Last time: quick intro to CLT: admin overview of learning models:
Key players: \( X, c, E; \) \( K \)-ONF, \( s \)-term ONF

\( \text{lit, conj, LTFs.} \rightarrow \mathcal{F}(\mathbb{R}^n \rightarrow \{0,1\}) \) is a LTF if
\( \exists w \in \mathbb{R}^n, \theta \in \mathbb{R} \) s.t. \( \{0,1\} \)
\( \forall x, f(x) = 1 \) if \( w \cdot x \geq \theta \), \( = 0 \) if \( w \cdot x < \theta \).

Today: Define OCMB model! "
 Few specific alg's: simple "init. int."; conj. (elim.); alg. for 1-decision lists.

Readings: Blum survey (sec. 3).

- Next week: watch lectures online.
- My OH: Fri 9-11. HW1: avail. today!

Online Mistake-Bound Learning Model

The life experience of a learner is a sequence of trials. Throughout process, learner always maintains hypothesis \( h: X \rightarrow \{0,1\} \).

A trial:
- Learner is given an \( x \in X \).
- "outputs \( h(x) \) (in \( \{0,1\} \))".
- "is told true value of \( c(x) \)."
If \( h(x) \neq c(x) \), learner is charged a mistake.

\( \star \) Learner can update \( h \) before next trial.

- No noise; learner always gets true \( c(x) \).
- No missing data.

We assess performance of a \( h \) by counting # mistakes on its seq. of examples.

Def: A learning alg. \( A \) in OCMB model has mistake bound \( M \) for \( c \in C \) if:

- For any target concept \( c \in C \) and any seq. of examples from \( X \), \( A \) makes \( \leq M \) mistakes.

Note:

- Sparsity \( |X| \) finite. There's an alg. \( A \) w/ MB \( M \leq |X| \) (memorization.)
- If \( |C| \) finite, \( \exists \) alg. \( A \) w/ MB \( \leq |C|-1 \) (try all concepts in \( C \)).

Example: \( X = \{0, 1, 2, \ldots, 2^n-1\} \)

- \( C \) = class of all initial intervals; \( 1 \in \leq 2^n \)
- e.g. \( c \in \{0, 1, 2, \ldots, 16, 17\} \).
- \( |X| = 2^n \)

Bin. search gives alg. \( A \) w/ MB \( \bigcirc \).

- Init. hyp. \( \{0, 1, \ldots, 2^n-1\} \)
- Update rule:
  - If \( h(x) = c(x) \), no change.
  - If \( h(x) \neq c(x) \), adjust.
h.s.t. new r+ endpt is in middle of current "uncertainty region."
Each mist. cuts & by at least 1/2, so after n mist. its length ≤ 1, so \( h \in \mathbb{C} \).

Tweak of above: \( X = [0, I] \subseteq \mathbb{R} \),
\( C = \text{init. intervals} \); e.g. \( C = [0, a] \subseteq \mathbb{R} \).
No finite MB for \( C \). (Any finite set of ex. can't specify \( C \) fully).

Next \( \approx 2 \) lect: some specific OC MB algs for specific \( C \) 's.

- Elim. alg. For disj. (disj. \( = 0 \cap \) \( X_3 \lor X_2 \lor X_6 \))
  - 1-OC
  - Sparse disj. (Winnow), Certain LTFs (Winnow)
  - Perceptron for LTFs (Kernel)

- Elim. Alg. For monotone disj.

\( C = \text{all} \); e.g. \( C(x) = X_2 \lor X_4 \lor X_6 \lor X_8 \)
\( X = \{0, 1\}^n \) \( \Rightarrow \) \( |I| = |X| = 2^n \)
The alg: Init. hyp \( h(x) \equiv x \lor X_2 \lor \ldots \lor X_n \).
Update rule: If \( h(x) = c(x) \), no update.
If \( h(x) = 1 \) but \( c(x) = 0 \) (false pos) : erase from \( h \) each \( x \); that was 1 in \( X \).
If \( h(x) = 0 \) but \( c(x) = 1 \) (false neg.) : output \text{FAIL}. 


Ex: Say \( n = 5 \). \( h_{\text{init}} \equiv x, vx_2 \ldots vx_5 \). Get \( ex = 01001 \). \( h_{\text{init}}(x) = 1 \). Told \( c(x) = 0 \).

Update \( h \) to \( x, vx_3, vx_4 \).

Claim: Any var \( x_i \) that's in \( c \) is never removed from \( h \). (If \( x_i = 2 \) in an \( ex \), \( c(x_i) = 2 \) \( x \) won't cause alg. to erase it.)

Claim: Alg never makes f.m. mistake given that \( c \) is a mon. disj. (\( h \) always includes all var \( s \) in \( c \), so if \( c(x) = 2 \), then \( h(x) = 2 \).)

Theorem: Elim. alg. has \( MB \leq n \).

PF: See \( h \) always incl. all vars in \( c \).
Each mist: f.p. \( h(x) = 1 \), so some \( x \) not in \( c \) was \( 1 \) in \( x \) \& that \( x \) is in \( h \); so each mist. removes \( \geq 1 \) \( x \) from \( h \). \( h \) init. has \( n \) vars, \( \geq \) so \( \leq n \) mistakes.

Notes: · MB \( n \) for Cover \( 0,1 \): \( n \)
    poly \((n)\)
· Runtime per trial, per update \( 0(n) \).
· We assumed no noise; noise DESTROYS this.
· Extends to \( c = \) all \( \bar{c} \)
    (poss. non-mon.) easily.
\( 2^n \) entries.
\[ h_{\text{init}} = x_1 \lor \overline{x_1} \lor x_2 \lor x_2 \ldots \lor x_n \lor x_n \]

as before.

1st \text{mist}: eliminates \( n \) of

\[ x = 10010 : \ c(x) = 0, \ h_{\text{init}} = 1 \]

as before.

* Can use same idea to learn conj.

Or, can negate each label!

if \( c(x) = x_2 \lor \overline{x_3} \lor \overline{x_4} \),

\[ c(x) = \overline{x_2} \land x_3 \land x_4 \]

\[ \text{Decision Lists (2-OL)} \]

A 2-OL: a func \( c : \{0,1\}^n \rightarrow \{0,1\} \)

\[
\begin{align*}
\text{if } l_1 \text{ then output } b_1 \\
\text{else if } l_2 \text{ then output } b_2 \\
\text{else } \ldots \text{ } \\
\text{else if } l_r \text{ then output } b_r \\
\text{else output } b_{r+1}
\end{align*}
\]

(K-OL: if \( C_i \) then \( b_i \) else...

each \( C_i \) is a conj of \( k \) literals)
Example: \[ X_7 \rightarrow X_4 \rightarrow X_2 \rightarrow X_5 \rightarrow 0 \]

is a 1-OC. Length 4

Next alg: OCMB alg for $C = \{ \text{all length-} r \text{ 1 OC's over } \{0,1\}^n \}$ with MB $O(n \cdot r)$. 