Last time:
- intro, admin, overview, topic/course overview
- Basic notions: \( X \) (instance space)
  \[ c : X \rightarrow \{-1, 1\} \text{ concept} \]
- concept class \( \mathcal{C} \): set of concepts
- Examples: disj., conj., s-term DNF, \( k \)-CNFs
  \[ s \text{-clause CNF, } k \text{-DNFs} \]

Today:
- \( \rightarrow \text{LTFs} \) (linear threshold functions)
- \emph{OLMB model}
- Elimination alg. for monotone disj., variants
- Decision lists, OLMB learning alg. for them

Admin:
- Rosco OH tomorrow;
- next week: video lectures
- PS 1 out Tues

Questions?

Last concept class:
\[ X = \mathbb{R}^n, \mathcal{C} = \text{all linear threshold functions (LTFs, halfspaces) over } \mathbb{R}^n \]
\[ c : X \rightarrow \{-1, 1\} \text{ is an LTF of the form} \]
\[ c(x) = \text{sign}(w \cdot x - \Theta) \]
\[ w = (w_1, \ldots, w_n) \in \mathbb{R}^n, \Theta \in \mathbb{R}, \text{sign}(t) = \begin{cases} +1 & t > 0 \\ -1 & t < 0 \end{cases} \]
\[ \{ x : w \cdot x = \Theta \} \]

\[ w \]
\[ + \]
\[ \{ x : w \cdot x > \Theta \} \]
\[ x : w \cdot x \in \Theta \]

\[ \text{sign}(3.3x_1 - 0.6x_2 + 14x_3 - 5) \]

\[ C = \text{LTFSs over } \mathbb{R}^n : \| \mathbf{w} \| = \infty. \]

Can also study \( X = \{0, 1\}^n \), \( C = \) all LTFSs over \( \{0, 1\}^n \).

First learning model: \textit{OLMB learning (Online Mistake Bound)}

A "learning session" consists of a sequence of trials. Throughout a learning session, learner maintains a hypothesis \( h : X \rightarrow \{0, 1\} \).

Learning a concept \( c \in C \) goes like this: in each trial,

1) learner is given an unlabeled \( x \in X \);
2) "outputs \( h(x) \in \{0, 1\} \);
3) "is given true value \( c(x) \in \{0, 1\} \).

If \( h(x) \neq c(x) \), learner is charged a mistake.

\( \bigcirc \) Before next trial, learner may update \( h \).

(\underline{This update rule defines the learning alg.})
No noise as defined above

Performance measure: total # mistakes made.

**Definition:** A learning alg. $A$ has mistake bound $M$ for concept class $C$ if: for any sequence of examples from $X$, any target concept $c \in C$, $A$ makes $\leq M$ mistakes on any seq. of trials as above. (w/ bound on # trials)

$c(11011) = 1$

$c : X \rightarrow \{0,1\}$

Easy:
- if $X$ finite, always poss. to achieve $M = |X|!$ (memorization).
- if $C$ finite, always poss. to achieve $M = |C|! - 1$ (try concept in $C$ until mistake; discard & try another; etc.)

**Ex:**

$X = \{0,1,2,\ldots,2^n-1\}$

$C =$ initial intervals

$c = \{0,1,2,\ldots, a\}$

$\not\in C \quad \forall a \in X$

$|X| = 2^n$, $|C| = 2^n$

Can achieve MB of $n$ via binary search:
Alg: use as h.p. \{0,1,...,b\} where b = midpt of "uncertainty interval" (largest +, smallest -)
initially b = 2^{-1}

Each mist. cuts uncertainty interval by at least \frac{1}{2}, so at most \log_2(2^n) = n mistakes.

Ex: X = [0,1] real interval
    C = init. intervals of [0,1]
    c = [0,0.5] \forall \epsilon \in [0,1]:

no finite mist. bound

Some Specific OLMB Algs for various Cs

Learning Disjunctions \( \checkmark \) \( \text{\textasciitilde} \text{\textasciitilde} 50 \text{\textasciitilde} \text{\textasciitilde} \)
Recall a disjunction is something like $c(x) = x_1 v x_2 v x_3 v x_4 v x_5$.

$C = \text{mon. disj.}$

OCMB alg. for $\text{"elimination algorithm"}$

- **Initial hyp**: $h(x) = x_1 v x_2 v \ldots v x_n$ (all vars)

- **When false pos mistake** $\left[ h(z) = 1 \text{ but } c(z) = 0 \right]$
  
  on example $z$:
  remove from $h$ all $x_i$ s.t. $z_i = 1$ in that example.

- **When false neg mistake** $\left[ h(z) = 0 \text{ but } c(z) = 1 \right]$
  
  on example $z$:
  stop and say FAIL (won't happen)

- if $h(z) = c(z)$ (no mistake): keep $h$ as is.

---

*Example*: say $n = 5$, $h(x)$ init. is $x_1 v x_2 v \ldots v x_5$.

- Get $z = 01001$; $h(z) = 1$

  Given $c(z) = 0$: remove $x_2, x_5$ to get

  new $h(x) = x_1 v x_3 v x_4$.

We'll show: mist bound $\leq n$ for this alg, for learning any mon. disj. $c$. 
Claim 1: no var that's in $c$ is ever removed from $h$.

PF: var $x_i$ only removed from $h$ if get $z \text{ s.t. }$
$z_i = 1, c(z) = 0, h(z) = 1; \text{ but means } x_i \text{ not in } c.$

Claim 2: alg only makes false pos mistakes

PF: by prev claim, $h$ always contains all vars in $c$.
So if $c(z) = 1$, also have $h(z) = 1$. So no false neg.

Thm: The elim alg has mistake bound $\leq n$
for $C = \text{mon disj over } \{0, 1\}^n$.

PF: Alg doesn't fail: \{vars in $h$\} always contains \{vars in $c$\}.
$h$ starts w/ $n$ vars; each mistake gets rid of \geq 2 var not in $c$,
and initially has $\leq n$ vars not in $c$.
So $\leq n$ mistakes.

Obs: This is a good obj:
$|X| = 2^n, |E| = 2^n, \text{ m.b. } \leq n$

Runtime per trial also $O(n)$

* This alg. would not be noise-robust.

* Easy extension to non monotone (general)
\text{disj : init } b = x, v \overline{x}, v x_2 v \overline{x}_2 \ldots v x_n v \overline{x}_n \}
\quad \text{m.b. } \leq 2^n \\
\text{Run same alg.}
\quad n=4
\text{First mistake: } z = 1010 \quad b(z) = 1
\quad c(z) = 0: \text{ cross off } x_i, \overline{x}_2, x_3, \overline{x}_4
\quad \text{so tot. mist. bound } \leq n+1.

\begin{itemize}
  \item Can learn mon. conj., non-mon. conj.
    similarly, or by just negating each \( c(z) \):
    \[
    \text{not } (x_1 \wedge \overline{x}_3 \wedge x_5) = \overline{x}_1 \vee x_3 \vee \overline{x}_5.
    \]
\end{itemize}

\underline{Learning \ I- \ decision \ lists}

\( X = \{0,1\} \)

\text{a I-decision list: a Bool fn } \{0,1\} \rightarrow \{0,1\}
\text{that's an ordered list of if-then rules of the form:}

\[
\begin{align*}
\text{if } \ell_i & \quad \text{then output } b_i \\
\text{else } & \quad \ell_2 \quad \text{else output } b_{i+1}.
\end{align*}
\]

\text{c(x): } \overline{X} \rightarrow x_4 \rightarrow \overline{x}_2 \rightarrow x_7 \rightarrow 0
\( z = 1010101 \) : \( c(z) = 2 \).

- Every conj. or disj. can be expressed as a 1-DL:
  \[ X_1 \lor \overline{X}_3 \lor X_4 \]
  \[ X_1 \rightarrow \overline{X}_3 \rightarrow X_4 \rightarrow 0 \]
  \[ \overline{0} \]
  \[ \overline{1} \]

- Length of a DL: \# of literals in it = 3

- WLOG, any DL has each var at most once.

  if reach here, \( x_3 = 0 \)

  \[ \ldots X_3 \ldots \rightarrow \overline{X}_3 \]
  \[ \downarrow \]
  \[ 0 \]
  \[ 1 \]

Next time: \( O(nr) \) for length-\( r \) 1-DLS over \( \{0, 1\}^3 \).