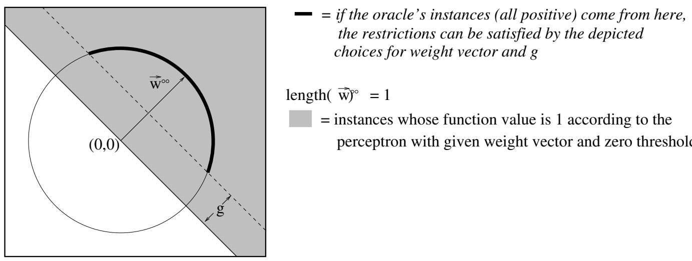
Computation, Information, and Intelligence (ENGRI/CS/INFO/COGST 172), Spring 2007
2/14/07: **Lecture 11 aid — The perceptron convergence theorem**

Topics: the gap condition; the perceptron learning algorithm (PLA); the perceptron convergence theorem

I. Restrictions on the oracle We assume that the labels assigned by the oracle turn out to be consistent with some single perceptron function $f_{\overrightarrow{w}^{\infty}, T^{\infty}}$.

1. The *simple-perceptron consistency* condition: $\mathrm{length}\left(\overrightarrow{w}^{\infty}\right) = 1$ and $T^{\infty} = 0$.
2. The *length restriction*: For all i, $\mathrm{length}\left(\overrightarrow{x}^{(i)}\right) = 1$.
3. The *gap condition*[1]: There is a $g > 0$ such that for all $\overrightarrow{x}^{(i)}$, $\overrightarrow{w}^{\infty} \cdot \overrightarrow{x}^{(i)} \geq g$.



II. The perceptron learning algorithm This is a "one-sided" version of Rosenblatt's algorithm.

1) Set $\overrightarrow{w}^{(0)}$ to all zeroes.
2) For each example $\overrightarrow{x}^{(i)}$ (i increasing from 1 on),
3) If $\overrightarrow{w}^{(i-1)} \cdot \overrightarrow{x}^{(i)} \leq 0$,
4) set $\overrightarrow{w}^{(i)}$ to $\overrightarrow{w}^{(i-1)} + \overrightarrow{x}^{(i)}$ ("update");
5) otherwise, set $\overrightarrow{w}^{(i)}$ to $\overrightarrow{w}^{(i-1)}$ ("no change").

Note: it can be shown that under the restrictions above, for all $i > 0$, $\mathrm{length}\left(\overrightarrow{w}^{(i)}\right) > 0$.

III. Outline of the proof of (our version of) the perceptron convergence theorem

Given all the constraints we have about the oracle and learner, and using various facts from the "Vector operations reference sheet",

- Use the cosine function to measure how "close" successive hypothesis vectors are to $\overrightarrow{w}^{\infty}$. Observe that it takes the form N/D (numerator over denominator).
- Show that at each *update* of the perceptron learning algorithm, i.e., where $\overrightarrow{w}^{(i)}$ is different from $\overrightarrow{w}^{(i-1)}$, the cosine increases by a non-negligible amount:
  - N increases by *at least* g, the *gap* quantity.
  - The square of D increases by *at most* 1.

  Hence, after u updates, the cosine must be at least $\sqrt{u}g$.
- Since cosines can't get bigger than one, we get that u can be at most $1/g^2$, which, since $g > 0$, must be finite.

---

[1] 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.