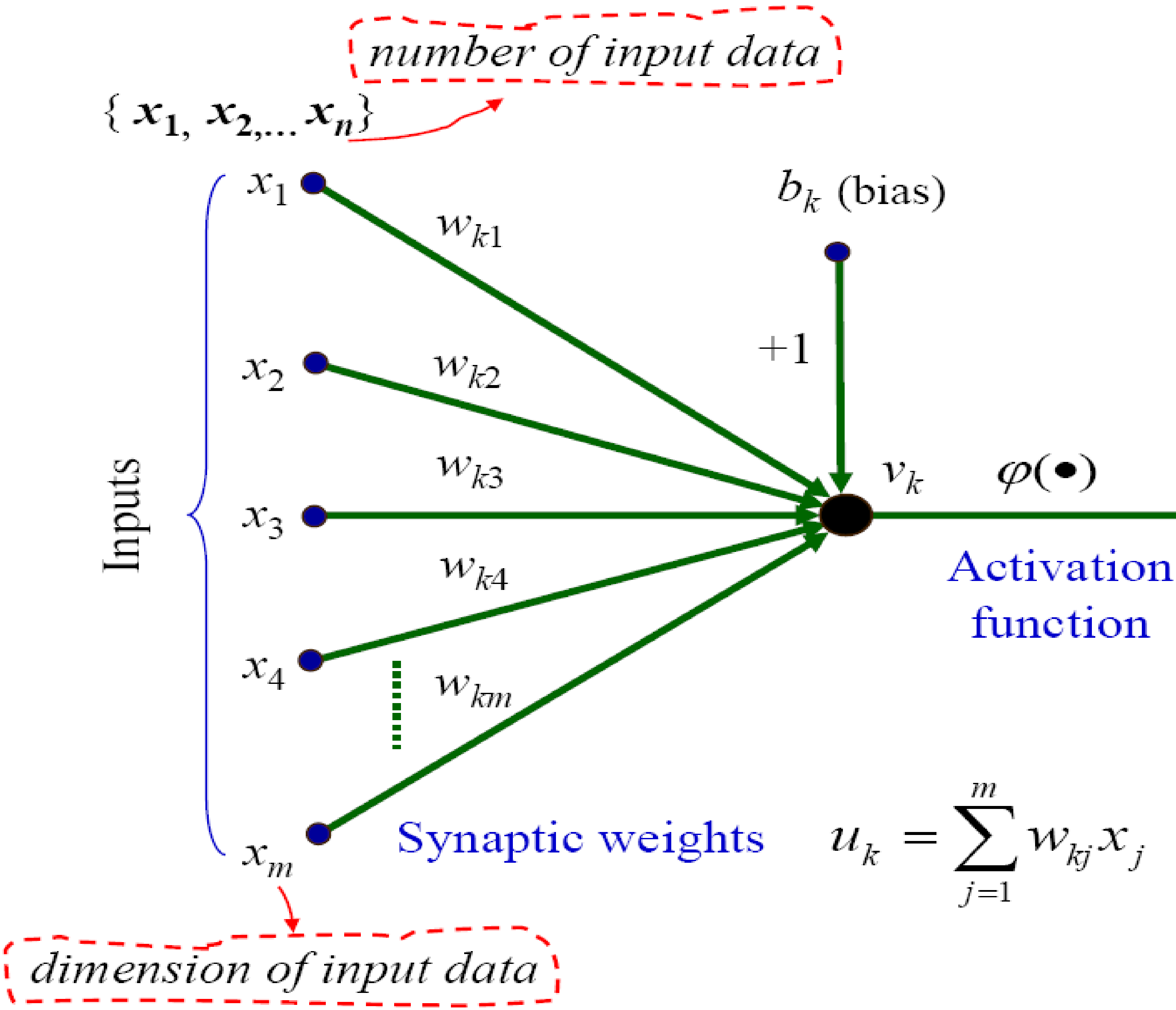
They can only excite the neuron!



How about inhibition?
How to model both excitation and inhibition?

The synaptic weights!

Model of a Neuron *k*



$$u_k = \sum_{j=1}^{m} w_{kj} x_j \qquad y_k = \varphi(u_k + b_k)$$

The era of artificial neural networks started with

Mathematical Model of a Neuron

Three basic components for the model of a neuron:

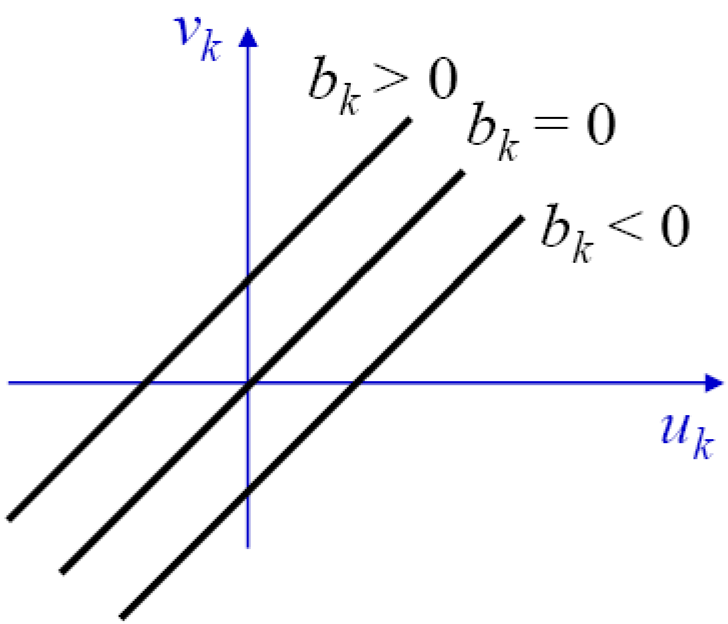
1. A set of synapses or connecting links: characterized by a weight or strength of its own.
2. An adder for summing the input signals, weighted by the respective synapses of the neuron (a linear combiner).
3. An activation function for limiting the amplitude of the neuron output, e.g., it limits the permissible amplitude range of the output signal, typically [0, 1] or [-1, 1].

Mathematically, for a neuron *k*,

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \quad \text{and} \quad y_k = \varphi(\overbrace{u_k + b_k}^{v_k})$$

Where $x_1, x_2, \ldots, x_m$ are the input signals; $w_{k1}, w_{k2}, \ldots, w_{km}$ are the synaptic weights of neuron *k*; u_k is the linear combiner output due to the input signals; b_k is the bias; $\varphi(\bullet)$ is the activation function; and y_k is the output signal of the neuron.

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \quad \text{and} \quad y_k = \varphi(\overbrace{u_k + b_k}^{v_k})$$



Definition: "induced local field"

$$v_k = u_k + b_k$$

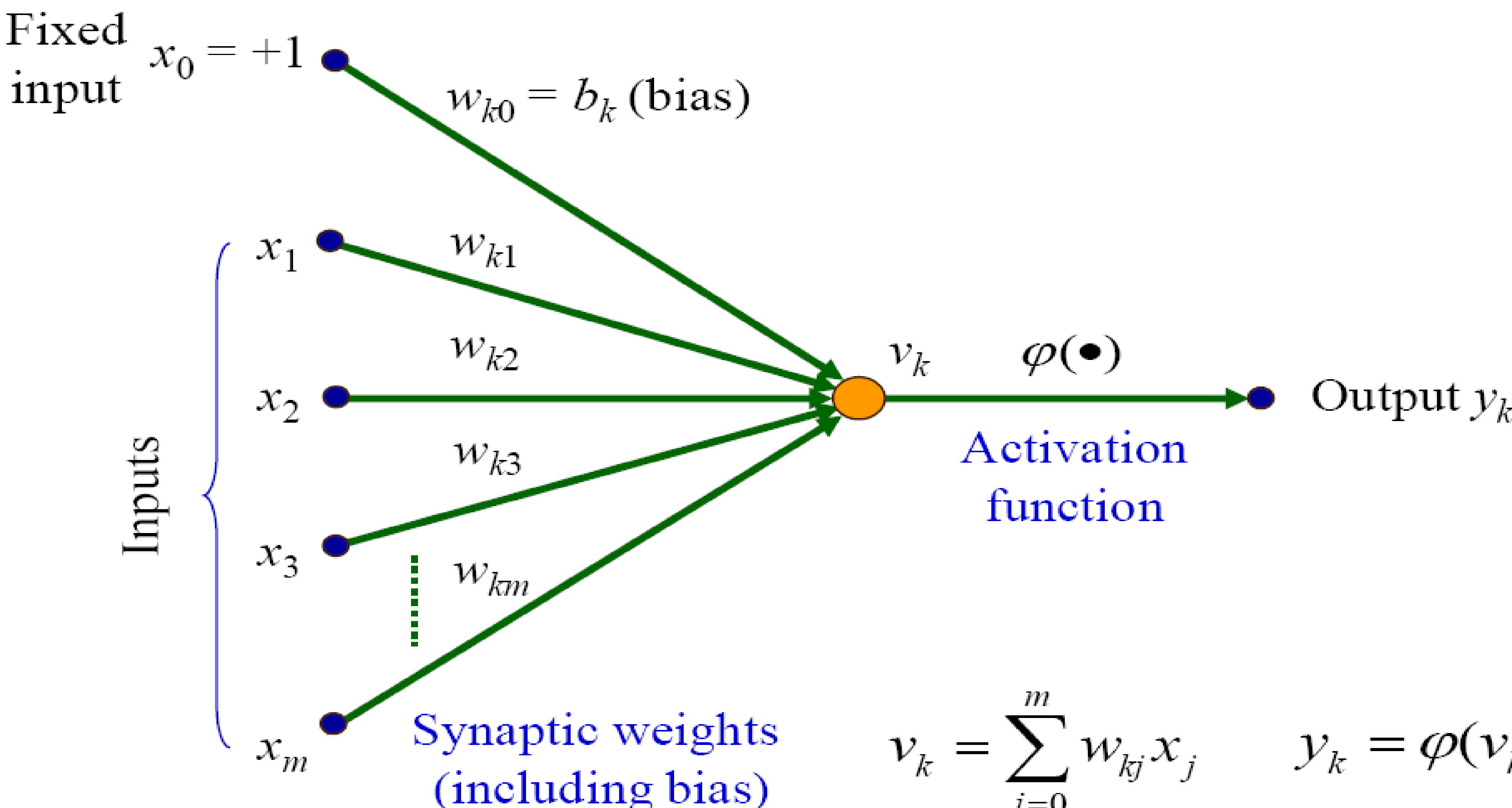
Potential induced by other neurons

Background potential

Alternatively, we may reformulate the model as follows:

$$v_k = \sum_{j=0}^{m} w_{kj} x_j \quad \text{and} \quad y_k = \varphi(v_k)$$

Note: we have added a new input: $x_0 = +1$ and its synapse weight is $w_{k0} = b_k$.



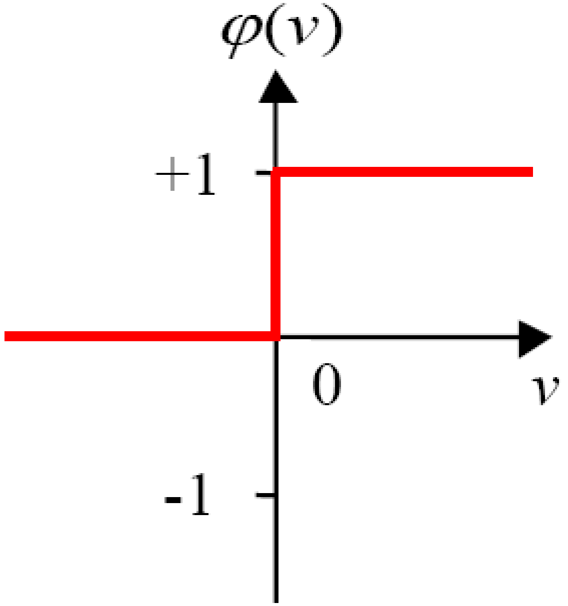
$$v_k = \sum_{j=0}^{m} w_{kj} x_j \qquad y_k = \varphi(v_k)$$

Type of Activation (Squash) Functions

Threshold function (hard limiter)

Note: McCulloch-Pitts model (1943) of neuron used this form of threshold function

$$\varphi(v) = \begin{cases} +1 & if \ v \geq 0 \\ 0 & if \ v < 0 \end{cases}$$



Is it continuous?

No.

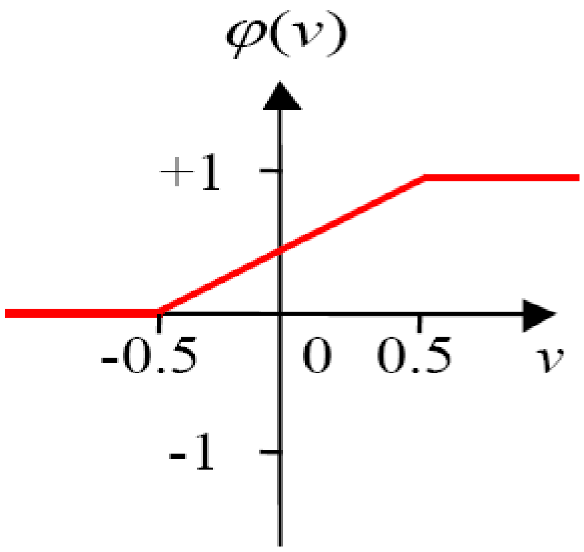
Does it distinguish between different excitation levels (firing rates) of the neuron?

No.

Type of Activation (Squash) Functions

Piecewise-linear function:

$$\varphi(v)=\begin{cases}1 & v\geq 0.5\\ v+0.5 & -0.5<v<0.5\\ 0 & v\leq -0.5\end{cases}$$



<u>Is it continuous?</u>

Yes.

<u>Is it differentiable (smooth)?</u>

No.

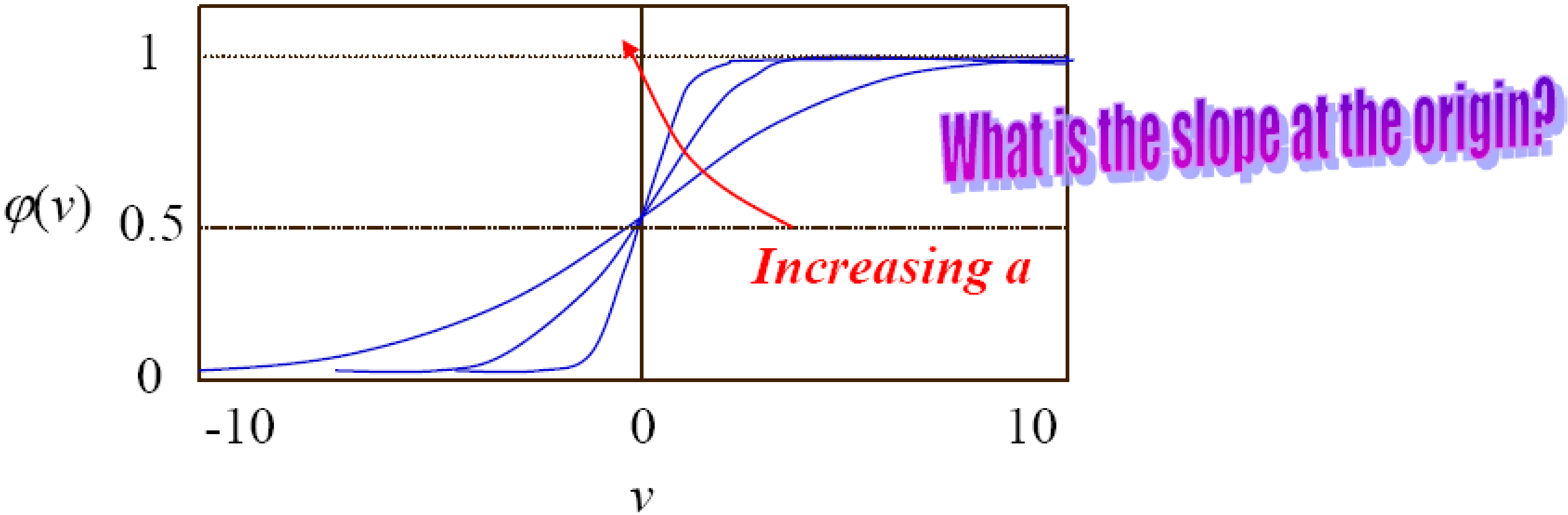
We may run into trouble if we try to compute the gradients (derivatives)!

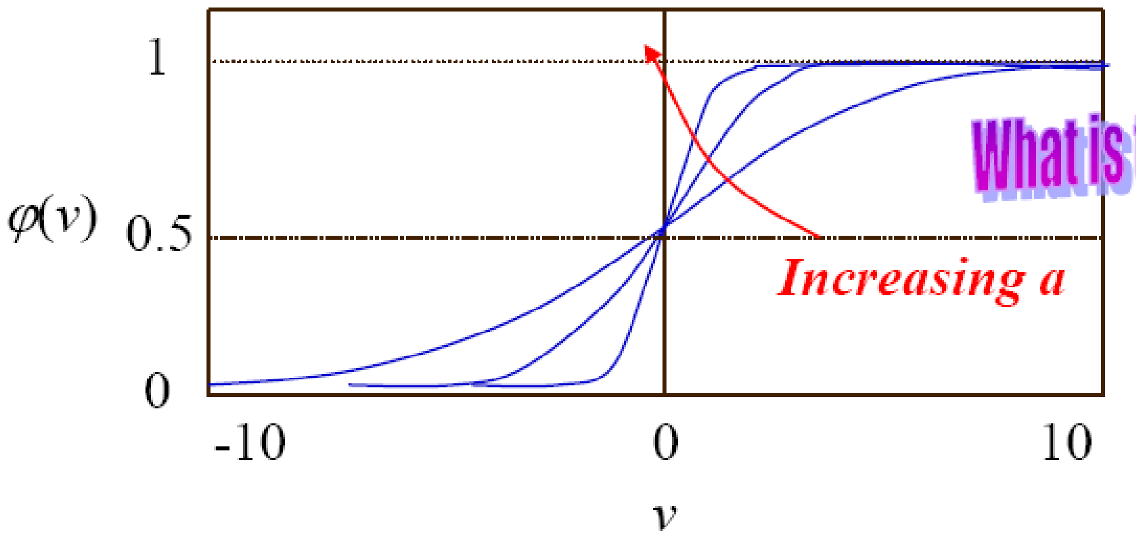
Sigmoid function (*s*-shaped)

- Most common
- Continuous & differentiable everywhere!
- Strictly increasing function
- Asymptotically approaches the saturation values
- Example: Logistic function

$$\varphi(v) = \frac{1}{1 + e^{-av}}$$

where a is the slope parameter.





What is the slope at the origin?

Solution:

$$\varphi(v) = \frac{1}{1+e^{-av}}$$

$$\varphi'(v) = \frac{ae^{-av}}{(1+e^{-av})^2}$$

$$= a\left(\frac{e^{-av}}{1+e^{-av}}\right)\frac{1}{1+e^{-av}}$$

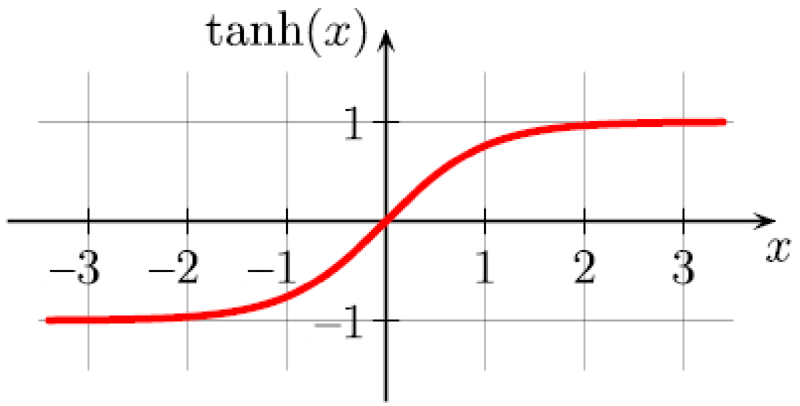
$$= a(1-\varphi(v))\varphi(v)$$

$$\boxed{\frac{d}{dx}\left(\frac{u}{v}\right) = \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2}}$$

So $\varphi'(0) = a(1\text{-}1/2)(1/2) = a/4$

- Logistic function assumes a continuous range of values from 0 to 1.
- As slope parameter approaches infinity, it becomes a threshold function.
- If the activation function range is required to be from –1 to +1, we may use hyperbolic tangent function:

$$\varphi(v) = \tanh(v) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$



In this model, the output can be both positive and negative.

Does negative value of the output make sense for biological neurons?

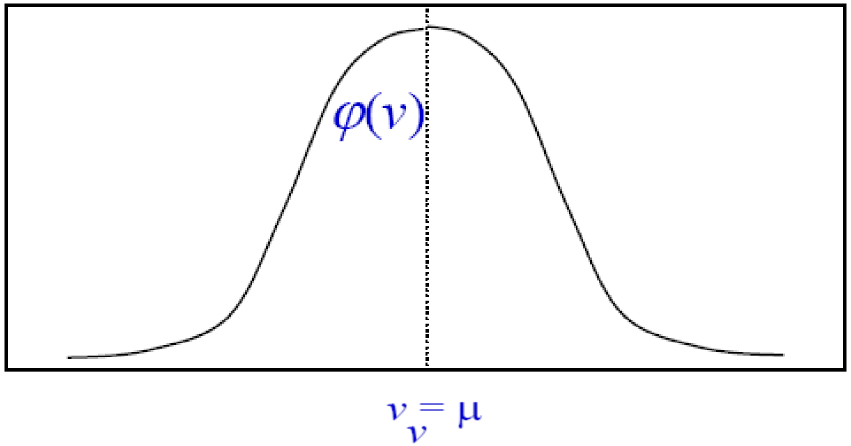
Biological neuron only fire “positive spikes”!

Although artificial neural networks originated from the biological neurons, they have gradually evolved into purely engineering tools, which may have no meaning for real biological neural networks at all!

Another example of activation function which is meaningless for real neurons!

Gaussian functions (Gaussian radial basis functions):

$$\varphi(v) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{v-\mu}{\sigma}\right)^2\right)$$

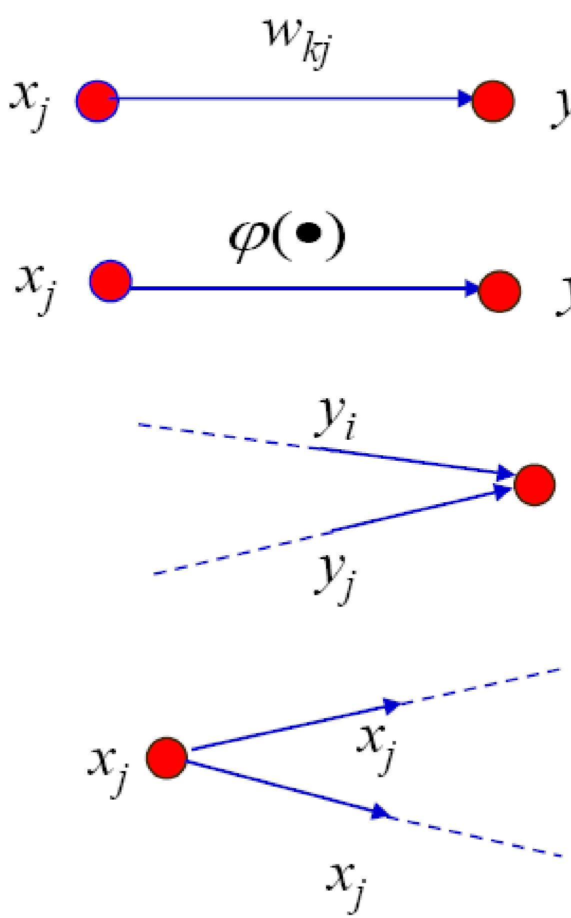


- Continuous and differentiable everywhere
- Asymptotically approaches 0 (or some constant) for $v \rightarrow \pm\infty$
- Single maximum at $v = \mu$
- Used in Radial Basis Function Networks

How do we model the connection of neurons?

Directed (Signal-Flow) Graphs

- A signal-flow graph is a network of directed lines that are interconnected at certain points called nodes.
- Signal flows are in the direction defined by the arrow of the directed link.



$y_k = w_{kj} x_j$ (Synaptic links)

$y_k = \varphi(x_j)$ (Activation links)

$y_k = y_i + y_j$ (Synaptic convergence)

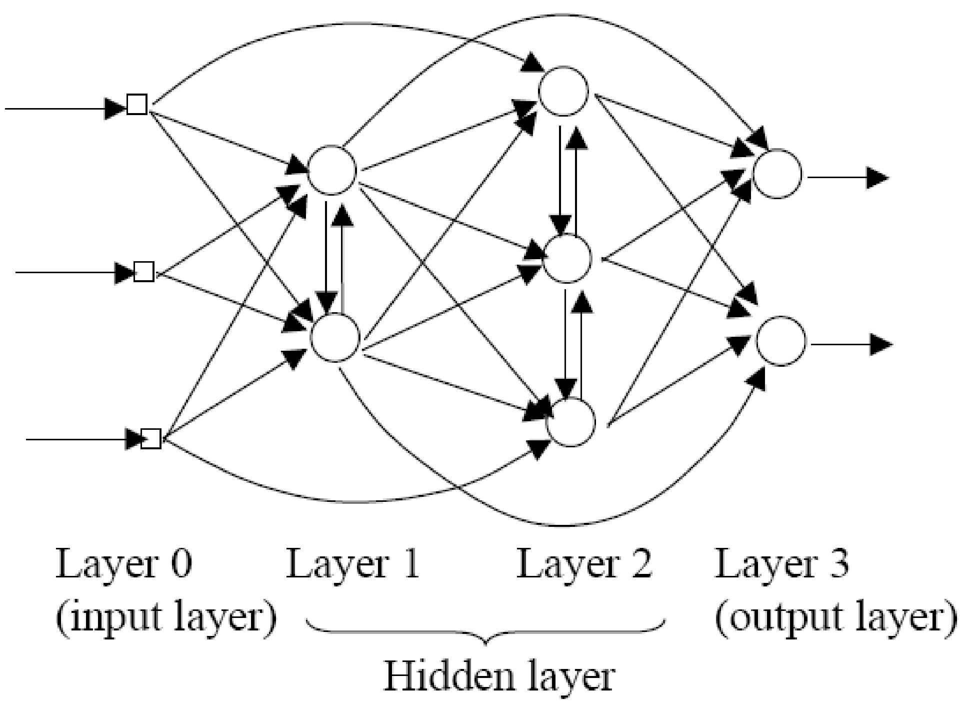
(Synaptic divergence)

Network Architectures

Single node is insufficient for many practical problems; networks with a large number of nodes are frequently used; Network architecture defines how nodes are connected;

Layered Feedforward Networks:

- Nodes are partitioned into subsets called layers;
- No connections lead from layer k to layer j if $k > j$;
- Intra-layer connections may exist.

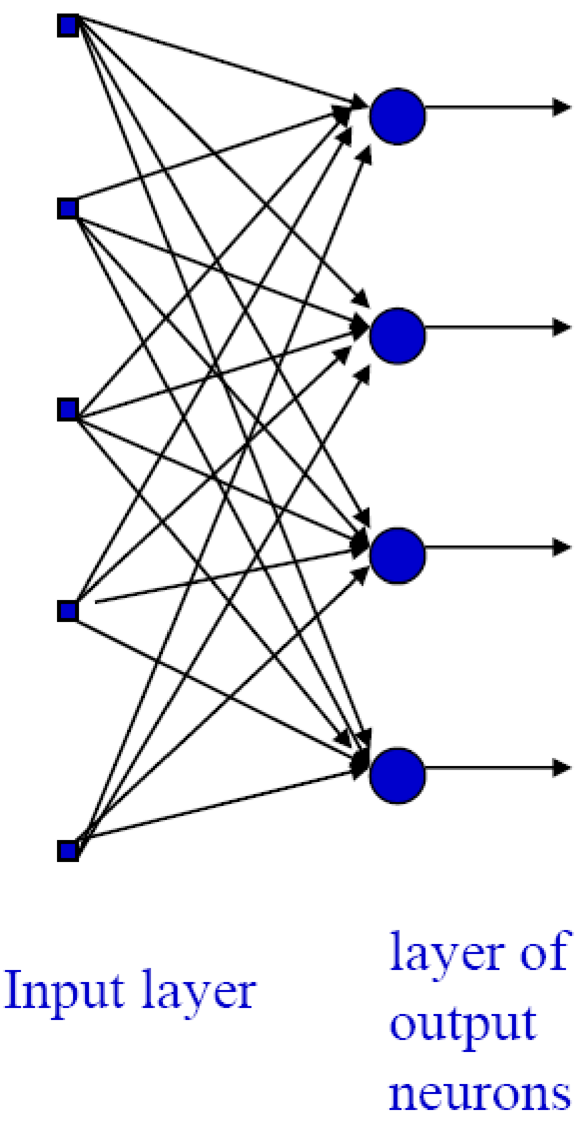


Single-Layer Feed-forward Networks

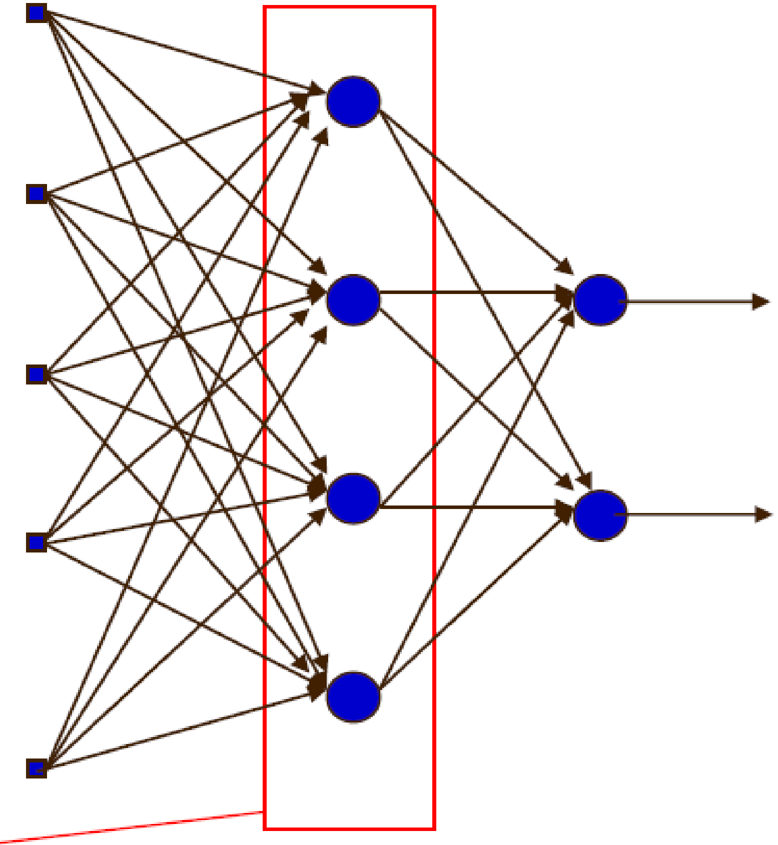
- **How many layers are there in the network?**

- Here, the layer refers to the output layer of computation nodes (neurons).

- Do not count the input layer of source nodes because no computation is performed there.

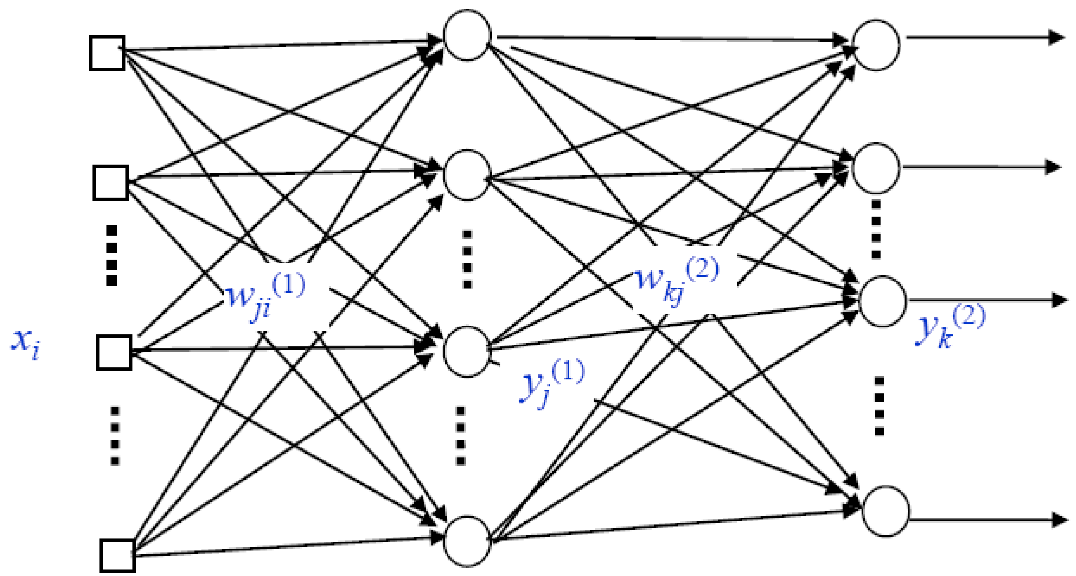


- **Multilayer Feedforward Networks**
  - Most common
  - Neurons in each layer of the network have as their inputs the output signals of the preceding layer only.



- One or more hidden layers whose computation nodes are called *hidden neurons.*
- Function of hidden neurons is *to intervene between the external input and the network output* in some useful manner.

Example: Consider a multilayer feedforward network, all the neurons of which operate in their linear regions. Justify the statement that such a network is equivalent to a single-layer feedforward network.



$\varphi(v) = v$

$$y_k^{(2)} = \sum_j w_{kj}^{(2)} y_j^{(1)} = \sum_j w_{kj}^{(2)} \left(\sum_i w_{ji}^{(1)} x_i \right)$$

$$y_k^{(2)} = \sum_j w_{kj}^{(2)} y_j^{(1)}$$

$$y_j^{(1)} = \sum_i w_{ji}^{(1)} x_i$$

$\Rightarrow$

$$= \sum_i \sum_j w_{kj}^{(2)} w_{ji}^{(1)} x_i$$

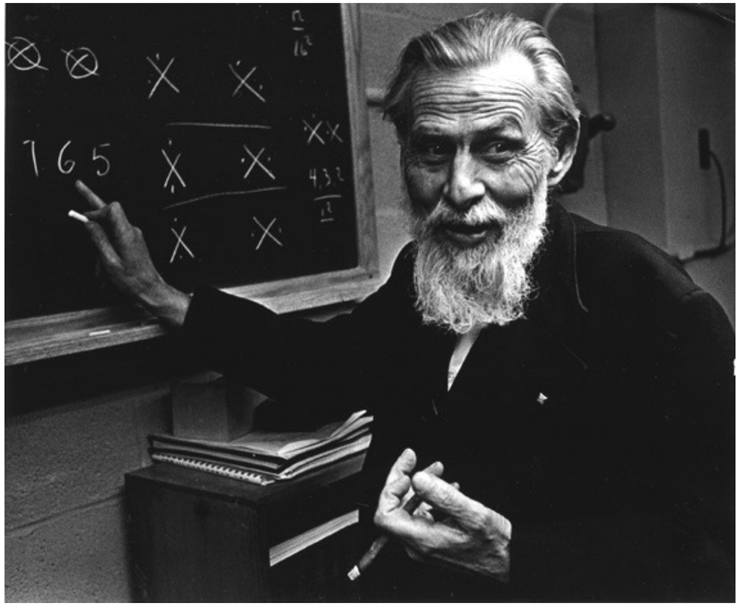
$$= \sum_i w_{ki} x_i \qquad \text{where} \qquad w_{ki} = \sum_j w_{kj}^{(2)} w_{ji}^{(1)}$$

(single-layer feedforward network)

The squash function (nonlinearity) is crucial for the success of neural networks! Without the squash function, the multilayer feedforward network would behave like a single layer feedforward network!

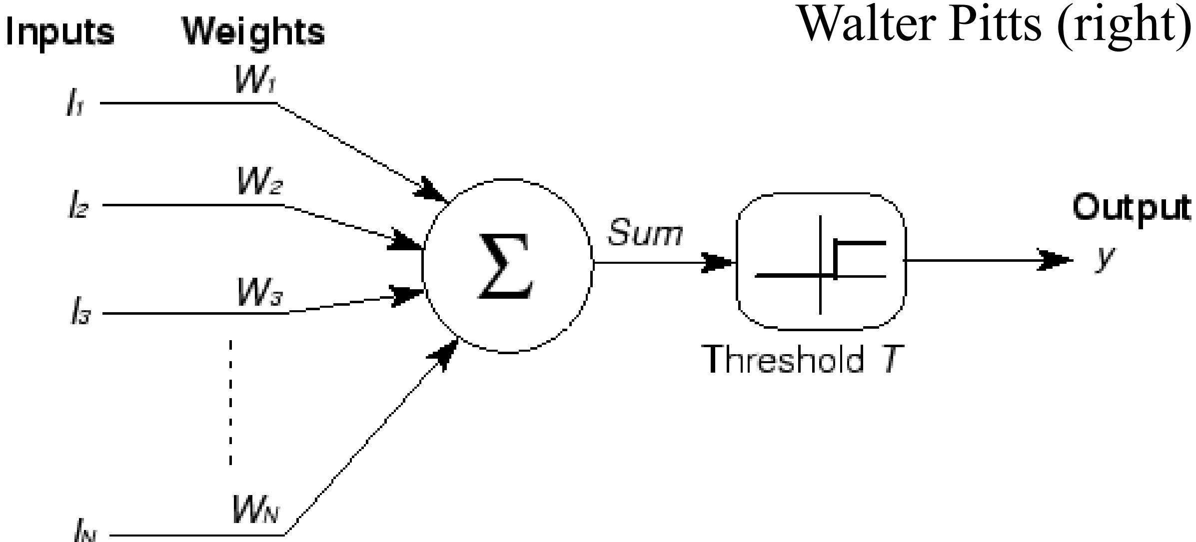
The beginning of the artificial neural networks

McCulloch and Pitts, 1943



Warren Sturgis McCulloch (1898-1969)

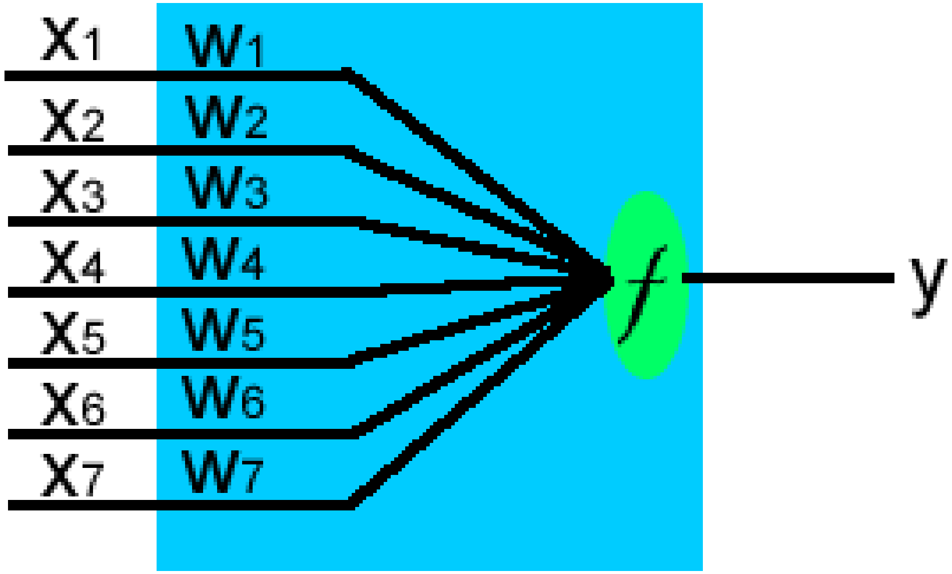
Walter Pitts (right) (1923-1969)



The next major step: Perceptron—single layer neural networks

Frank Rosenblatt, 1958

Frank Rosenblatt
(1928-1969)



Supervised learning:

$$w(k+1) = w(k) + a(y_d - y)x(k)$$

The weights were initially random. Then it could alter them so that it could be taught to perform certain simple tasks in pattern recognition.

Rosenblatt proved that for a certain class of problems, it could learn to behave correctly!

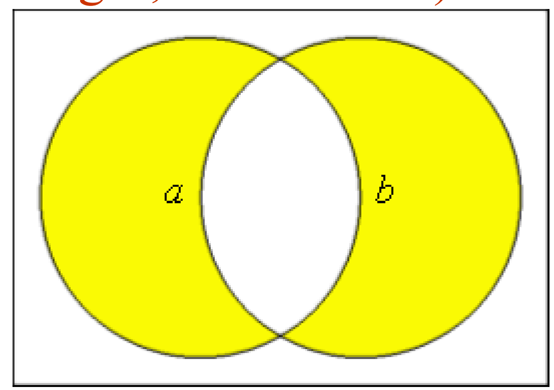
During 1960s', it seemed that neural networks could do anything.

The dark age of neural networks: 1969-1982

Marvin Minsky and Seymour Papert, 1969

Marvin Minsky
(1927-2016)

They proved that the perceptron could not execute simple logic like "exclusive OR" (i.e., apples OR oranges, but not both)

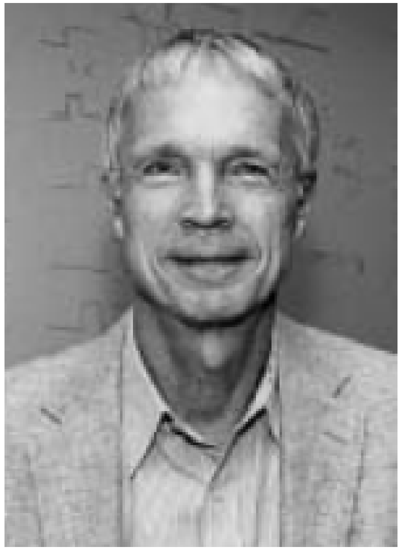


This killed interest in perceptrons for over 10 years! Rosenblatt died in a boating accident in 1969.

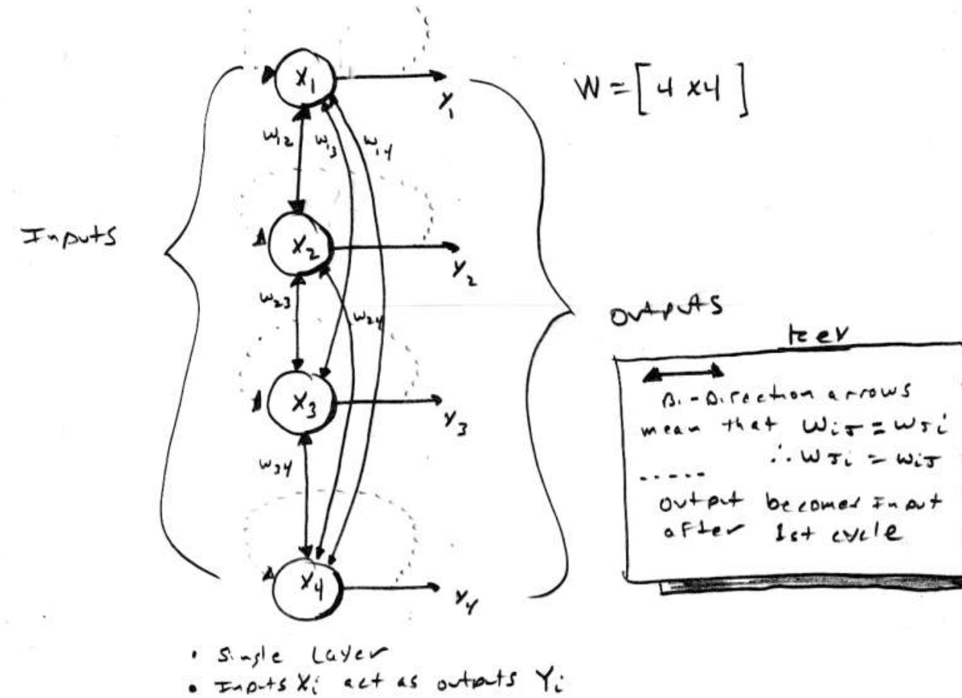
Minsky and Papert would have contributed more if they had produced a solution to this problem rather than beating the perceptron to death.

Three years later Stephen Grossberg proposed networks capable of modelling differential, contrast-enhancing and XOR functions. Nevertheless the often-cited Minsky/Papert text caused a significant decline in interest and funding of neural network research.

Then came the Hopfield Network which revitalized this area
John Hopfield, 1982



John Hopfield
(1933-present)



It is a simple network that feeds back on itself -- recurrent network

Hopfield nets serve as content-addressable memory systems with binary threshold units.

If the corrupted pattern is presented to the network, it will, after running around a few times, settle down to the whole pattern. demo

Why is such kind of memory called "content addressable"?

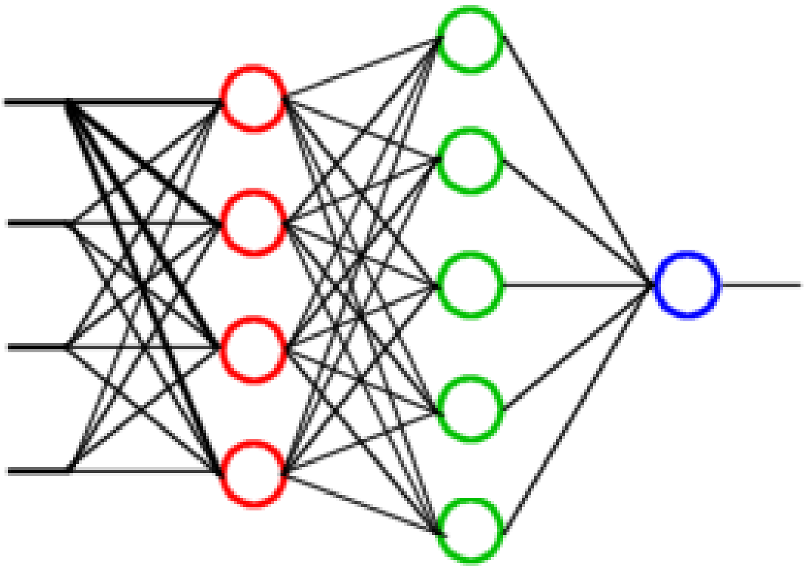
For digital computer, there is always an address to store and retrieve the information.

Is there any such unique address in the Hopfield net?

Any appreciable part of the input pattern will act as an address. This begins to have some faint resemblance to human memory, which draws lots of attention.

Multilayer Perceptron (MLP) and Back Propagation Algorithm
David Rumelhart and his colleagues, 1986

David Rumelhart (1942-2011)



Paul Werbos (1947-present)

It was proved later that MLP is a universal approximator, which can approximate any continuous function. It can solve the XOR problem easily.

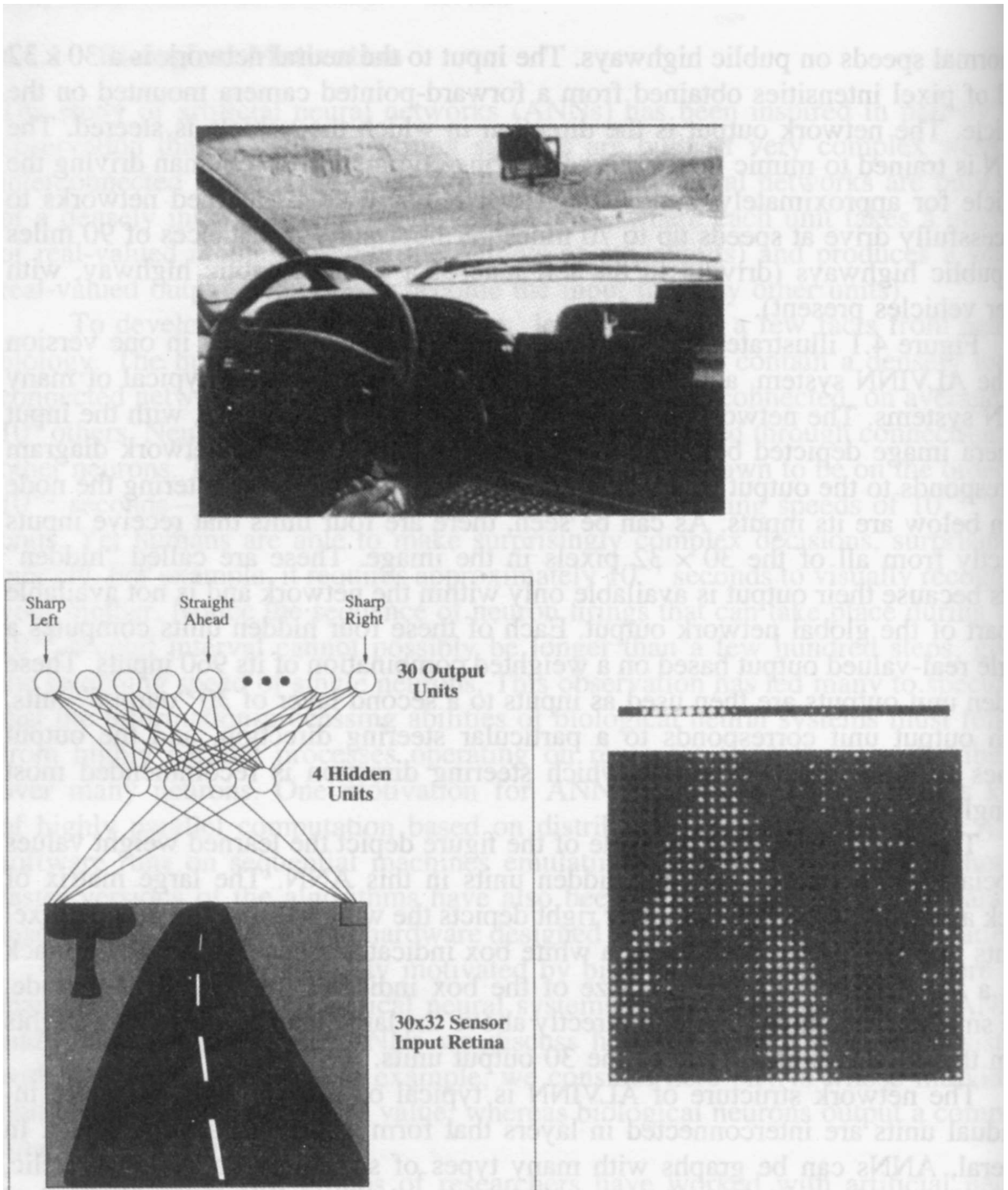
The Back Propagation Algorithm can be easily implemented to adjust the synaptic weights.

After the BP was publicized in the 1980s. It turned out that the BP was described in the Ph.D. thesis of Paul Werbos in 1974 at Harvard University.

A lot of research since the late 1980's.

NNs have now been used successfully in many areas and applications.

"Autonomous Land Vehicle In a Neural Network" (ALVINN)
by Dean Pomerleau and Todd Jochem, CMU



ALVINN was successfully trained by human beings to drive.

The teacher was simply driving the vehicle while the neural networks were learning by back propagation.

On a highway north of Pittsburgh, ALVINN successfully drove autonomously for distances of over 90 miles (150 km) and reached speeds of up to 70 mph, (117km/h)

Does the biological neural net also learn in a similar fashion?

What is Neural Network (NN)?

A neural network is a *massively parallel distributed processor* made up of simple processing unit, which has a natural propensity for *storing experiential knowledge* and making it available for use.

It employs a massive inter-connection of "simple" computing units - *neurons*.

It is capable of organizing its structure consists of many neurons, to perform tasks that are many times faster than the fastest digital computers nowadays.

Knowledge is obtained from the data/input signals provided.

Knowledge is learned by adjusting the synapses!

What is Artificial Neural Network (ANN)?

It resembles the brain in two respects:

1. Knowledge is acquired by the network through a *learning process*.
2. Inter-neuron connection strengths known as *synaptic weights* are used to store the knowledge.

Artificial neural networks are parameterized *computational nonlinear algorithms* for (numerical) data/signal/image processing. These algorithms are either implemented on a general-purpose computer or are built into a dedicated hardware.

Benefits of Neural Networks

High computational power

1. *Generalization*: Producing reasonable outputs for inputs not encountered during training (learning).

2. *Has a massively parallel distributed structure.*

Useful properties and capabilities

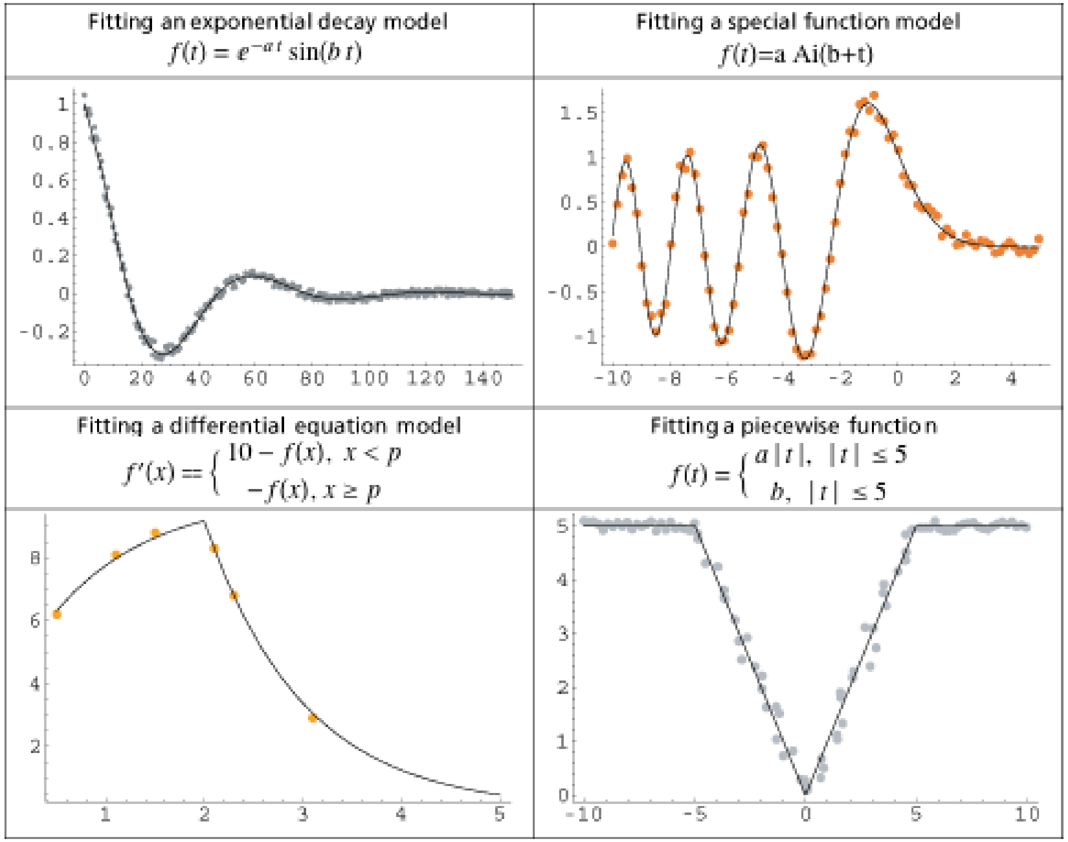
1. *Adaptivity (plasticity)*: Has built-in capability to adapt their synaptic weights to changes in the environment.
2. *Fault tolerance*: If a neuron or its connecting links are damaged, the overall response may still be ok (due to the distributed nature of information stored in a network).

Applications of NNs

NNs are mainly used for solving two types of problems:

Pattern Recognition (Pattern Classification)

Regression (Data Fitting, Function Approximation)



<u>Can you fit the data without knowing the mathematical form of the models?</u>

The neural networks can fit the data without any knowledge of the models!

Many real-world examples at http://www.ics.uci.edu/~mlearn/MLRepository.html

Course Outline

Part I:

1. **What are neural networks and why? (introduction)**
2. **Single Layer Perceptron (chapters 1,2,3)**
3. **Multilayer Perceptron (chapter 4)**
4. **Radial-Basis Function Networks (chapter 5)**
5. **Self-Organizing Networks (chapter 9)**

Part II (**by Dr Peter Chen (ME))** ***:***

Support Vector Machines

and

Reinforcement Learning

Lecture notes will be provided separately for part II.

Q & A…

THANK YOU !