Lecture 28 - The Perceptron

The Perceptron - Artificial Neuron

Before we discuss the perceptron and related algorithms in more detail, let us take a brief tour through the early beginnings of machine learning. Trying to understand how the biological brain works to design artificial intelligence, Warren McCullock and Walter Pitts published the first concept of a simplified brain cell, the so-called McCullock-Pitts (MCP) neuron, in 1943.

- W. S. McCulloch and W. Pitts. A Logical Calculus of the Ideas Immanent in Nervous Activity. The bulletin of mathematical biophysics, 5(4):115–133, 1943

Neurons are interconnected nerve cells in the brain that are involved in the processing and transmitting of chemical and electrical signals.

Neurons and Neural Networks in the Human Brain

The brain is a highly complex, nonlinear and parallel information-processing system composed of *neurons*.

At birth, a brain has the ability to build up its own rules through experience. The most dramatic development of the human brain takes place during the first two years from birth.

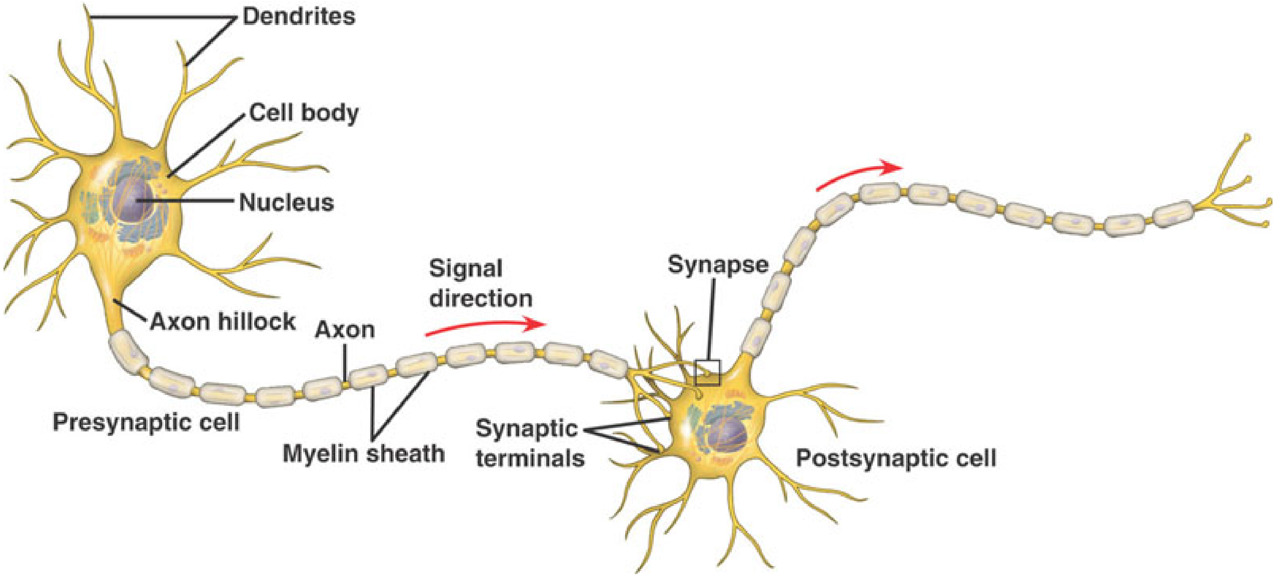
The neuron in the human brain is typically five to six orders of magnitude slower than silicon logic gates (10^{-3}s/millisecond vs. 10^{-9}s/nanosecond) but makes it up by having an enormous number of massively interconnected neurons

- Neurons in the human brain:

In [1]:

```python
from IPython.display import Image
Image('figures/neuron_structure.jpg', width=900)
```

Out[1]:



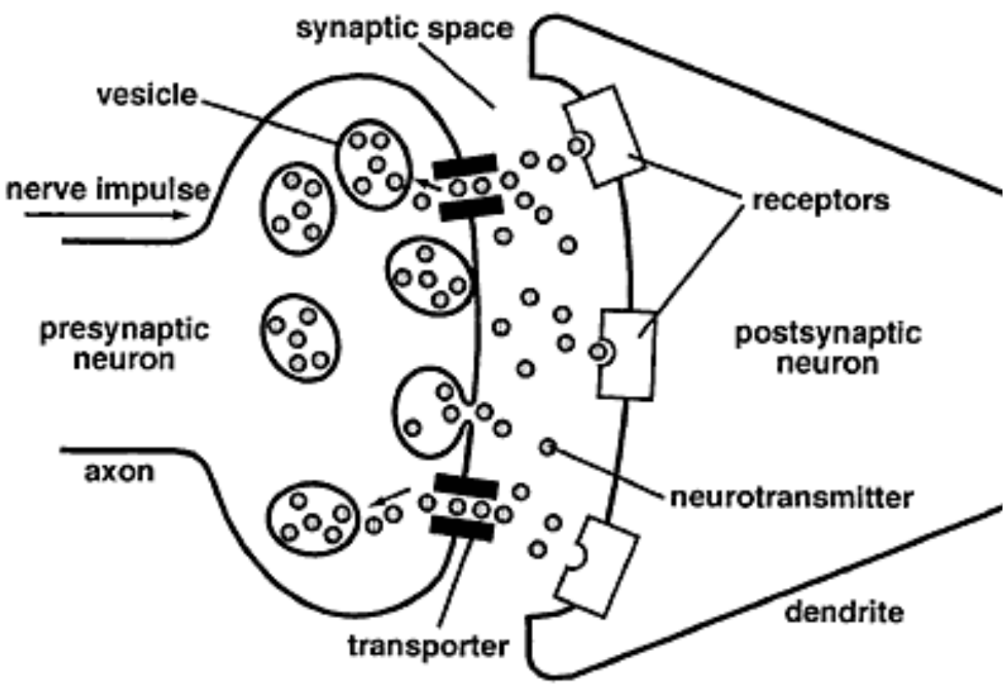
The **synapse** is an elementary structural and functional unit that controls *interaction* between neurons.

- *Chemical synapse*, most common kind of synapse, converts a pre-synaptic electrical signal into a chemical signal and then back into a post-synaptic electrical signal.

Traditional descriptions of assume that a synapse can either impose **excitation** or **inhibition** (but not both) on the receptive neuron.

In [2]:
```
Image('figures/synapse.png', width=500)
```

Out[2]:

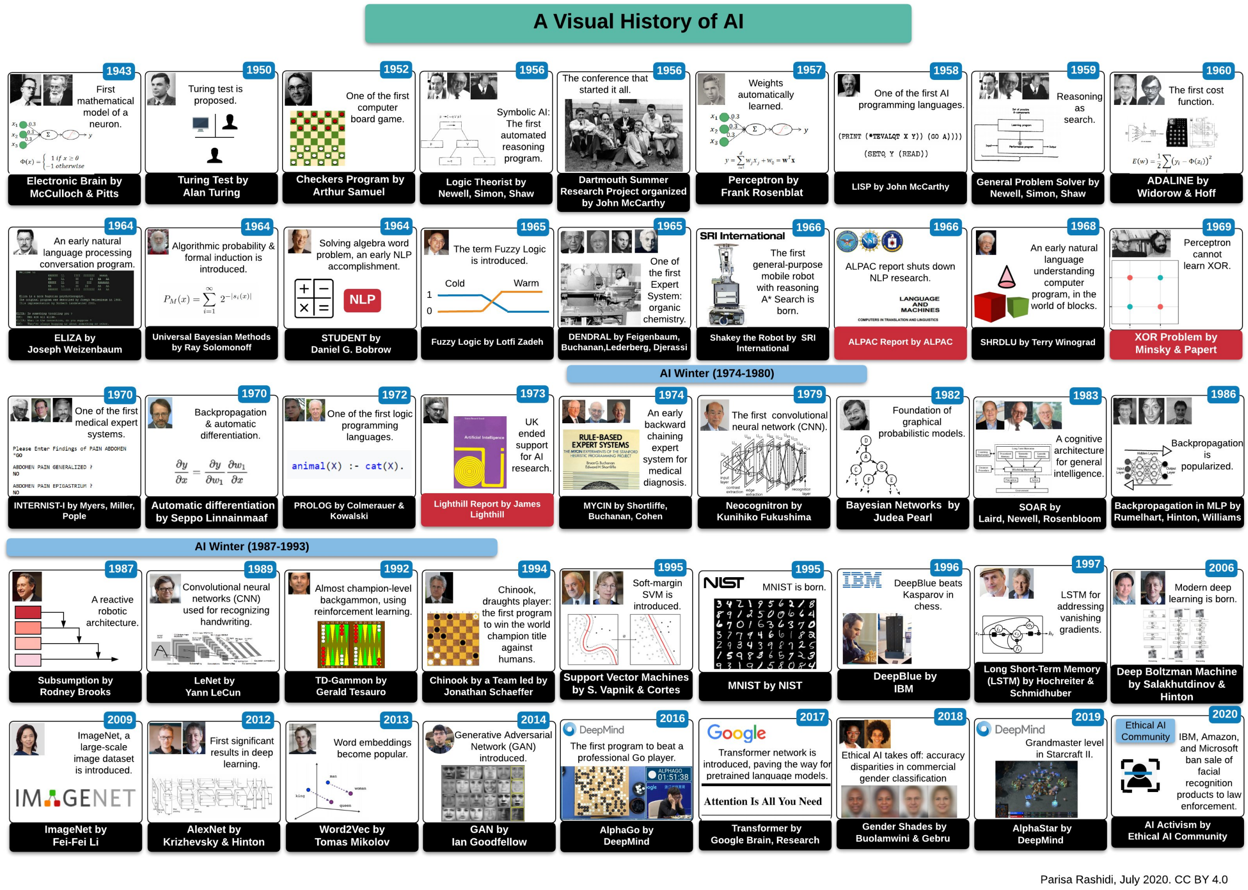


Brief History of Artificial Neural Networks

In [3]:
```
# Image('figures/NN_timeline.png',width=1000)

Image('figures/A_Visual_Histoy_of_AI.jpg',width=1000)
```

Out[3]:



- 1943 McCulloch and Pitts. McCulloch was a psychatrist and neuroanatomist. Pitts was a mathematician. They published a widely read article that introduced the idea of neural networks as computing machines. Their goal was to develop a model/understand how neurons in the brain might work. They showed a range of arithmetic and logical functions their neuron could compute.

- 1949 Hebb wrote "The Organization of Behavior" which postulated (among many other things): "When an axon of cell A is near enough to excite cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased" or "when neurons fire together they wire together". This has come to be known as *Hebbian learning*.

- 1954 Minsky wrote a "neural network" doctoral thesis at Princeton.

- 1957 Rosenblatt introduced his work in the perceptron and he came up with the perceptron convergence theorem. First model for learning with a teacher (i.e., "supervised learning").

- 1969 Minsky and Papert demonstrated the limits of the perceptron. They introduce multi-layer perceptrons but the published limits had the biggest influence - and interest dropped away. The AI winter began.

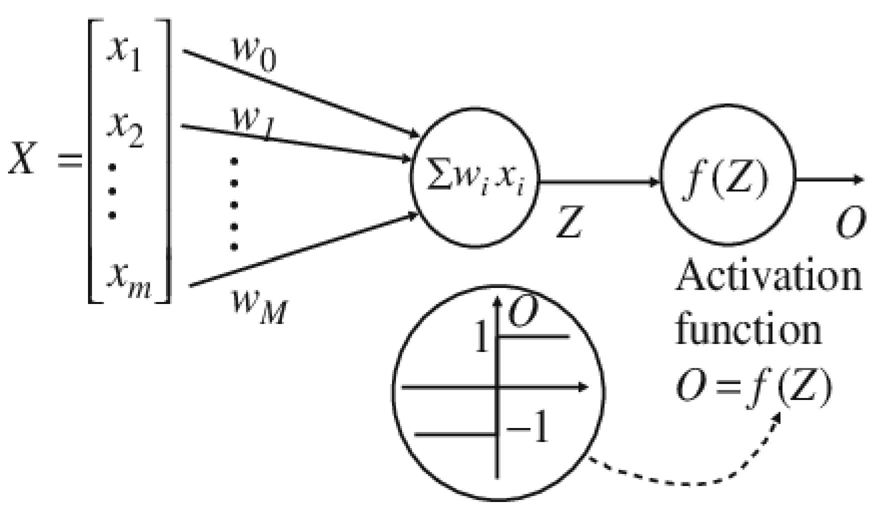
- 1974 Werbos' Ph.D. thesis at Harvard developed back-propagation.

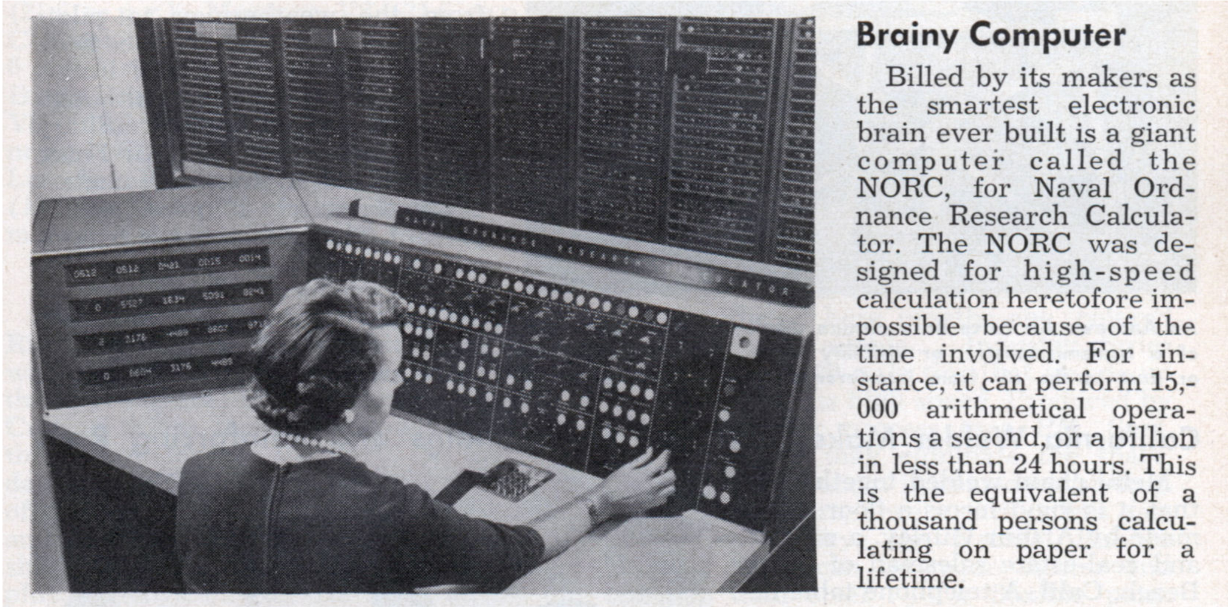
- 1986, the book "Parallel Distributed Processing: Explorations in the Microstructures of Cognition" was published and it covered back-propagation. This made Neural Networks popular again.
- In 1989, Yann LeCun used back-propagation to learn the convolution kernel coefficients directly from images of hand-written numbers.
- Early 90's, Support Vector Machines (SVMs) overtook ANNs in popularity due to a number of challenges/downsides to ANNs in comparison to SVMs. This included that SVMs were less likely to overtrain and easier to get good results on. Also, ANNs were very slow to train and had issues when they became "deep".
- 2012 ImageNet challenge won by Hinton's team using a deep CNN (based on top 5 error rate, given an image, the model does not output the correct label within its top 5 predictions). They had an error rate of 15.4\% (which was way better than 2nd place at 26.6\%). This started the current DL/ANN resurgance. Now it's huge.
- *What do you think will happen next?*

Electronic Brain (McCullock and Pitts, 1943)

McCullock and Pitts (1943) described such a nerve cell as a simple logic gate with binary outputs; multiple signals arrive at the dendrites, are then integrated into the cell body, and, if the accumulated signal exceeds a certain threshold, an output signal is generated that will be passed on by the axon.

- McCullock and Pitts Neuron (MCP):





Brainy Computer

Billed by its makers as the smartest electronic brain ever built is a giant computer called the NORC, for Naval Ordnance Research Calculator. The NORC was designed for high-speed calculation heretofore impossible because of the time involved. For instance, it can perform 15,000 arithmetical operations a second, or a billion in less than 24 hours. This is the equivalent of a thousand persons calculating on paper for a lifetime.

Popular Mechanics 5-1955

The Perceptron (Frank Rosenblatt, 1957)

In **1957, Frank Rosenblatt** published the first concept of the perceptron learning rule based on the MCP neuron model.

- F. Rosenblatt, The Perceptron, A Perceiving and Recognizing Automaton. Cornell Aeronautical Laboratory, 1957

With his perceptron rule, Rosenblatt proposed an algorithm that would **automatically learn the optimal weight coefficients** that are then multiplied with the input features in order to make the decision of whether a neuron fires or not. In the context of supervised learning and classification, such an algorithm could then be used to predict if a sample belonged to one class or the other.

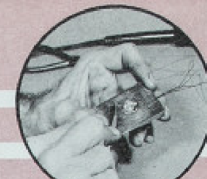
More formally, we can pose this problem as a **binary classification task** where we refer to our two classes as 1 (positive class) and -1 (negative class) for simplicity.

This worked produced the **Mark I Perceptron**.

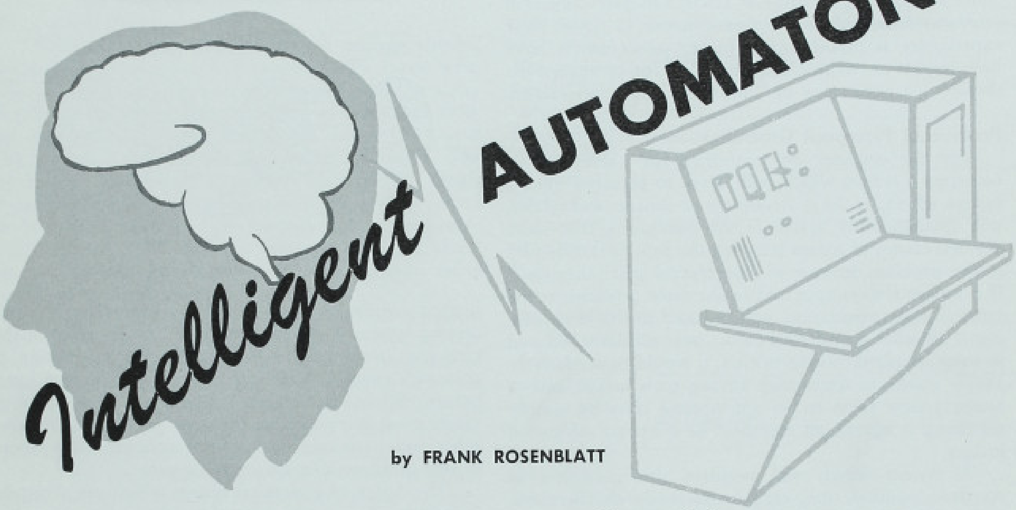
Vol. VI, No. 2, Summer 1958

research trends

CORNELL AERONAUTICAL LABORATORY, INC., BUFFALO 21, NEW YORK



The Design of an *Intelligent* AUTOMATON



by FRANK ROSENBLATT

Introducing the perceptron — A machine which senses, recognizes, remembers, and responds like the human mind.

STORIES about the creation of machines having human qualities have long been a fascinating province in the realm of science fiction. Yet we are now about to witness the birth of such a machine — a machine capable of perceiving, recognizing, and identifying its surroundings without any human training or control.

Development of that machine has stemmed from a search for an understanding of the physical mechanisms which underlie human experience and intelligence. The question of the nature of these processes is at least as ancient as any other question in western science and philosophy, and, indeed, ranks as one of the greatest scientific challenges of our time.

Our understanding of this problem has gone perhaps as far as had the development of physics before Newton. We have some excellent descriptions of the phenomena to be explained, a number of interesting hypotheses, and a little detailed knowledge about events in the nervous system. But we lack agreement on any integrated set of principles by which the functioning of the nervous system can be understood.

We believe now that this ancient problem is about to yield to our theoretical investigation for three reasons:

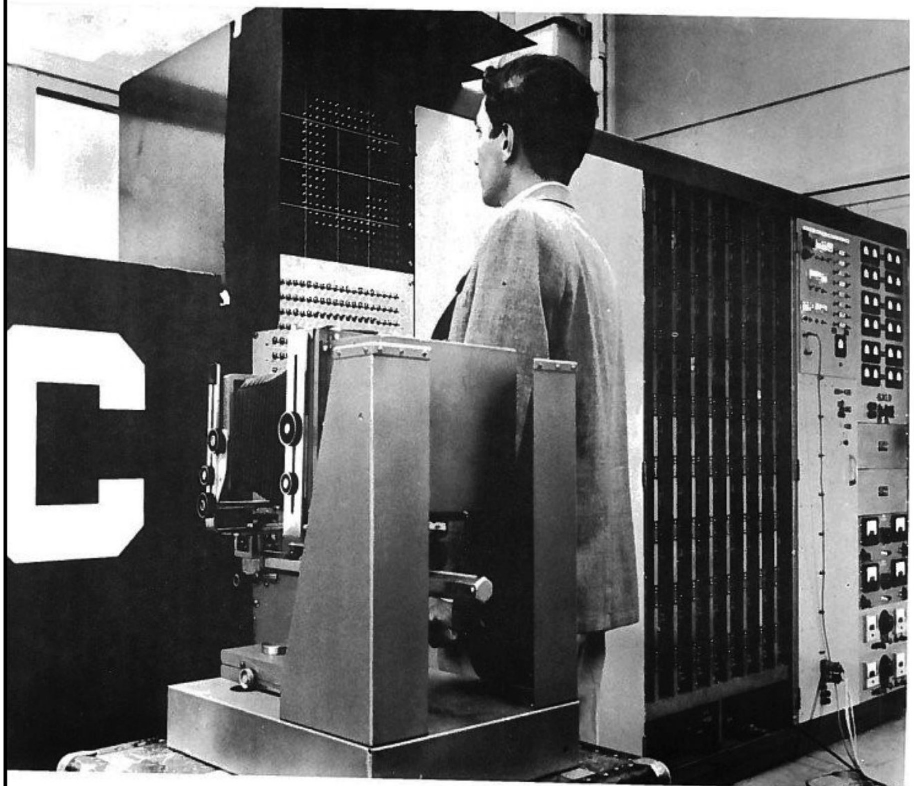
First, in recent years our knowledge of the functioning of individual cells in the central nervous system has vastly increased.

Second, large numbers of engineers and mathematicians are, for the first time, undertaking serious study of the mathematical basis for thinking, perception, and the handling of information by the central nervous system, thus providing the hope that these problems may be within our intellectual grasp.

Third, recent developments in probability theory and in the mathematics of random processes provide new tools for the study of events in the nervous system, where only the gross statistical organization is known and the precise cell-by-cell "wiring diagram" may never be obtained.

Receives Navy Support

In July, 1957, Project PARA (Perceiving and Recognizing Automaton), an internal research program which had been in progress for over a year at Cornell Aeronautical Laboratory, received the support of the Office of Naval Research. The program had been concerned primarily with the application of probability theory to



THE MARK I PERCEPTRON

NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI) —The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

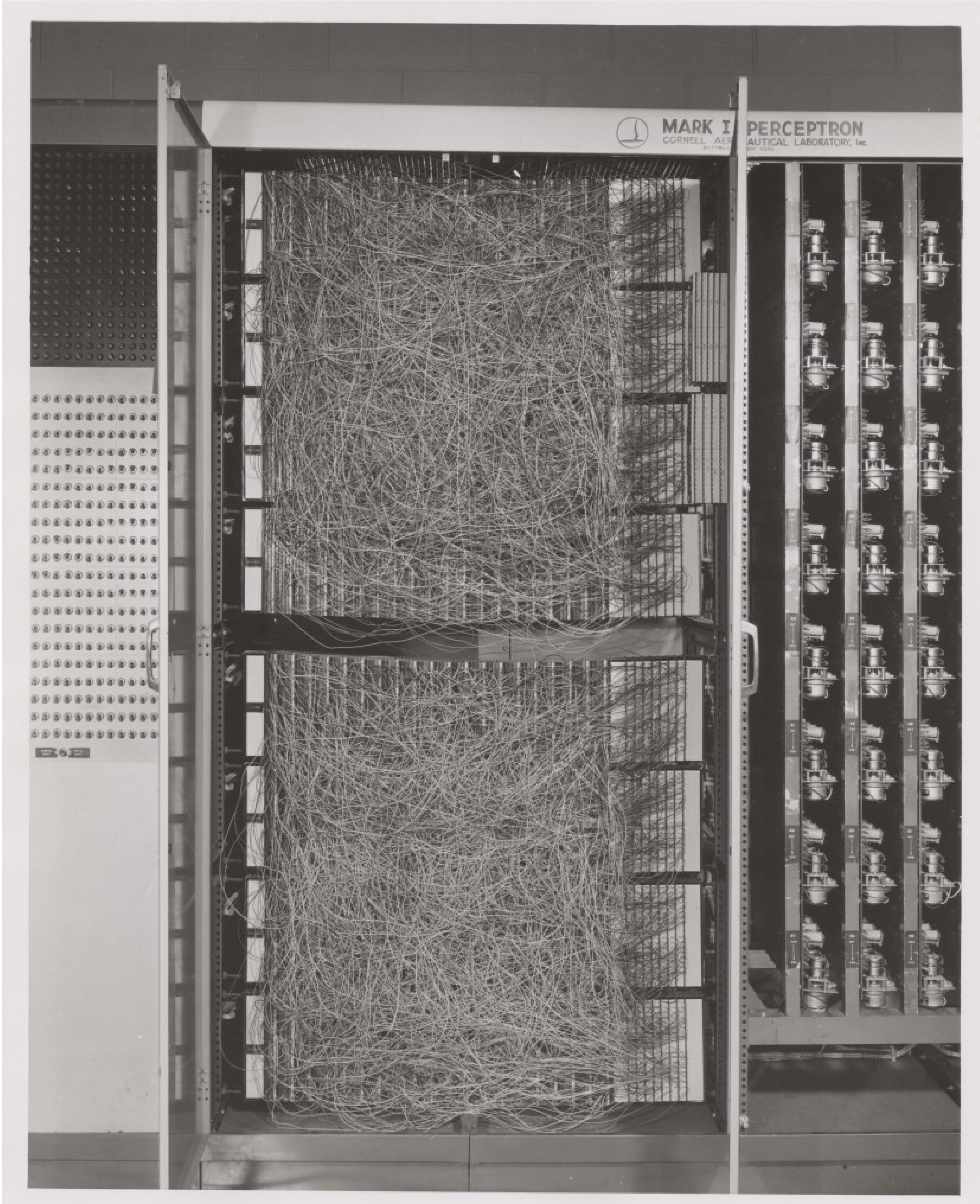
The embryo—the Weather Bureau's $2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.

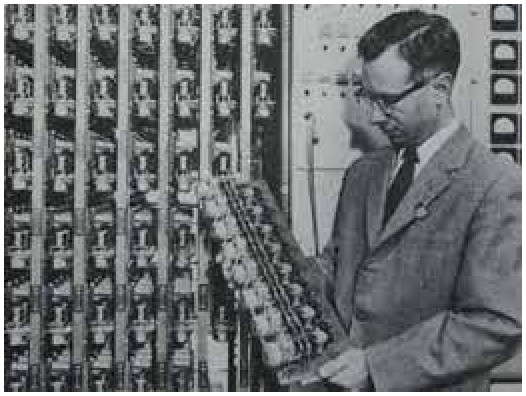
Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human beings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

UPI wire report, New York Times, July 8, 1958



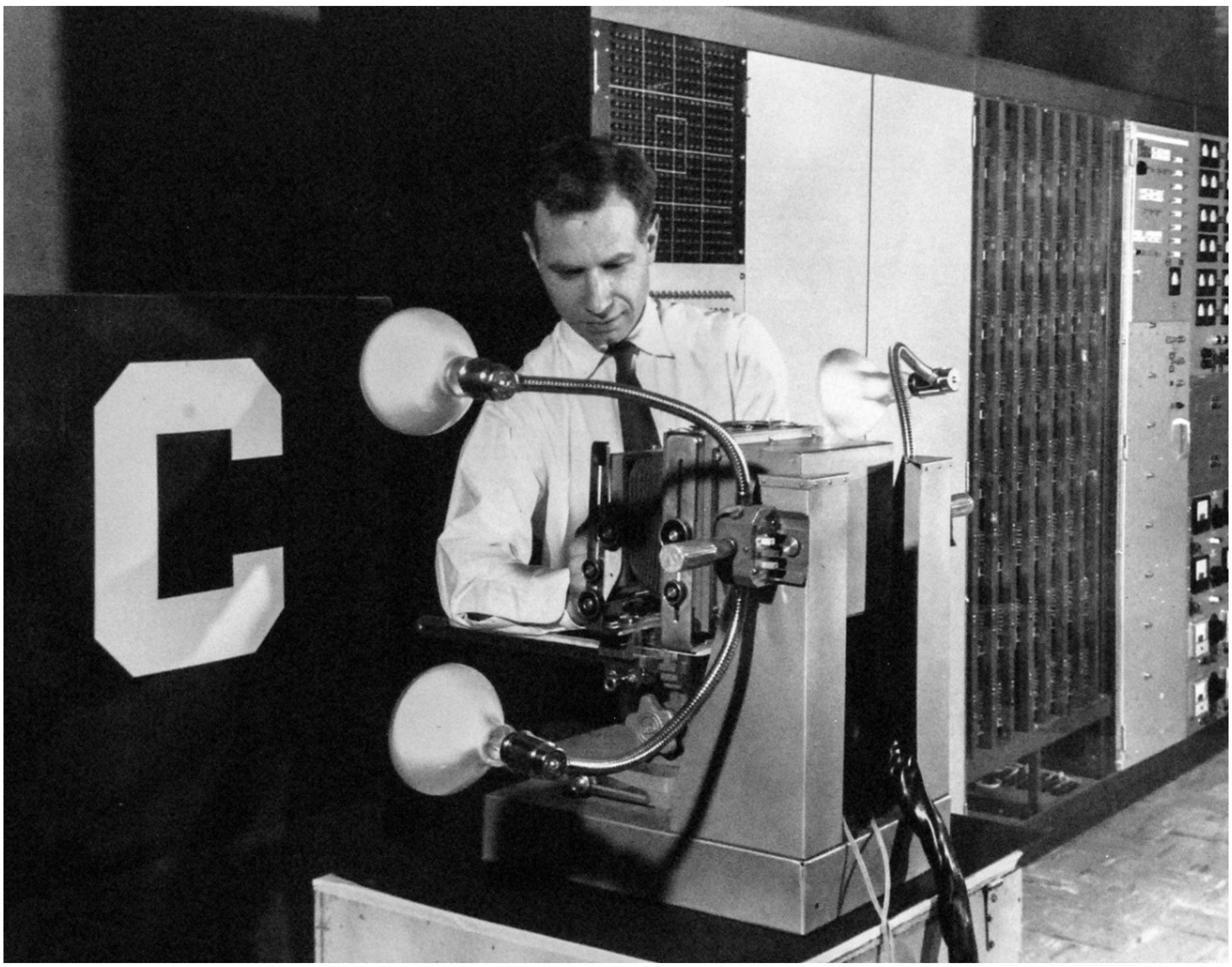
Mark I Perceptron



Frank Rosenblatt holding an array of potentiometers

The perceptron was implemented in hardware that got the name of **Mark I Perceptron**.

The weights were encoded in potentiometers, and weight updates were done by electric motors.

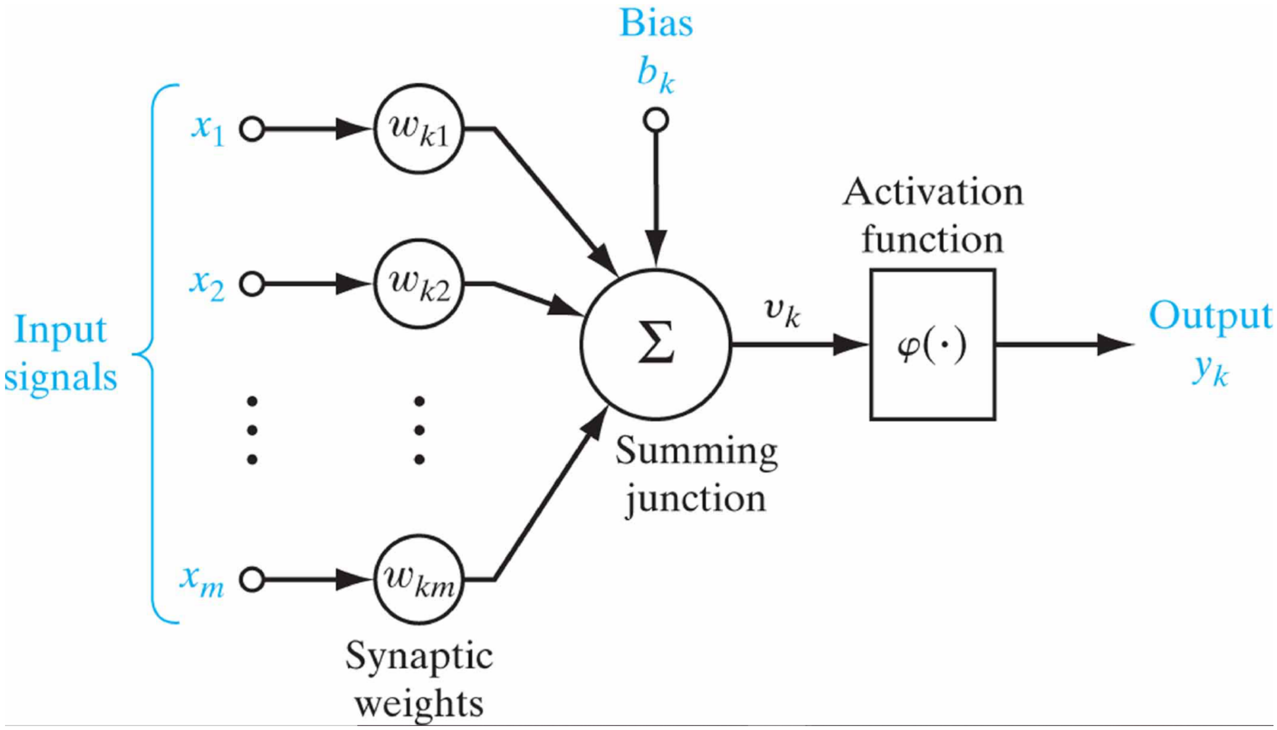


A basic model for a neuron consists of the following:

- A set of *synapses* each of which is characterized by a *weight* (which includes a *bias*).
- An *adder*.
- An *activation function* (e.g. linear function)

In [4]:
```
Image('figures/modelneuron.png', width=700)
```

Out[4]:



We can write this mathematically as:

$$y = \varphi\left(v\right)$$

where

$$v = \sum_{j=1}^{m} w_j x_j + b = \mathbf{w}^T \mathbf{x} + b$$

and $\varphi(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$

- *What does this look like graphically?*

The Perceptron Algorithm

Consider an alternative error function known as the *perceptron criterion*. To derive this, we note that we are seeking a weight vector $\mathbf{w}$ such that patterns x_i in class C_1 will have $\mathbf{w}^T x_i + b > 0$, whereas the patterns x_i in class C_2 have $\mathbf{w}^T x_i + b < 0$. Using the $t \in \{-1, 1\}$ target coding scheme it follows that we would like all patterns to satisfy

$$(\mathbf{w}^T x_i + b)t_i > 0$$

- The perceptron criterion associates zero error with any pattern that is correctly classified, whereas for a misclassified pattern x_i it tries to minimize the quantity $-(\mathbf{w}^T x_i + b)t_i$.

- The perceptron criterion is therefore given by:

$$E_p(\mathbf{w}, b) = -\sum_{n \in \mathcal{M}} (\mathbf{w}^T \mathbf{x}_n + b)t_n$$

where $\mathcal{M}$ denotes the set of all misclassified patterns.

- We now apply the *stochastic gradient descent* algorithm to this error function. The change in the weight vector $\mathbf{w}$ is then given by:

$$\mathbf{w}^{(t+1)} \leftarrow \mathbf{w}^{(t)} - \eta \frac{\partial E_p(\mathbf{w}, b)}{\partial \mathbf{w}} = \mathbf{w}^{(t)} + \eta \mathbf{x}_n t_n \tag{1}$$

$$b^{(t+1)} \leftarrow b^{(t)} - \eta \frac{\partial E_p(\mathbf{w}, b)}{\partial b} = b^{(t)} + \eta t_n \tag{2}$$

where η is the **learning rate** parameter and t is an integer that indexes the iteration steps of the algorithm.

- Note that, as the weight vector evolves during training, the set of patterns that are misclassified will change.

In [5]:

```python
Image('figures/PerceptronLearning.png', width=700)
```

Out[5]:

Algorithm 1: Perceptron Learning Algorithm

Data: Training data matrix $\mathbf{X}$, Truth Values $y \in \{-1, 1\}$, Parameter η
Result: Weight vector $\mathbf{w}$ and bias b
Initialize weight vector and bias;
$errorDetected \leftarrow True$;
while $errorDetected$ **do**
  $errorDetected \leftarrow False$;
  for $n = 1 : N$ **do**
    $v \leftarrow \mathbf{w}^T \mathbf{x}_n + b$;
    if $sign(v) == y_n$ **then**
      $\mathbf{w} \leftarrow \mathbf{w}$
      $b \leftarrow b$
    else
      $errorDetected \leftarrow True$;
      $\mathbf{w} \leftarrow \mathbf{w} + \eta y_n \mathbf{x}_n$
      $b \leftarrow b + \eta y_n$