Last time: More OLMB alg's for specific concept classes.

- OLMB alg. for learning DL's: $O(nr)$ mist. bound for learning 1-DL's of length r over $\{0,1\}^n$
- Winnow1 alg. for learning length-k monotone disj. over $\{0,1\}^n$ with $O(k \cdot \log n)$ mist. bound

LTFs

LTFs

Today:
- discuss Winnow 1
- Winnow 2 alg: OLMB alg. for certain LTF's (with positive weights, $X = \{0,1\}^n$)
- Perceptron alg. for LTFs "with a margin" over $\mathbb{R}^n$
- start dual Perceptron, kernelization

Questions?

Discuss. of W1:

- "brittle": set wts to $\underline{\underline{0}}$.
  Bad if noise...
  could have set $w_i \leftarrow \frac{w_i}{2}$.

- W1: uses LTF hyp.
  ↳ can we use to learn LTFs?

Not as is: poss. wts "too coarse" $\boxed{1, 2, 4, 8, 16, \ldots}$

e.g. $c(x) = \mathbb{1}\left[\sum_{i=1}^{n} i x_i \geq n\right]$ $\quad$ $x_1 + 2x_2 + \cdots + nx_n \geq n$

$\star$

More gentle updates... ?

Winnow2: variant which learns "δ-separable monotone LTFs"

$\downarrow$ pos wts

Def: Let $0<\delta<1$. The class $F(\delta)$ of over $\{0,1\}^n$ is the class of all LTFs $c:\{0,1\}^n \to \{0,1\}$ s.t.

$$\exists \text{ wts } u_1,\dots,u_n \geq 0 \text{ s.t. } \forall x \in \{0,1\}^n,$$

- $u \cdot x \geq 1$ & $c(x)=1$, or
- $u \cdot x \leq 1-\delta$ & $c(x)=0$.

Ex: "r-out-of-k threshold function":

$c(x)=1$ iff $x_{i_1}+\dots+x_{i_k} \geq r$.

i.e.

$$\underbrace{\tfrac{1}{r}(x_{i_1}+\dots+x_{i_k})}_{\text{either } \geq 1 \text{ or } \leq 1-\frac{1}{r}} \geq 1 : \quad \text{in } F(\tfrac{1}{r}).$$

Ex: ☆ is $\equiv$ to $\underbrace{\tfrac{1}{n}(x_1+2x_2+3x_3+\dots+nx_n)}_{\geq 1 \text{ or } \leq 1-\frac{1}{n}} \geq 1$

in $F(\tfrac{1}{n})$.

Winnow2: Requires init. param. $\alpha > 1$.

Hyp is $w \cdot x \geq n$ $\quad$ ($\Theta = n$, as before)

$w_{init} = (1, \ldots, 1)$ as before

Update rule: $h(x) = c(x) \Rightarrow$ no change

False pos: $\overset{\shortparallel}{1}$ $\overset{\shortparallel}{0}$ $\Rightarrow$ for each i s.t. $x_i = 1$, set $w_i \leftarrow w_i / \alpha$

False neg: $\overset{\shortparallel}{0}$ $\overset{\shortparallel}{1}$ $\Rightarrow$ for each i s.t. $x_i = 1$, set $w_i \leftarrow w_i \cdot \alpha$

Thm: Let $0 < \delta < 1$.
Let target $c \in F(\delta)$, & let $u_1, \ldots, u_n$ be wts as above.
If you run W2 with $\alpha = 1 + \frac{\delta}{2}$ on $\{0,1\}^n$ inputs labeled by c, then W2 makes

$$\leq \frac{8}{\delta^2} + \left(\frac{5}{\delta} + \frac{14 \ln n}{\delta^2}\right) \sum_{i=1}^{n} u_i$$

$$= O\left(\frac{\log n}{\delta^2} \cdot \sum_{i=1}^{n} u_i\right) \qquad \left(\sum u_i \geq 1\right)$$

mistakes.

Ex: r-out-of-k fn: $\delta = \frac{1}{r}$, $\sum_{i=1}^{n} u_i = \frac{k}{r}$.

Bd is $O\left(r^2 \cdot \ln(n) \cdot \frac{k}{r}\right) = O(k \cdot r \cdot \log n)$.

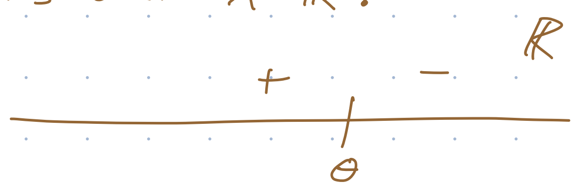
Ex 2: ✪ try yourself :)

Ex 3: Say c is a length-r OL.
Can show any " " " is a δ-sep. LTF for $\delta = \frac{1}{2^{\Theta(r)}}$.

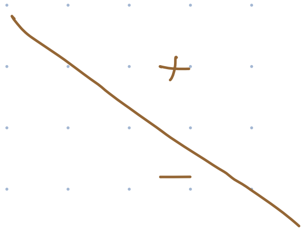
Work it out: W2 learns length-r OL with mist. bd. $2^{\Theta(r)} \cdot \log n$

Learning LTFs w/ the Perceptron Alg.
(mid-50's; Winnow: late 80s)

Perceptron: learns LTFs over $X = \mathbb{R}^n$.



Let target LTF be $\text{sign}(v \cdot x - \theta)$ $v \in \mathbb{R}^n$, $\theta \in \mathbb{R}$.



Assumption 1: $\theta = 0$.

This is wlog: given each $x \in \mathbb{R}^n$, build augmented $x' = (x_1, \ldots, x_n, 1)$

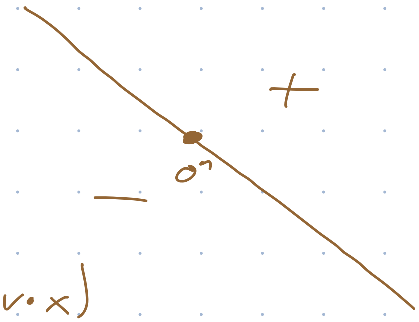
Learning $v \cdot x \geq \theta$ $(\mathbb{R}^n)$
" " "

|||

learning $\boxed{v' \cdot x' \geq 0}$ $\quad v' = (v_1, \dots, v_n, -\theta)$

↓ zero threshold.

New picture:



$\sqrt{x \cdot x}$

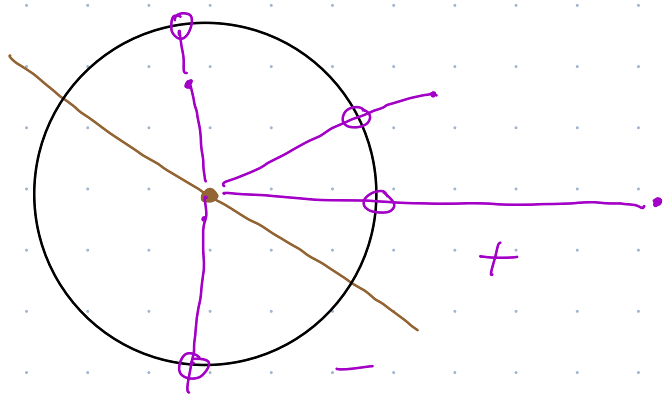
$\parallel$

$\sqrt{x_1^2 + \dots + x_n^2}$

Target: learn $\mathrm{sign}(v \cdot x)$

Assump. #2: each example x has $\|x\| = 1$.

WLOG! I can rescale any data pt x to become a unit vector, & this doesn't change how $\mathrm{sign}(v \cdot x)$ classifies x.



Assump. 3: v is unit vector, $\|v\| = 1$.
WLOG.

Here's Perc. alg:

Maintains hyp. vector $w \in \mathbb{R}^n$;

$$\mathrm{sign}(t) = \begin{cases} +1 & \text{if } t \geq 0 \\ -1 & \text{if } t < 0 \end{cases}$$

hyp. is $h(x) = \text{sign}(w \cdot x)$

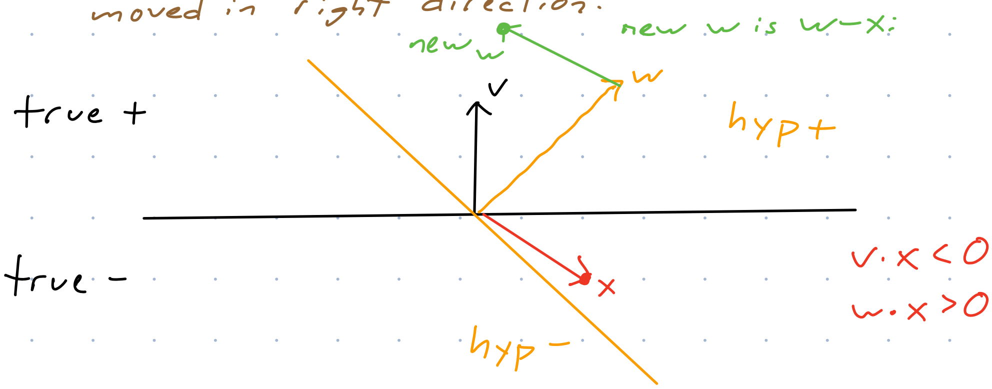
$w_{init} = (0, \dots, 0)$.

Update rule:

- if $h(x) = c(x)$, no change.
- $(w \cdot x \geq 0)$ $(v \cdot x < 0)$
  if $h(x) = 1$ & $c(x) = -1$: set $w \leftarrow w - x$
- $(w \cdot x < 0)$ $(v \cdot x \geq 0)$
  if $h(x) = -1$ & $c(x) = 1$: set $w \leftarrow w + x$.

→ when mistake, & $y = c(x) = \text{sign}(v \cdot x)$,
update is $w \leftarrow \underbrace{w + yx}_{w + \text{sign}(v \cdot x) x}$

Reasonable: e.g. on false pos $(\underline{w \cdot x > 0}, \underline{v \cdot x < 0})$:
new w is $w - x$; if got x again, now
new $w \cdot x$ is $(w - x) \cdot x = w \cdot x - x \cdot x = w \cdot x - 1$ ☺
moved in right direction.

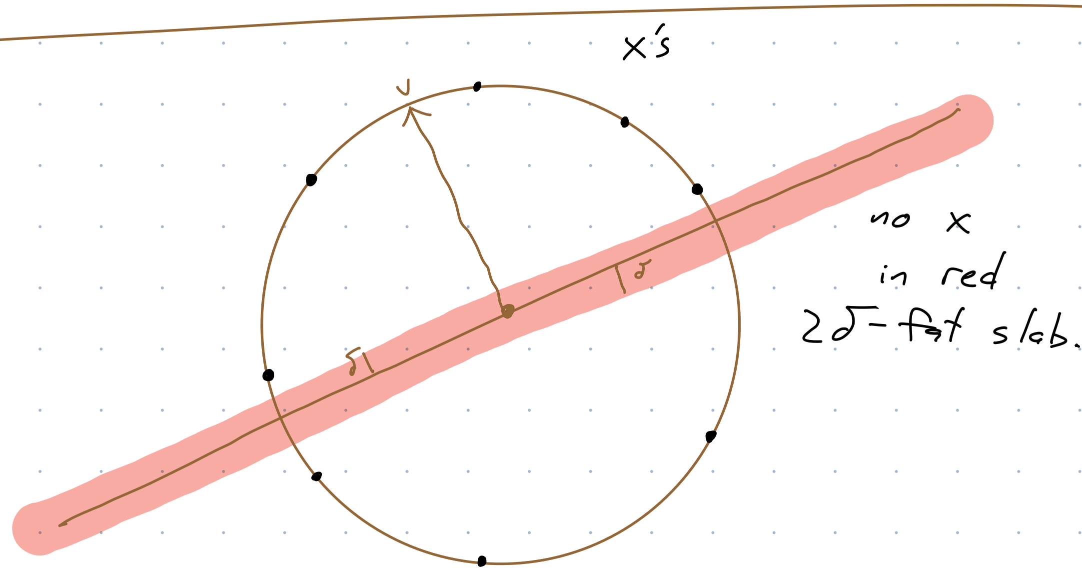


Thm: (Perc. Convergence Theorem):

Suppose run Perc. on seq of ex labeled by $\underbrace{\text{sign}(v\cdot x)}_{a1}$, where $\underbrace{\|v\|}_{a3}$, each $\underbrace{\|x\|}_{a2}$ are $= 1$.

Let $\delta := \min\limits_{\substack{\text{all} \\ \text{ex. } x \\ \text{we get}}} |v \cdot x|$. (= min Euclidean dist. from x to v's hyperplane)

Then # mist Perc. makes is $\leq \frac{1}{\delta^2}$.



2 lemmas. Track $v \cdot w$ + $w \cdot w$

L1: After M mistakes, $v \cdot w \geq \delta \cdot M$

Pf: $w_{init} = (0, \ldots, 0)$ so $v \cdot w = 0$.

Each mist. causes $v \cdot w$ to $\uparrow$ by $\geq \delta$:

w = old hyp vector

$w + \text{sign}(v\cdot x)x$ = new hyp vector

$$\underbrace{\big(w + \text{sign}(v\cdot x)x\big)\cdot v}_{\text{"new } w\text{"}\cdot v} = w\cdot v + \overbrace{\text{sign}(v\cdot x)\, v\cdot x}^{=|v\cdot x| \geq \delta}$$

$$\geq w\cdot v + \delta. \qquad \blacksquare$$

<u>L2:</u> After M mist., $w\cdot w \leq M$.

<u>Pf</u>: $w_{init} = \vec{0}$ so $w\cdot w = 0$ at start.

Each mist. causes $w\cdot w$ to $\uparrow$ by ≤ 1:

$w_{new} = w + \text{sign}(v\cdot x)x$ so after mist. on x,

$$\overset{\text{new}}{w\cdot w} = (\ \cdot\ ) = w\cdot w + \underbrace{2\,\text{sign}(v\cdot x)\, w\cdot x}_{\substack{<0 \\ \text{b/c} \\ \text{mistake!}}} + \underbrace{\big(\text{sign}(v\cdot x)\big)^2}_{=1} \underbrace{x\cdot x}_{\substack{=1 \\ \text{by} \\ \text{assump.}}}$$

$$\leq \overset{\text{old}}{w\cdot w} + 1. \qquad \blacksquare$$

Next time: use L1 + L2 to prove PCT :)
