Lecture 18 Part 2 - Linear Discriminant Function: The Perceptron

The Perceptron, 1957

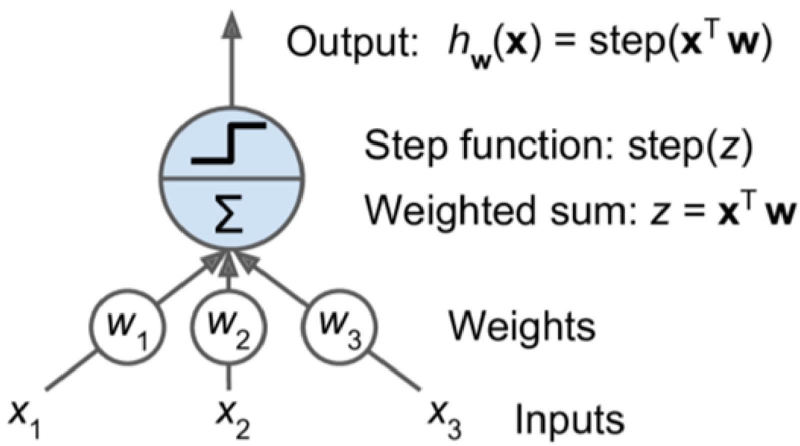
A basic model for a neuron consists of the following:

- 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 [1]:

```
from IPython.display import Image
Image('figures/perceptron.png', width=400)
```

Out[1]:



The Perceptron is one of the simplest ANN architectures, invented in 1957 by Frank Rosenblatt. 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.

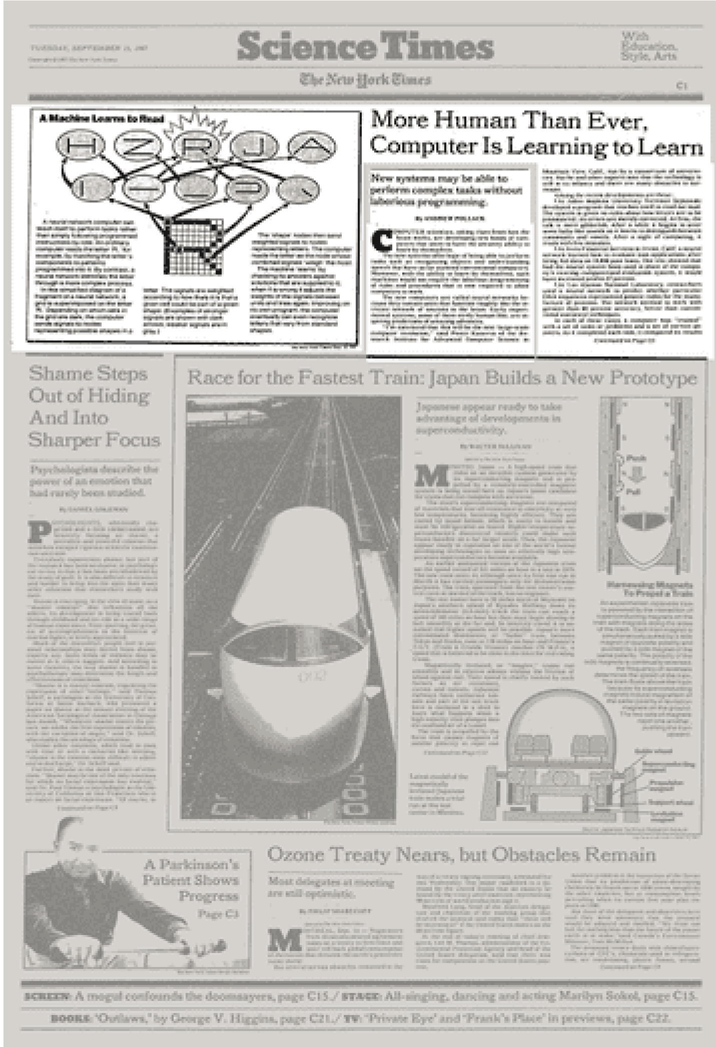
Second, large numbers of engineers and mathematicians are, for the first time, undertaking serious study of the mathematical basis for thinking, perception, and the handling of information by the central nervous system, thus providing the hope that these problems may be within our intellectual grasp.

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

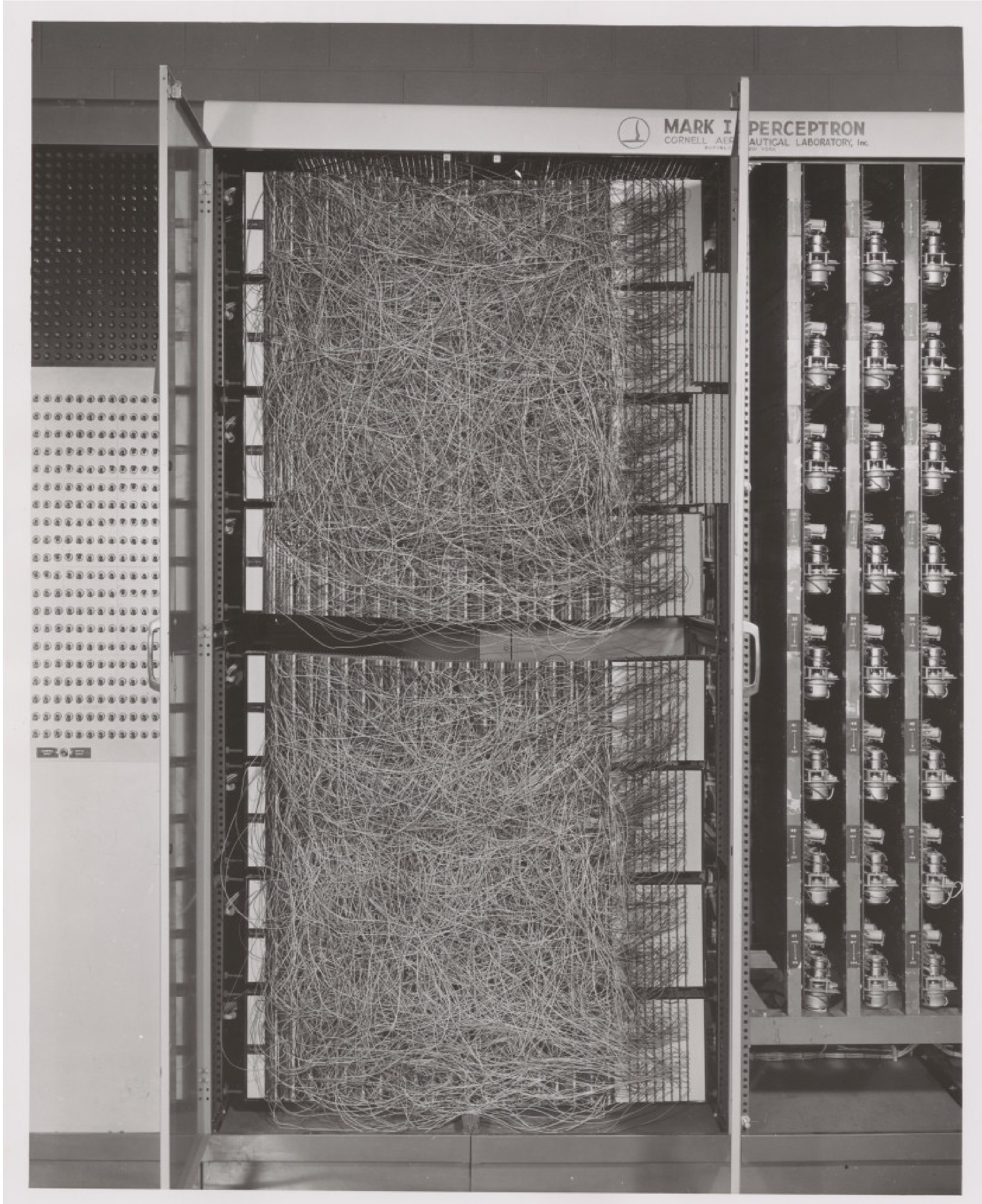
In July, 1957, Project PARA (Perceiving and Recognizing Automaton), an internal research program which had been in progress for over a year at Cornell Aeronautical Laboratory, received the support of the Office of Naval Research. The program had been concerned primarily with the application of probability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

Research Trends, Cornell Aeronautical Laboratory, 1958

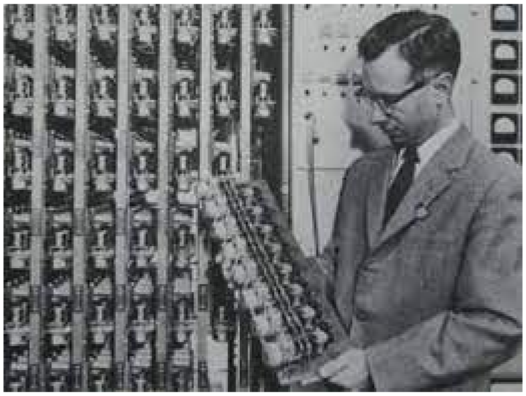


"More Human than Ever, Computer is Learning to Learn", New York Times, September 15, 1987, Section C, Page 1

Perceptron, Mark I. National Museum of American History.



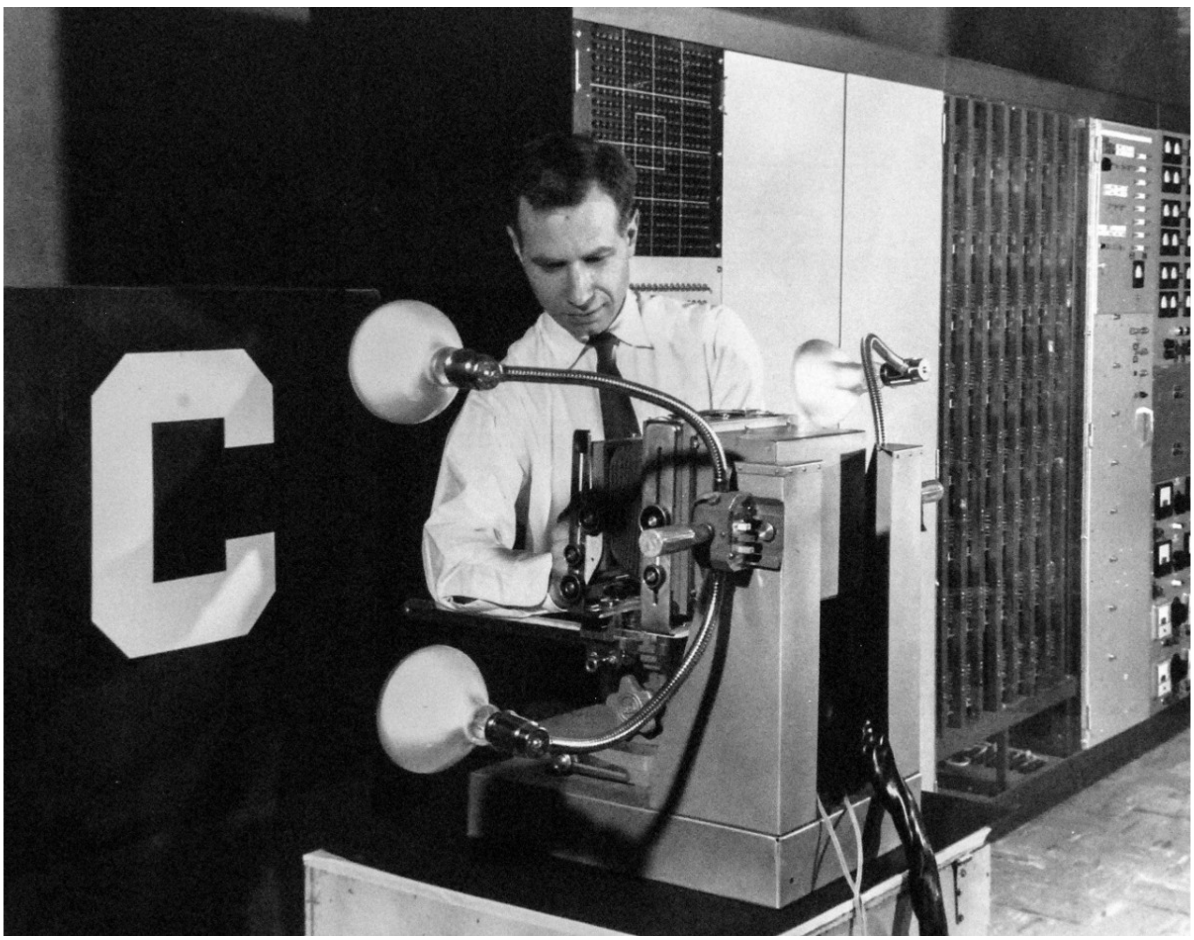
Mark I Perceptron



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.



We can write this mathematically as:

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

where

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

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

- *What does this look like graphically?*

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.

to be continued...
