Computation, Information, and Intelligence (ENGRI/CS/INFO/COGST 172), Spring 2007
2/16/07: **Lecture 12 aid — Proof of the perceptron convergence theorem**

Topics: further exploration of the perceptron learning algorithm (PLA); proof of the perceptron convergence theorem.

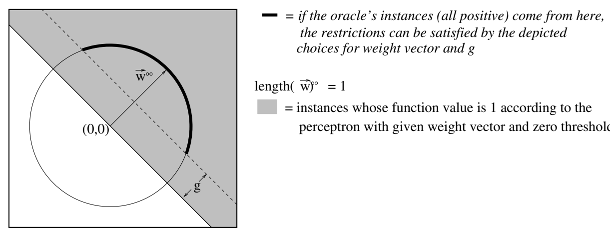
Announcements: Rafael Frongillo's office hours for this Sunday 8-9pm only will be held in the Clara Dickson computer lab, located in the basement of Dickson Hall. (The usual room in Balch is unavailable this particular weekend.)

I. Reminders Here are the oracle restrictions again; we are employing some slight rephrasings for compactness, since we've discussed these at length already.

1. *Simple-perceptron consistency*: there exists a $\overrightarrow{w}^{\infty}$ such that $\text{length}(\overrightarrow{w}^{\infty}) \stackrel{\text{def}}{=} \sqrt{\overrightarrow{w}^{\infty} \cdot \overrightarrow{w}^{\infty}} = 1$ and for all i, $\ell^{(i)} = f_{\overrightarrow{w}^{\infty},0}(\overrightarrow{x}^{(i)})$.

2. *Length restriction*: For all i, $\text{length}\left(\overrightarrow{x}^{(i)}\right) \stackrel{\text{def}}{=} \sqrt{\overrightarrow{x}^{(i)} \cdot \overrightarrow{x}^{(i)}} = 1$.

3. *Gap condition (one-sided version)* There is a $g > 0$ such that for all i, $\overrightarrow{w}^{\infty} \cdot \overrightarrow{x}^{(i)} \geq g$.



II. (One-sided) perceptron learning 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").

III. A "normalized" perceptron "learning" algorithm The update step sets $\overrightarrow{w}^{(i)}$ to the *unit-length-normalized* version of $\overrightarrow{w}^{(i-1)} + \overrightarrow{x}^{(i)}$. What happens when $\overrightarrow{x}^{(1)} = (1,0)$, $\overrightarrow{x}^{(2)} = (0,1)$, $\overrightarrow{x}^{(3)} = (-\sqrt{2}/2, \sqrt{2}/2)$, $\overrightarrow{x}^{(4)} = (1,0)\ldots$, all positive examples?

IV. Self-check exercises Answers in footnotes.

Q: Can a restriction-compliant oracle choose $\overrightarrow{x}^{(1)} = (-2,-4)$, $\ell^{(1)} = +1$, $\overrightarrow{x}^{(2)} = (-1,0)$, $\ell^{(2)} = -1$?[1]

Q: Suppose $\overrightarrow{x}^{(1)} = \left(\sqrt{2}/2, \sqrt{2}/2\right)$ and $\ell^{(1)} = +1$. Give an instance $\overrightarrow{x}^{(2)} \neq \overrightarrow{x}^{(1)}$ and $\ell^{(2)}$ such that the two instances (with labels) satisfy our restrictions and such that the PLA sets $\overrightarrow{w}^{(2)} = \overrightarrow{w}^{(1)} \left(= \overrightarrow{x}^{(1)}\right)$ in response to seeing these two examples (in order).[2]

Q: Now give an $\overrightarrow{x}^{(3)}$ and $\ell^{(3)}$ such that $\overrightarrow{x}^{(1)}$, $\overrightarrow{x}^{(2)}$, and $\overrightarrow{x}^{(3)}$ collectively obey the oracle restrictions and such that the PLA produces a $\overrightarrow{w}^{(3)}$ different from $\overrightarrow{w}^{(2)}$.[3]

---

[1] No, only simple-perceptron consistency would hold.
[2] One solution of many: $\overrightarrow{x}^{(2)} = (\sqrt{3}/2, 1/2)$, $\ell^{(2)} + 1$.
[3] One solution of many: $\overrightarrow{x}^{(3)} = (-1,0)$, $\ell^{(3)} = +1$.

(OVER)

V. 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.