Computation, Information, and Intelligence (ENGRI/CS/INFO/COGST 172), Spring 2007
2/7/07: **Lecture aid — Perceptrons**

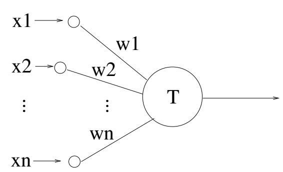
Topics: perceptrons as simple models of neurons; mathematical characterization as linear separators.
Announcements:

- We are beginning a fairly sophisticated unit of the course. It is recommended that you (continue to) make the effort to review and understand the topics of each lecture before the next one comes around.
- Please look over the "Vector-operations reference sheet" handout before next lecture.

I. Comments on Deep Blue

1. Deep Blue [had] ingenious counterattacks .... [It engaged in] original play in the opening .... No one had foreseen its scintillating method of certifying the draw. – Robert Byrne (Grandmaster and NYT chess columnist), *New York Times*, May 13, 1997.
2. The Brain's Last Stand. — *Newsweek*, May 5, 1997 cover.
3. [Computers] never learn.... Deep Blue plays chess better ... but only because human beings have carefully programmed Deep Blue to play chess. Left on its own, Deep Blue wouldn't even know to come in out of the rain. — Joel Achenbach, *Washington Post*, May 10, 1997. Emphasis added.

II. Perceptrons: A simple model of a neuron Perceptrons are characterized by a weight vector $\overrightarrow{w}$ and a threshold value T.



Letting n be the number of components in $\overrightarrow{w}$, a perceptron computes the following function on n-component vectors $\overrightarrow{x}$

$$f_{\overrightarrow{w},T}(\overrightarrow{x}) = \begin{cases} +1, & \overrightarrow{w} \cdot \overrightarrow{x} \geq T \\ -1 & \text{otherwise} \end{cases}$$

where the *inner product* (or *dot product*) between $\overrightarrow{w}$ and $\overrightarrow{x}$ is defined as $w_1x_1 + w_2x_2 + \cdots + w_nx_n$.

III. Example binary functions

1. Disease diagnosis. Input vector: (weight relative to population average, platelet count relative to population average, Chemical-X level relative to population average).

$$f(\overrightarrow{x}) = \begin{cases} +1 & \text{if the patient has the disease} \\ -1 & \text{otherwise} \end{cases}.$$

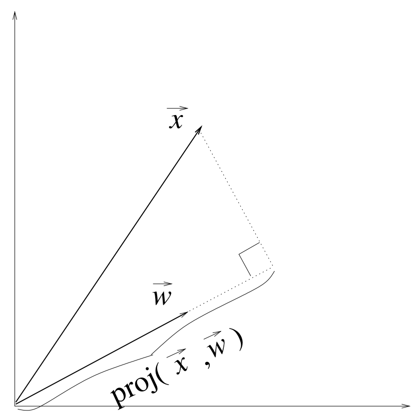
2. The "turn right?" function. Input vector: a camera view $\overrightarrow{x}$, where x_i is the brightness of pixel i.
3. Spam detection. Input vector: $\overrightarrow{x}$, where each x_i is the number of appearances of the i^{th} predetermined suspicious or auspicious phrase (e.g., "free" or "stock tip" vs. "172 announcement").

(OVER)

IV. Example perceptron Let $\overrightarrow{w} = (-1, 1), T = \frac{1}{2}$.

	x_1	x_2	$\overrightarrow{w} \cdot \overrightarrow{x}$	$f_{\overrightarrow{w},T}(\overrightarrow{x})$
(a)	1	3		
(b)	-3	2		
(c)	3	-2		
(d)	$\frac{3}{4}$	1		

V. The inner product and length of projections Here is a schematic showing the *projection* $\mathrm{proj}(\overrightarrow{x}, \overrightarrow{w})$ of $\overrightarrow{x}$ onto $\overrightarrow{w}$, which is a *signed* length: it is negative if the cosine of the angle between $\overrightarrow{x}$ and $\overrightarrow{w}$ is negative, that is, if $\overrightarrow{x}$ and $\overrightarrow{w}$ are more than $90°$ apart.[1]



The inner product relates to vector projections in the following way.

$$\begin{aligned} \overrightarrow{w} \cdot \overrightarrow{x} &= \mathrm{length}\,(\overrightarrow{w})\,\mathrm{length}\,(\overrightarrow{x})\cos(\angle(\overrightarrow{w}, \overrightarrow{x})) \\ &= \mathrm{length}\,(\overrightarrow{w})\,\mathrm{proj}(\overrightarrow{x}, \overrightarrow{w}) \end{aligned}$$

because “cosine = adjacent over hypotenuse”.

[1]We are abusing the conventional definition of the term projection for notational convenience.