Artificial Neural Network
Artificial Neural Network

The Mark I Perceptron patchboard
What is artificial neural networks (ANNs)?

- ANNs are biologically inspired; that are composed of elements that perform in a manner that is analogous to most elementary functions of the biological neuron.
What is ANNs? (Cont.)

• A parallel, distributed information processing structure consisting of processing elements (which can possess a local memory and carry out localized information processing operations) interconnected together.
What can artificial neural networks do?

• **Learning**: ANNs can modify their behavior in response to their environment.

• **Generalization**: Once trained, a network's response can be, to a degree, insensitive to minor variations in its input.

• **Abstraction**: Some ANNs are capable of abstracting the essence of a set of inputs. In one sense, it has learned to produce something that it has never seen before!

• **Applicability**: It appears that they will, however, become the preferred technique for a large class of pattern-recognition tasks that conventional computers do poorly, if at all.
Brain VS Computer Processing

- **Processing Speed**: Milliseconds VS Nanoseconds.
- **Processing Order**: Massively parallel. VS serially.
- **Abundance and Complexity**: $10^{11}$ and $10^{14}$ of neurons operate in parallel in the brain at any given moment, each with between $10^3$ and $10^4$ abutting connections per neuron.
Brain VS Computer Processing (Cont.)

- **Knowledge Storage**: Adaptable VS New information destroys old information.

- **Fault Tolerance**: Knowledge is retained through the redundant, distributed encoding information VS the corruption of a conventional computer's memory is irretrievable and leads to failure as well.

- **Processing Control**: The brain is an anarchic system - do not have any specific area with dictatorial control!
Computer Processing VS ANNs

- Computer: based on algorithms and rules
Computer Processing VS ANNs

- ANNs: based on transformations
Traditional Classifier

Parameters estimated from training data

Input Symbols

Compute Matching Score

Stored Value & Select Maximum

Output

Most likely class
Neural Net Classifier

- Input Symbols
- Compute Matching Score
- Select & Enhance Maximum
- Adapt weights given both O/Ps and correct class
- Output
Where are *Neural Nets* being used?

- **Signal Processing:**
  - *Adaptive echo cancellation* (Widrow and Stearns)

- **Control:**
  - *Truck backer-upper* (Nguyen and Widrow)

- **Pattern Recognition:**
  - *Recognizing handwritten characters* (Le Cun et al.)
Where are Neural Nets being used? (Cont.)

• Medicine:
  – *Instant Physician* (Anderson)

• Speech Production:
  – *NETtalk* (Sejnowski and Rosenberg)

• Speech Recognition:
  – *Speaker-independent recognition* (Kohonen)

• Business:
  – *Mortgage assessment* (Collins, Ghosh, and Scofield)
Motivation for ANN Research

• Efficiency and ease of biological systems at some tasks
  • Derive meaningful information from complex images
  • Ability to adapt and learn from previous mistakes

• Parallelism

• Fault tolerance

• Generalization

• Abstraction
Neuron VS ANN
Overview of Neurocomputing

1946-late 1980s

Digital Computer: All information processing applications used a single basic approach: *programmed computing* (involves devising an algorithm and/or a set of rules for solving the problem and then correctly coding these in software)

If the required algorithmic procedure and/or set of rules are not known, they must be developed (time consuming and costly-if possible).
Overview of Neurocomputing

Present

Neurocomputing: A new approach to information processing that does not require algorithm or rule development.

Typically in areas such as sensor processing, pattern recognition, data analysis, and control (mostly the algorithms or rules are not known and/or too expensive, time consuming, or inconvenient to develop).

The primary information processing structures of interest in neurocomputing are neural networks
History of Neurocomputing

1943  Warren McCulloch and Walter Pitts showed that even simple types of NN could, in principle, compute any arithmetic or logical function.

1949  Donald Hebb wrote a book entitled "The organization of Behavior" which pursued the idea that classical psychological conditioning is ubiq1uitous in animals because it is a property of individual neurals. He proposed a specific learning law for the synapses of neurons.
**History of Neurocomputing**

1951  Marvin Minsky designed and constructed the first neurocomputer (the Snark). The Snark did operate successfully from a technical standpoint (adjusted its weights automatically), but it never actually carried out any particularly interesting information processing function.

1957–58  Frank Rosenblatt, Charles Wightman, and other developed the first successful neurocomputer (the Mark I Perceptron). Many people see Rosenblatt as the founder of neurocomputing as we know it today. He wrote a book "Principle of Neurodynamics"
History of Neurocomputing

1958- Bernard Widrow and Marcian E. Hoff developed a different type of neural network processing element called the ADALINE which was equipped with a powerful new learning law.

Mid 1960s
Neural network research had been proven to be a dead end.
History of Neurocomputing

1969 Perceptrons (Minsky and Papert)
Showed single layered networks to be incapable of solving problems that were not linearly separable

<table>
<thead>
<tr>
<th>N</th>
<th>Number of functions</th>
<th>Linear separable functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>256</td>
<td>104</td>
</tr>
<tr>
<td>4</td>
<td>65,536</td>
<td>1,882</td>
</tr>
<tr>
<td>5</td>
<td>4,300,000,000</td>
<td>94,572</td>
</tr>
<tr>
<td>6</td>
<td>18,000,000,000,000,000,000,000</td>
<td>5,028,134</td>
</tr>
</tbody>
</table>

Data from Windner (1960)
History of Neurocomputing


1983 Defense Advanced Research Projects Agency (DARPA) funded neurocomputing research.
# Neurocomputers

## II. Neurocomputers built to date

<table>
<thead>
<tr>
<th>Neurocomputer</th>
<th>Year Introduced</th>
<th>Technology</th>
<th>Number of processing elements</th>
<th>Number of connections</th>
<th>Number of network(s)</th>
<th>Connections per second</th>
<th>Developers</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>perceptron</td>
<td>1957</td>
<td>Electromechanical and electronic</td>
<td>8</td>
<td>512</td>
<td>1</td>
<td>$10^8$</td>
<td>Frank Rosenblatt, Charles Wightman, Cornell Aeronautical Laboratory</td>
<td>Experimental</td>
</tr>
<tr>
<td>Adaline/Madaline</td>
<td>1960/62</td>
<td>Electrochemical (now electronic)</td>
<td>$1/8$</td>
<td>16/128</td>
<td>1</td>
<td>$10^4$</td>
<td>Bernard Widrow, Stanford U.</td>
<td>Commercial</td>
</tr>
<tr>
<td>Electro-optic crossbar</td>
<td>1984</td>
<td>Electro-optic</td>
<td>32</td>
<td>$10^4$</td>
<td>1</td>
<td>$10^8$</td>
<td>Demetri Prallitis, California Inst. of Technology</td>
<td>Experimental</td>
</tr>
<tr>
<td>Mark III</td>
<td>1985</td>
<td>Electronic</td>
<td>$8 \times 10^3$</td>
<td>$4 \times 10^5$</td>
<td>1</td>
<td>$3 \times 10^4$</td>
<td>Robert Hecht-Nielsen, Todd Gutschow, Michael Myers, Robert Kuczewski, TRW</td>
<td>Commercial</td>
</tr>
<tr>
<td>Neural emulation processor</td>
<td>1985</td>
<td>Electronic</td>
<td>$4 \times 10^4$</td>
<td>$1.6 \times 10^8$</td>
<td>1</td>
<td>$4.9 \times 10^8$</td>
<td>Claude Cruz, IBM</td>
<td>Experimental</td>
</tr>
<tr>
<td>Optical resonator</td>
<td>1985</td>
<td>Optical</td>
<td>$6.4 \times 10^4$</td>
<td>$1.6 \times 10^8$</td>
<td>1</td>
<td>$1.6 \times 10^8$</td>
<td>Bernard Soffer, Yuri Owechko, Gilbert Dunning, Hughes Malibu Research Labs</td>
<td>Experimental</td>
</tr>
<tr>
<td>Mark IV</td>
<td>1986</td>
<td>Electronic</td>
<td>$2.5 \times 10^9$</td>
<td>$5 \times 10^9$</td>
<td>1</td>
<td>$5 \times 10^9$</td>
<td>Robert Hecht-Nielsen, Todd Gutschow, Michael Myers, Robert Kuczewski, TRW</td>
<td>Experimental</td>
</tr>
<tr>
<td>Odyssey</td>
<td>1986</td>
<td>Electronic</td>
<td>$8 \times 10^4$</td>
<td>$2.5 \times 10^9$</td>
<td>1</td>
<td>$2 \times 10^{10}$</td>
<td>Andrew Penc, Richard Wiggins, Texas Instruments Central Research Labs</td>
<td>Commercial</td>
</tr>
<tr>
<td>Crossbar chip</td>
<td>1986</td>
<td>Electronic</td>
<td>256</td>
<td>$6.4 \times 10^4$</td>
<td>1</td>
<td>$8 \times 10^9$</td>
<td>Larry Jackel, John Denker and others, AT&amp;T Bell Labs</td>
<td>Experimental</td>
</tr>
<tr>
<td>Optical novelty filter</td>
<td>1986</td>
<td>Optical</td>
<td>$1.6 \times 10^4$</td>
<td>$2 \times 10^8$</td>
<td>1</td>
<td>$2 \times 10^9$</td>
<td>Dana Anderson, U. of Colorado</td>
<td>Experimental</td>
</tr>
<tr>
<td>Anza</td>
<td>1987</td>
<td>Electronic</td>
<td>$3 \times 10^5$</td>
<td>$5 \times 10^8$</td>
<td>No limit</td>
<td>$2.5 \times 10^9$ (1.4 \times 10^9)</td>
<td>Robert Hecht-Nielsen, Todd Gutschow, Hecht-Nielsen Neurocomputer Corp.</td>
<td>Commercial</td>
</tr>
<tr>
<td>Parilen 2</td>
<td>1987</td>
<td>Electronic</td>
<td>$10^4$</td>
<td>$5.2 \times 10^9$</td>
<td>No limit</td>
<td>$1.5 \times 10^8$ (3 \times 10^8)</td>
<td>Sam Bogoch, Oren Clark, Iain Bason, Human Devices</td>
<td>Commercial</td>
</tr>
<tr>
<td>Parilen 2x</td>
<td>1987</td>
<td>Electronic</td>
<td>$9.1 \times 10^4$</td>
<td>$3 \times 10^8$</td>
<td>No limit</td>
<td>$1.5 \times 10^8$ (3 \times 10^8)</td>
<td>Richard Kasbo, Science Applications Int'l Corp.</td>
<td>Commercial</td>
</tr>
<tr>
<td>Delta floating-point processor</td>
<td>1987</td>
<td>Electronic</td>
<td>$10^4$</td>
<td>$10^8$</td>
<td>No limit</td>
<td>$2 \times 10^9$ (10^9)</td>
<td>George A. Works, William L. Hicks, Stephen Deiss, Richard Kasbo, Science Applications Int'l Corp.</td>
<td>Commercial</td>
</tr>
<tr>
<td>Anza plus</td>
<td>1988</td>
<td>Electronic</td>
<td>$10^4$</td>
<td>$1.5 \times 10^8$</td>
<td>No limit</td>
<td>$1.5 \times 10^9$ (6 \times 10^9)</td>
<td>Robert Hecht-Nielsen, Todd Gutschow, Hecht-Nielsen Neurocomputer Corp.</td>
<td>Commercial</td>
</tr>
</tbody>
</table>
# Best Known NNs

<table>
<thead>
<tr>
<th>Name of Network</th>
<th>Inventors and developers</th>
<th>Years Introduced</th>
<th>Primary applications</th>
<th>Limitations</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive resonance theory</td>
<td>Gall Carpenter, Northeastern U.; Stephen Grossberg, Boston U.</td>
<td>1976–80</td>
<td>Pattern recognition, especially when pattern is complicated or unfamiliar to humans (radar or sonar readouts, voiceprints)</td>
<td>Sensitive to translation, distortion, changes in scale</td>
<td>Very sophisticated; not yet applied to many problems</td>
</tr>
<tr>
<td>Avalanche</td>
<td>Stephen Grossberg, Boston U.</td>
<td>1967</td>
<td>Continuous-speech recognition; teaching motor commands to robotic arms</td>
<td>Literal playback of motor sequences—no simple way to alter speed or interpolate movements</td>
<td>Class of networks—no single network can do all these tasks</td>
</tr>
<tr>
<td>Back propagation</td>
<td>Paul Werbos, Harvard U.; David Parker, Stanford U.; David Rumelhart, Stanford U.</td>
<td>1974–85</td>
<td>Speech synthesis from text; adaptive control of robotic arms; scoring of bank loan applications</td>
<td>Supervised training only—correct input-output examples must be abundant</td>
<td>The most popular network today—works well, simple to learn</td>
</tr>
<tr>
<td>Bidirectional associative memory</td>
<td>Bart Kosko, U. of Southern California</td>
<td>1985</td>
<td>Content-addressable associative memory</td>
<td>Low storage density; data must be properly coded</td>
<td>Easiest network to learn—good educational tool; associates fragmented pairs of objects with complete pairs</td>
</tr>
<tr>
<td>Boltzmann and Cauchy machines</td>
<td>Jeffrey Hinton, U. of Toronto; Terry Sejnowsky, Johns Hopkins U.; Harold Szu, Naval Research Lab</td>
<td>1985–6</td>
<td>Pattern recognition for images, sonar, radar</td>
<td>Boltzmann machine: long training time. Cauchy machine: generating noise in proper statistical distribution</td>
<td>Simple networks in which noise function is used to find a global minimum</td>
</tr>
<tr>
<td>Model</td>
<td>Year</td>
<td>Description</td>
<td>Applications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brain state in a box</td>
<td>1977</td>
<td>Extraction of knowledge from data bases</td>
<td>One-shot decision making—no iterative reasoning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carebelltron</td>
<td>1969-82</td>
<td>Controlling motor action of robotic arms</td>
<td>Requires complicated control input</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Counterpropagation</td>
<td>1986</td>
<td>Image compression; statistical analysis; loan application scoring</td>
<td>Large number of processing elements and connections required for high accuracy for any size of problem</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hopfield</td>
<td>1982</td>
<td>Retrieval of complete data or images from fragments</td>
<td>Does not learn—weights must be set in advance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Madaline</td>
<td>1960-62</td>
<td>Adaptive nulling of radar jammers; adaptive modems; adaptive equalizers (echo cancellers) in telephone lines</td>
<td>Assumes a linear relationship between input and output</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neocognitron</td>
<td>1978-84</td>
<td>Handprinted-character recognition</td>
<td>Requires unusually large number of processing elements and connections</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceptron</td>
<td>1957</td>
<td>Typed-character recognition</td>
<td>Cannot recognize complex characters (such as Chinese); sensitive to differences in scale, translation, rotation; able to identify complex characters (such as Chinese)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-organizing</td>
<td>1980</td>
<td>Maps one geometrical region (such as a rectangular grid) onto another (such as an aircraft)</td>
<td>Requires extensive training</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Richard P. Lippmann: An Introduction to Computing with Neural Nets, IEEE ASSP Magazine, 4/97*
Classification of ANN Paradigms

Neural Networks

- **Supervised Learning**
  - Back propagation

- **Unsupervised Learning**
  - Kohonen's,
  - Cognitron,
  - ART
Classification of ANN Paradigms

Artificial Neural Networks Training Paradigms

- Supervised
  - Recurrent
    - Hopfield net (B)
    - Hamming net (B)
    - Bidirectional associative memory (B)
  - Nonrecurrent
    - Delta rule (C)
    - Back propagation (C)
    - Statistical methods

- Unsupervised
  - Recurrent
    - Carpenter/Grossberg: ART (B)
  - Nonrecurrent
    - Kohonen: Self organizing feature maps (C)

(B) - Normally used with Binary inputs
(C) - Normally used with Continuous inputs
Artificial Neural Networks?

- Maps an input pattern to an output Pattern
  - Like an expert system
- Parallel distributed computing
- Use examples instead of rules
- Works with incomplete information
  - Complete results are unknown or hard to compile
Parallel Distributed Processing Model
Parallel Distributed Processing Model

- A set of processing Units
- A state of activation
- An output function for each unit
- A pattern of connectivity among units
- A propagation rule for propagating patterns of activities through the network of connectivities
Parallel Distributed Processing Model

• An activation rule for combining the inputs impinging on a unit with the current state of that unit to produce a new level of activation for the unit

• A learning rule whereby patterns of connectivity are modified by experience

• An environment within which the system must operate