INTRODUCTION TO ARTIFICIAL NEURAL NETWORKS

Source: Haykin, 2009

What is the motivation to study these networks?
Answer: Humans outperform the most advanced computers in such tasks as:

- Image recognition and processing {100-200 ms}
- Voice recognition
- Pattern recognition
- Motor control
- Heuristic optimization (examples: manual material handling, routing)

Computers are good at tasks based on precise and fast arithmetic operations, outperforming biological systems (Churchland, 1986)

What is so unique about biological systems?
Answer: "Method" of Information Processing

Biological Systems:
Massively parallel
Analog in nature
Adaptive and flexible
10^{-16} joules/operation/sec

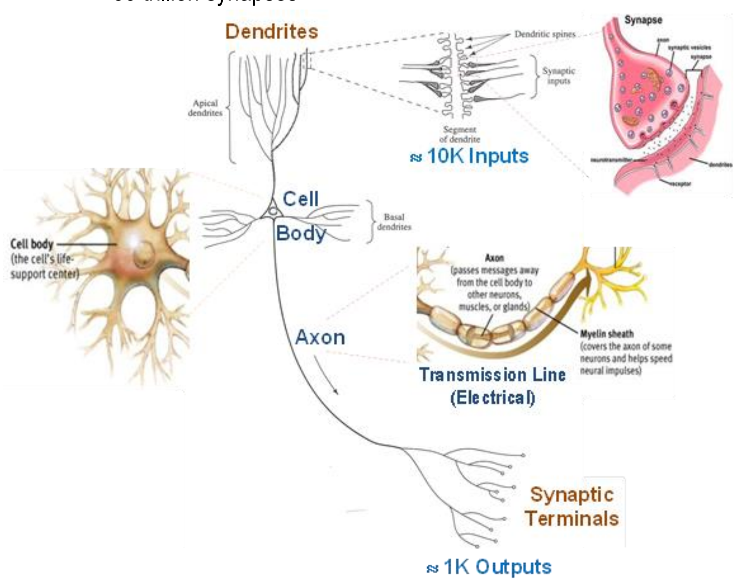
Most Computers: (Von Neuman machines)
Serial machines
Digital processing
10^{-6} joules/operation/sec (Foggin, 91)

Research efforts have been made in the last several decades to imitate or reproduce the flexibility and power of the human brain by artificial means. (Founding father: Raymon Cajal, 1911)
Results:

- Advances have been made in applying such systems (artificial neural networks) for problems found intractable or difficult for traditional computation. Examples: Approximators, classifiers, restoration of patterns, discriminators, optimization
- Artificial neural networks can supplement the enormous processing power of the Von Neuman digital computer
- This is just the tip of the ice-berg
  - Most ANN's are being simulated on serial machines
  - 1000 or 2000 nodes in comparison with 10 billion neurons and 60 trillion synapses
  - Connectivity
  - Complexity of the neuron

Basic Operation of a Real Neuron (Source: Dr. Kevin Gurney, Psychology Department, University of Sheffield):

- Signals are transmitted between neurons by electrical pulses (action-potentials or `spike' trains) traveling along the axon.
- These pulses impinge on the downstream (afferent) neuron at terminals called synapses.
- Synapses are found principally on a set of branching processes emerging from the cell body (soma) known as dendrites.
- Each pulse occurring at a synapse initiates the release of a small amount of chemical substance or neurotransmitter which travels across the synaptic cleft and which is then received at post-synaptic receptor sites on the dendritic side of the synapse.
- The neurotransmitter becomes bound to molecular sites, which, in turn, initiates a change in the dendritic membrane potential.
- This post-synaptic-potential (PSP) change may serve to increase (hyperpolarize) or decrease (depolarize) the polarization of the post-synaptic membrane. In the former case, the PSP tends to inhibit generation of pulses in the downstream (afferent) neuron, while in the latter, it tends to excite the generation of pulses.
- The size and type of PSP produced will depend on factors such as the geometry of the synapse and the type of neurotransmitter.
- Each PSP will travel along its dendrite and spread over the soma, eventually reaching the base of the axon (axon-hillock).
- The downstream (afferent) neuron sums or integrates the effects of thousands of such PSPs over its dendritic tree and over time.
- If the integrated potential at the axon-hillock exceeds a threshold, the cell `fires' and generates an action potential or spike which starts to travel along its axon. This then initiates the whole sequence of events again in neurons contained in the outward (efferent) pathway.
- Cells come in different sizes/shapes and exhibit plasticity
- Human cortex is estimated to carry ~10 billion neurons and ~60 trillion synapses



Pyramidal Neuron Cell

Characteristics of most ANN's:

- Function as parallel distributed computing networks
- Most basic characteristic is their architecture
- Provide either:
  - Instantaneous responses
  - Time-domain behavior (need time to respond), dynamic
- Resort to different learning modes – different learning rules and modes
- Exhibit different speeds, efficiency of learning, and effectiveness
- Contrast to conventional computers (which are programmed to perform specific tasks) most NN's must be taught or trained (supervised learning or unsupervised learning)
- Learn new associations, new patterns, functional dependencies
- Different types of neuron models are popular
- Acquires knowledge (or) assembles information that can be later recalled

ANN's have attracted to attention of scientists and technologists from a number of disciplines

- Neuroscientists – interested in modeling biological neural networks
- Physicists – interested in analogies between neural network models and nonlinear dynamical systems
- Mathematicians – fascinated by the potential of mathematical modeling applied to complex large systems phenomena
- Psychologists – ANN's as possible prototype structures of human-like information processing
- Engineers – applications; control, signal processing, opportunities opened by massively parallel computational networks in the areas of AI computational theory, simulation ..

Neural Network Applications:

"It would be impossible to cover the total range of applications for which neural networks have provided outstanding solutions. The following table provides an idea of the diversity of applications for which neural networks provide **state-of-the-art** solutions."
– MathWorks, 2012 (http://www.mathworks.com/help/toolbox/nnet/gs/f9-36282.html)

Industry	Business Applications
Aerospace	High-performance aircraft autopilot, flight path simulation, aircraft control systems, autopilot enhancements, aircraft component simulation, and aircraft component fault detection
Automotive	Automobile automatic guidance system, and warranty activity analysis
Banking	Check and other document reading and credit application evaluation
Defense	Weapon steering, target tracking, object discrimination, facial recognition, new kinds of sensors, sonar, radar and image signal processing including data compression, feature extraction and noise suppression, and signal/image identification
Electronics	Code sequence prediction, integrated circuit chip layout, process control, chip failure analysis, machine vision, voice synthesis, and nonlinear modeling
Entertainment	Animation, special effects, and market forecasting
Financial	Real estate appraisal, loan advising, mortgage screening, corporate bond rating, credit-line use analysis, credit card activity tracking, portfolio trading program, corporate financial analysis, and currency price prediction
Industrial	Prediction of industrial processes, such as the output gases of furnaces, replacing complex and costly equipment used for this purpose in the past
Insurance	Policy application evaluation and product optimization
Manufacturing	Manufacturing process control, product design and analysis, process and machine diagnosis, real-time particle identification, visual quality inspection systems, beer testing, welding quality analysis, paper quality prediction, computer-chip quality analysis, analysis of grinding operations, chemical product design analysis, machine maintenance analysis, project bidding, planning and management, and dynamic modeling of chemical process system
Medical	Breast cancer cell analysis, EEG and ECG analysis, prosthesis design, optimization of transplant times, hospital expense reduction, hospital quality improvement, and emergency-room test advisement
Oil and gas	Exploration
Robotics	Trajectory control, forklift robot, manipulator controllers, and vision systems
Securities	Market analysis, automatic bond rating, and stock trading advisory systems
Speech	Speech recognition, speech compression, vowel classification, and text-to-speech synthesis
Telecommunications	Image and data compression, automated information services, real-time translation of spoken language, and customer payment processing systems
Transportation	Truck brake diagnosis systems, vehicle scheduling, and routing systems

Benefits of Artificial Neural Networks:

- Non-linearity
- Non-parametric
- Input-Output Mapping (or) Approximation
- Adaptivity – to surrounding environment while maintaining overall system "stability"
  – adaptive control } ideal for
  – adaptive signal processing } non-stationary
  – adaptive pattern classification } environment
  Adaptivity does not always lead to robustness; Example: an adaptive system with short time constants responds to spurious disturbances, causing a drastic degradation in performance
  Stability-Plasticity dilemma (Grossberg, 1988)
- Confidence intervals
- Fault Tolerance (or) graceful degradation of performance if implemented in hardware
- VLSI Implementability
- Uniformity of Analysis and Design

Definition of an Artificial Neural Network [Haykin, 1994]:

A neural network is a massively parallel distributed processor that has a natural property for storing experiential knowledge and making it available for use. To achieve good performance, a neural network employs a massive interconnection of simple computing cells referred to as "neurons."
It resembles the brain in two aspects:

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

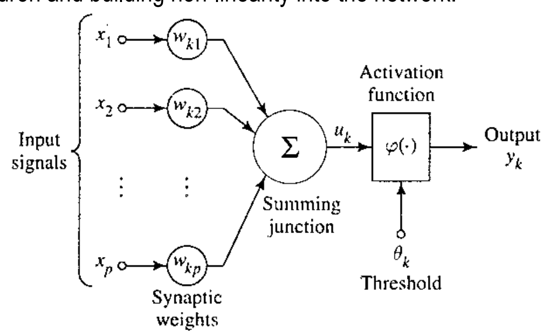
Learning Process:

The procedure used to perform the learning process is called a *learning algorithm*, the function of which is to modify the synaptic weights of the network in an orderly fashion so as to attain a desired design objective.

Models of a Neuron:

Most neuron models have three basic elements:

1. A set of *synapses* or connecting links, each of which is characterized by a *weight* or strength of its own. The weight is positive if the associated synapse is excitatory; it is negative if the synapse is inhibitory.
2. An *adder* for summing the input signals, weighted by the respective weights.
3. An *activation function* for limiting the amplitude of the output of a neuron and building non-linearity into the network.



Nonlinear model of a neuron.

Mathematical Terms:

Linear Combiner Output: $u_k = \sum_{j=1}^{p} w_{kj} x_j$

Activation Potential: $v_k = u_k + \theta_k$ or $v_k = \sum_{j=0}^{p} w_{kj} x_j$ where $x_0 = +1$ and $w_{k0} = \theta_k$

Neuron Output: $y_k = \varphi(v_k)$

Bias allows the net activation potential to be nonzero even if all the inputs are zeroes.

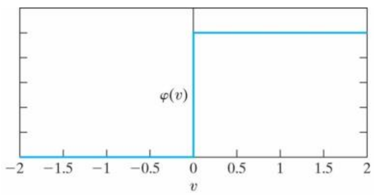
Activation Functions:

The activation function defines the output of a neuron in terms of the activity level at its input. Five popular types of activation functions are: 1) Threshold (Hard-Limiter) Function, 2) Piecewise-Linear Function, 3) Sigmoid Function, 4) Radial Basis Function, and 5) Softmax Function.

Threshold Functions

An example is:

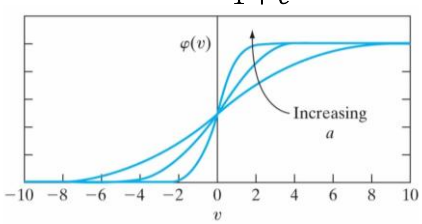
$$\varphi(\vartheta) = \begin{cases} 1 & \text{if } \vartheta \geq 0 \\ 0 & \text{if } \vartheta < 0 \end{cases}$$



Sigmoid Functions

Are functions that are differentiable and monotone. Exhibit smoothness and asymptotic properties. An example is the logistic function:

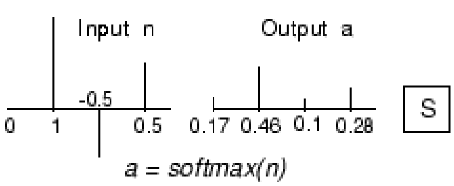
$$\varphi(\vartheta) = \frac{b}{1 + e^{-a\vartheta}}$$



Softmax Function

Soft-max neuron (or normalized exponential) is a generalization of the logistic function that "squashes" a *K*-dimensional vector of arbitrary real values to a *K*-dimensional vector of real values in the range (0,1) that add up to 1):

$$P(a_j|\boldsymbol{n}) = \frac{e^{n_j}}{\sum_{k=1}^{K} e^{n_k}} \text{ for } j = 1, \ldots, K$$



Softmax Transfer Function

Radial Basis Functions

A radial basis function (RBF) is a real-valued function whose value depends only on the distance from the origin, so that $\phi(\mathbf{x}) = \phi(\|\mathbf{x}\|)$; or alternatively on the distance from some other point c, called a center, so that $\phi(\mathbf{x}, \mathbf{c}) = \phi(\|\mathbf{x} - \mathbf{c}\|)$. Any function ϕ that satisfies the property $\phi(\mathbf{x}) = \phi(\|\mathbf{x}\|)$ is a radial function. The norm is usually Euclidean distance, although other distance functions are also possible. Commonly used examples include (writing $r = \|\mathbf{x} - \mathbf{x}_i\|$):

- Gaussian:

  The first term—that which is used for normalisation of the Gaussian—is missing because every Gaussian has a weight in our sum, thus the normalisation is not necessary.

  $$\phi(r) = e^{-(\varepsilon r)^2}$$

- Multiquadric:

  $$\phi(r) = \sqrt{1 + (\varepsilon r)^2}$$

- Inverse quadratic:

  $$\phi(r) = \frac{1}{1 + (\varepsilon r)^2}$$

- Inverse multiquadric:

  $$\phi(r) = \frac{1}{\sqrt{1 + (\varepsilon r)^2}}$$

- Polyharmonic spline:

  $$\phi(r) = r^k, \; k = 1, 3, 5, \ldots$$
  $$\phi(r) = r^k \ln(r), \; k = 2, 4, 6, \ldots$$

- Thin plate spline (a special polyharmonic spline):

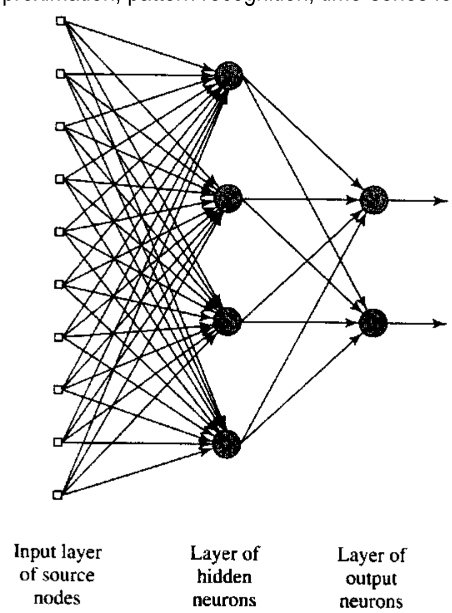
  $$\phi(r) = r^2 \ln(r)$$

Network Architectures:

The manner in which the neurons of a network are connected is referred to as network architecture. In general, there are two different classes of network architectures:

Feed-forward Networks

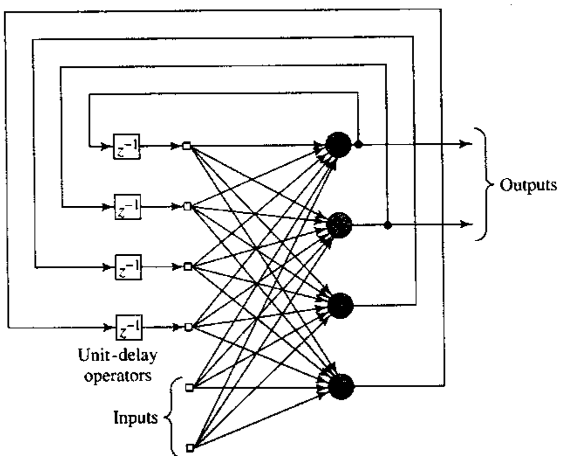
In these networks, the neurons are organized in the form of layers. Typically there is an input layer of source nodes, multiple hidden layers (with multiple neurons per layer), and an output layer of neurons. The information strictly flows forward from the input layer to the output layer. These networks are popular for applications such as: function approximation, pattern recognition, time-series forecasting etc.



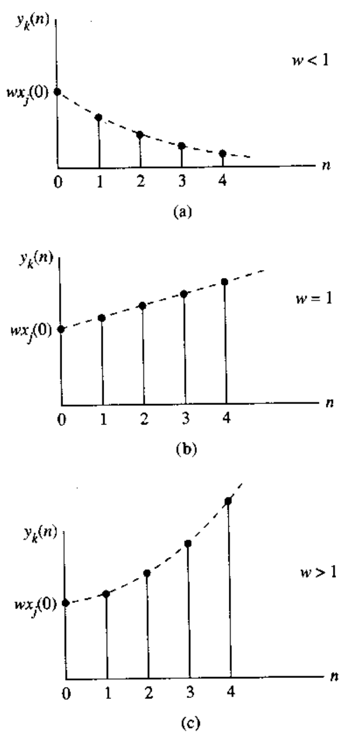
A feed-forward network.

Recurrent Networks

A recurrent network distinguishes itself from a feed-forward network in that it has at least one feedback loop. These networks have strong dynamic properties and are popular for applications such as: modeling dynamic systems, combinatorial optimization.



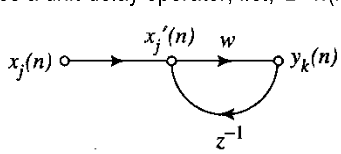
A recurrent network.
(z^{-1} denotes a unit-delay operator,
i.e., $z^{-1}x(n) = x(n-1)$)



Time response for 3 different values of weight *w*. (a) Stable with infinite memory. (b) Linear divergence. (c) Exponential divergence.

Feedback:

- Feedback is said to exist in a dynamic system when the output of an element in the system influences in part the input applied to that particular element.
- Feedback occurs in almost every part of the nervous system of every mammal (Freeman, 1975).
- Plays a major role in the study of recurrent networks.
- If we consider the following single-loop feedback system (here z^{-1} denotes a unit-delay operator, i.e., $z^{-1}x(n) = x(n-1)$):



$y_k(n) = w\left[x_j^{'}(n)\right]$ and $x_j^{'}(n) = x_j(n) + z^{-1}y_k(n)$

by eliminating $x_j^{'}(n)$, we get:

$$
\begin{aligned}
y_k(n) &= \frac{w}{1-wz^{-1}}\left[x_j(n)\right] \\
&= w\left(1-wz^{-1}\right)^{-1}\left[x_j(n)\right]
\end{aligned}
$$

using Binomial expansion, we get

$$
\begin{aligned}
y_k(n) &= w\sum_{l=0}^{\infty} w^l z^{-l}\left[x_j(n)\right] \\
&= \sum_{l=0}^{\infty} w^{l+1} x_j(n-l) \\
&\equiv \text{First-order, Infinite-duration Impulse Response Filter}
\end{aligned}
$$

Unfortunately, the analysis of the dynamic behavior of recurrent networks is complicated by virtue of the fact that the processing units, i.e. the neurons, are nonlinear.

Knowledge Representation in ANNs:

- Knowledge can be classified into two kinds of information:
  1. Known facts (*prior* information)
  2. Observations and measurement (inherently noisy)
- The subject of knowledge representation in ANNs is very complicated, and can only be addressed effectively one application at a time.
- In general, neural network design involves the following steps:
  1. Selection of appropriate architecture
  2. Collection of example patterns representative of the environment (both +ve and -ve examples)
  3. Training the network by means of a suitable algorithm
  4. Testing the network performance for generalization
- Nevertheless, there are four rules of knowledge representation that are of a general commonsense nature (in pattern classifier metaphor) [Anderson, 1988]:

  Rule I: Inputs from "similar" classes should usually produce similar representations inside the network.

  Rule II: Items from "separate" classes should be given widely different representations in the network.

  Rule III. Important features should be allocated a large number of neurons in the network for that feature's representation.

  Rule IV. *Prior information* and *invariances* should be built into the design of a neural network, thereby not having to learn them. Why? Results in fewer network parameters, a faster network, and a cheaper network.
  Image Recognition Example:
  - Prior Information – Two dimensional image, BW/Color, Critical Features, …
  - Invariance Requirements – Invariance to rotation, sharpness, brightness, …
- In general, prior information can be built into ANNs in two ways:
  1. Through network architecture restrictions
  2. Through synaptic weight constraints
- In general, invariances can be built into the ANNs in three ways:
  1. Through structure (architecture restrictions and synaptic weight constraints)
  2. Through training
  3. Through invariant feature space