Computation, Information, and Intelligence (ENGRI/CS/INFO/COGST 172), Spring 2007
2/12/07: **Lecture aid — Online learning; obstacles to perceptron learning**

Topics: formalizing the learning problem; possible restrictions on the oracle to make learning feasible/

I. Self-check: A (pseudo-)sensory example Suppose that we have a perceptron whose input comes from a small patch of 12 photoreceptors, each of which are either in an "on" or an "off" state:

x_1	x_4	x_7	x_{10}
x_2	x_5	x_8	x_{11}
x_3	x_6	x_9	x_{12}

and we want the perceptron to "fire" whenever it sees a "darkness to the left pattern": at least half of the "left-hand" photoreceptors x_1 through x_6 are on, and all of the remaining right-hand photoreceptors are off. Give one choice of weight vector and threshold that achieves our goal.[1]

II. Identification in the limit, a success criterion for on-line learning For any (infinite) sequence of labeled *examples* (a.k.a. *instances*) the oracle presents, after some finite period of time,

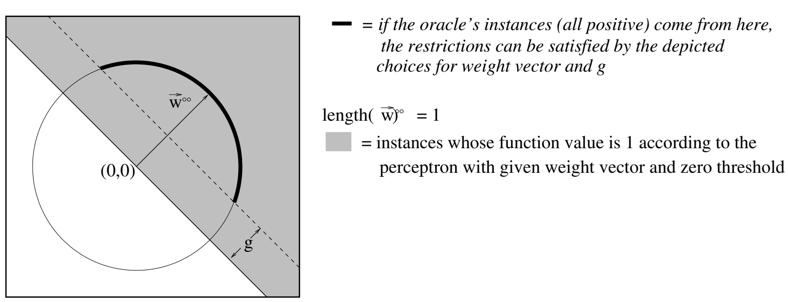
- the learner always outputs the same predictor from then on, and
- this predictor correctly predicts the label of every subsequent example.

In order for identification in the limit to be feasible at all in the perceptron-learning case, we will further require that the labels assigned by the oracle to all of the infinite number of examples it produces turn out to be consistent with some single perceptron function $f_{\overrightarrow{w}^{\infty}, T^{\infty}}$.

III. Initial oracle restrictions We impose the following restrictions just for simplicity of presentation; they can be relaxed considerably.

1. The *simple-perceptron consistency* condition: $\mathrm{length}\,(\overrightarrow{w}^{\infty}) = 1$ and $T^{\infty} = 0$.
2. The *length restriction*: For all i, $\mathrm{length}\left(\overrightarrow{x}^{(i)}\right) = 1$.

IV. The gap condition [2]: There is a $g > 0$ such that for all $\overrightarrow{x}^{(i)}$, $\overrightarrow{w}^{\infty} \cdot \overrightarrow{x}^{(i)} \geq g$.



---

[1]Answer: Set $w_1 = w_2 = \cdots = w_6 = 1$ and $w_7 = w_8 = \cdots = w_{12} = -6$; set $T = 3$.

[2]The "real" but (slightly) harder to work with version of this condition is "double-sided", requiring only that $|\overrightarrow{w}^{\infty} \cdot \overrightarrow{x}^{(i)}| \geq g$. This corresponds to having a gap, or *margin*, between the positive and negative examples, and eliminates "cheating" solutions (such as setting the weight vector to all-zeroes always) on the part of the learner.