Recap from last time

- Winnow algorithm for learning a disjunction of $r$ out of $n$ variables. \( f(x) = x_3 \lor x_9 \lor x_{12} \)
- \( h(x) \): predict pos iff \( w_1 x_1 + \ldots + w_n x_n \geq n \).
- Initialize \( w_i = 1 \) for all \( i \).
  - Mistake on pos: \( w_i \leftarrow 2w_i \) for all \( x_i = 1 \).
  - Mistake on neg: \( w_i \leftarrow 0 \) for all \( x_i = 1 \).
- Thm: Winnow makes at most \( O(r \log n) \) mistakes.

Winnow for general LTFs

More generally, can show the following (will do the analysis on hwk2):

Suppose \( \exists w^* \) s.t.:
- \( w^* \cdot x \geq c \) on positive \( x \),
- \( w^* \cdot x \leq c - \gamma \) on negative \( x \).

Then mistake bound is
- \( O((L_1(w^*)/\gamma)^2 \log n) \)

Multiply by \( L_2(X) \) if examples not in \( \{0,1\} \)

Perceptron algorithm

An even older and simpler algorithm, with a bound of a different form.

Suppose \( \exists w^* \) s.t.:
- \( w^* \cdot x \geq \gamma \) on positive \( x \),
- \( w^* \cdot x \leq -\gamma \) on negative \( x \).

Then mistake bound is
- \( O(L_2(w^*)(L_2(x))/\gamma^2) \)

L_2 margin of examples

Perceptron algorithm

Thm: Suppose data is consistent with some LTF \( w^* \cdot x > 0 \), where \( ||w^*|| = 1 \) and
\[
\gamma = \min_x |w^* \cdot x|/||x||
\]

Then # mistakes \( \leq 1/\gamma^2 \).

Algorithm:
- Initialize \( w=0 \). Use \( w \cdot x > 0 \).
  - Mistake on pos: \( w \leftarrow w + x \).
  - Mistake on neg: \( w \leftarrow w - x \).

(Pre-scale examples to be in unit ball)
**Mistakes:**

So, in $M$ mistakes, each mistake increases $|w \cdot x|$.

Initialize $w = 0$. Use $w \cdot x > 0$.
- Mistake on pos: $w \leftarrow w + x$.
- Mistake on neg: $w \leftarrow w - x$.

**Proof:**

Suppose data is consistent with some LTF $w^* \cdot x > 0$, where $||w^*|| = 1$ and

$\gamma = \min_x |w^* \cdot x|$  (after scaling so all $||x|| = 1$)

Then # mistakes $\leq 1/\gamma$.

**Proof:** consider $|w \cdot w^*|$ and $||w||$
- Each mistake increases $|w \cdot w^*|$ by at least $\gamma$.
  - $(w + x) \cdot w^* = w \cdot w^* + x \cdot w^* \geq w \cdot w^* + \gamma$.
- Each mistake increases $w \cdot b$ by at most 1.
  - $(w + x) \cdot (w + x) = w \cdot w + 2(w \cdot x) + x \cdot x \leq w \cdot w + 1$.
- So, in $M$ mistakes, $\gamma M \leq |w \cdot w^*| \leq ||w|| \leq M^{1/2}$.
- So, $M \leq 1/\gamma^2$.

**What if no perfect separator?**

In this case, a mistake could cause $|w \cdot w^*|$ to drop. The $\gamma$-hinge-loss of $w^* = \sum_x \max[0, 1 - l(x)(x \cdot w^*)/\gamma]$ (by how much, in units of $\gamma$, would you have to move the points to all be correct by $\gamma$)

Prove: consider $|w \cdot w^*|$ and $||w||$
- Each mistake increases $|w \cdot w^*|$ by at least $\gamma$.
  - $(w + x) \cdot w^* = w \cdot w^* + x \cdot w^* \geq w \cdot w^* + \gamma$.
- Each mistake increases $w \cdot b$ by at most 1.
  - $(w + x) \cdot (w + x) = w \cdot w + 2(w \cdot x) + x \cdot x \leq w \cdot w + 1$.
- So, in $M$ mistakes, $\gamma M \leq |w \cdot w^*| \leq ||w|| \leq M^{1/2}$.
- So, $M \leq 1/\gamma^2$.

**Kernel functions**

See board...