Last time:
- "Chernoff bounds" / tail bounds (for hypoth. testing) (CHF)
- learning by finding consistent hyp. from a fixed $\mathcal{H}$
- "Occam's Razor" ("cardinality" version)

Today:
- Applic. of ↑: learning sparse disj., using few examples, with hyp. that's a (pretty) sparse disj.
  → via "greedy set cover" heuristic
- (start) proper vs improper learning:
  - ☺ can efficiently improperly learn 3-term DNF;
  - ☹ can't eff. properly learn 3-term DNF (unless $NP \subseteq RP$).

PS 2 due, PS 3 out.

Midterm Mon Oct 23 (on mat. thru Mon Oct 16)

Questions?

Thm "Occam's razor, cardinality version":

Fix $m, \mathcal{C}, \mathcal{H}$. Suppose there's a subset $\mathcal{H}_m \subseteq \mathcal{H}$ s.t. $\boxed{m \geq \frac{1}{\varepsilon}\left(\ln|\mathcal{H}_m| + \ln\frac{1}{\delta}\right)}$ ⟷ $|\mathcal{H}_m| \leq \delta e^{\varepsilon m}$

s.t. given any set S of m lab. ex. acc. to some $c \in \mathcal{C}$, there's an $h \in \mathcal{H}_m$ consistent with the m ex.

If L is any alg. that finds such an $h \in \mathcal{H}_m$ when given a sample of size m, then L run on m ex from $EX(c, \mathcal{D})$ is an (ε, δ)-PAC learner.

Applic.: Using "greedy set cover" heuristic to eff. learn sparse disjunctions.

$\mathcal{C}$ = {all mon. disj. of length $\le k$ over $x_1, \ldots, x_n$}
$X = \{0,1\}^n$

Winnow: OLMB $\underbrace{O(k \log n)}^{=z}$ mist.
OLMB $\rightarrow$ PAC conv: $\frac{z}{\varepsilon} \ln\left(\frac{z}{\delta}\right)$ = # ex is enough;

LTF hyp "n"

Now: use Occam stuff to $\overset{\text{PAC}}{\text{learn}}$ with $\approx$ samples, now using "pretty short" mon disj as hyp's.

Idea: we'll show how, given $\underline{m}$ ex, we can eff find a (length s) mon disj consistent with our m ex.

$s \approx k \cdot \ln m$

"greedy set cover" heuristic

$\mathcal{H}_m$ = all length$-s$ mon disj
$|\mathcal{H}_m| \le n^s$: so as long as

$n^s \le \delta e^{\varepsilon m}$, can apply Occam result.

$$\boxed{n^{k \ln m} \le \delta e^{\varepsilon m}}$$

small m will work! :)

Consider following problem:

SET COVER:

Input: family of r sets $S_1, \ldots, S_r$; each $S_i \subseteq [m]$.

Output: use min # of S_i's to cover all of $[m]$.

OPT = size of optimal solution

Ex: $m = 7$

$S_1 = \{1, 2, 3, 4, 5\}$
$S_2 = \{1, 3, 5, 7\}$
$S_3 = \{2, 3, 4, 6\}$
$S_4 = \{3, 4, 7\}$

opt sol: S_2, S_3

:(NP-complete: almost certainly, there's no poly-time alg. that always finds opt. solution

:) The greedy alg/heuristic is efficient & provably finds an almost-optimal solution!

GH:

- First step: Pick S_i s.t. $|S_i|$ largest. Delete all elts in S_i from every S_j.
- Repeat using new S_j's. Stop when all $[m]$ covered.

Ex of GH:

$S_1 = \{1, 2, 3, 4, 5\}$ ← first set
$S_2 = \{\cancel{1}, \cancel{3}, \cancel{5}, \cancel{7}\}$
$S_3 = \{\cancel{2}, \cancel{3}, \cancel{4}, 6\}$
$S_4 = \{\cancel{3}, \cancel{4}, \cancel{7}\}$

can choose S_2, S_3 or S_4 as second set; say S_2

$S_3 = 3^{rd}$ set

Thm: Greedy heuristic always finds a sol. that contains $\le \mathrm{OPT} \cdot \ln m$ sets.

Pf:

Fact: Consider any $U \subseteq [m]$. ("uncovered" elts)

There's some S_j, $1 \le j \le r$, covering $\ge \frac{|U|}{\mathrm{OPT}}$ points in U.

[Some OPT many S_i's cover $[m]$, hence cover U, so at least one of them covers $\ge |U|/\mathrm{OPT}$ elts.]

Let $U_i \subseteq [m]$ be set of elts not yet covered after i stages.

FACT $\Rightarrow$ next step covers $\ge |U_i|/\mathrm{OPT}$ elts of U_i

(whichever S_j of orig. sets covered this many of U_i's elts still covers them; uncovered elt's don't get removed as we go.)

So $$|U_{i+1}| \le |U_i| - \frac{|U_i|}{\mathrm{OPT}} = |U_i|\left(1 - \frac{1}{\mathrm{OPT}}\right).$$

So $$|U_i| \le m \cdot \left(1 - \frac{1}{\mathrm{OPT}}\right)^i$$

$\left(\left(1 - \frac{1}{x}\right)^x < e^{-1} \text{ for } x > 1\right)$

$i = \mathrm{OPT} \cdot \ln m$: $$\le m \cdot \left(1 - \frac{1}{\mathrm{OPT}}\right)^{\mathrm{OPT} \cdot \ln m}$$

$$< m \cdot e^{-\ln m} < 1$$

so $|\check{U}_i| = 0$

Learning again: why is this useful for learning sp. mon. disj.?

It'll let us find, given a set S of m ex lab. by a mon disj of size k, a mon disj of length $\le k \cdot \ln m$ that's consistent.

<u>The alg:</u>

1. <u>Run elim alg on sample</u>.
   (start w/ $x_1 \vee \dots \vee x_n$; for each neg ex in S, eliminate all x_i that are 1 in the ex.)
   This gives a mon disj h_1 consistent w/ the sample S; say it's $x_1 \vee \dots \vee x_r$.
   (Note: any sub-disj of $\hookrightarrow$ is consistent w/ neg ex.)

2. We want to find a small sub-disj. of $x_1 \vee \dots \vee x_r$ consist. w/ all the (at most m) <u>pos</u> ex.
   I.e. want to find small subset of $x_1, \dots, x_r$ covering all m pos ex.
   (x_i covers j^{th} pos ex $\iff$ $x_i = 1$ in that ex.)

This is SET COVER!

x_i ; pos example i

$\updownarrow$ $\updownarrow$

S_i elt $i \in [m]$

Run greedy set cover heur. on the pos. ex.

$\rightarrow$ Finds a coll. of $\leq OPT \cdot \ln m$ many vars x_i

Have $OPT \leq k$ by def. of $\mathcal{C}$.

So... alg has prop. that, given a set S of m ex. lab. by a mon disj of size k, it finds a mon disj of size $\leq k \cdot \ln m$ that's consistent.

So $\mathcal{H}_m$ = all mon disj over $x_1, \ldots, x_n$ of size $\leq k \cdot \ln m$.

$\hookrightarrow$ $|\mathcal{H}_m| = \binom{n}{0} + \cdots \binom{n}{k \cdot \ln m} \leq n^{k \cdot \ln m}$.

So as long as $\underline{\underline{|\mathcal{H}_m|}} \leq \underline{\underline{\delta e^{\varepsilon m}}}$ holds, :) PAC learner using m ex.

How big does m have to be to give us

$n^{k \cdot \ln m} \leq \delta e^{\varepsilon m}$?

Taking

$$m \geq 2\left(\frac{1}{\varepsilon}\cdot k\cdot \ln n\cdot \ln\frac{1}{\delta}\right)\cdot \ln\left(\frac{1}{\varepsilon}\cdot k\cdot \ln n\cdot \ln\frac{1}{\delta}\right)$$

φ

is enough.

take ln:

B/c: $\boxed{k\cdot \ln m\cdot \ln n \leq \ln\delta + \varepsilon m}$

Need $k\cdot \ln m\cdot \ln n + \ln\frac{1}{\delta} \leq \varepsilon m$

$A+B \leq AB$ for $A, B \geq 2$, so suff to have

$k\cdot \ln m\cdot \ln n\cdot \ln 1/\delta \leq \varepsilon m$, i.e.

suff to have $\boxed{\underbrace{\frac{k\cdot \ln n\cdot \ln 1/\delta}{\varepsilon}}_{A} \leq \frac{m}{\ln m}}$

$A \leq \frac{m}{\ln m}$: m ?

$m = A\cdot \ln A$? not quite good enough:

$$A \overset{?}{\leq} \frac{A\cdot \ln A}{\ln(A\ln A)}$$

not true

frac < 1 !!

But

$m = 2A \cdot \ln A$?

$$A \overset{?}{\leq} \frac{2A\ln A}{\ln(2A\ln A)} \; ?$$

Yes:

$$= \frac{A \cdot \ln(A^2)}{\ln(2A\ln A)}$$

frac > 1

!! :)

Proper vs Improper Learning

Sometimes insisting on <u>proper</u> learning makes learning hard!

$\mathcal{C}$ = all 3-term DNFs:

$T_1 \vee T_2 \vee T_3$

T_i = some conj.

eff. PAC Learnable?

Good news: Y, using $\mathcal{H}$ = all 3-CNFs.

Fact 1:

- Recall: The class of all mon. conj. over N vars is PAC learnable with

$$m = O\left(\frac{1}{\varepsilon}\left(N + \ln\frac{1}{\delta}\right)\right) \text{ ex, in}$$

poly$(N, \frac{1}{\varepsilon}, \ln 1/\delta)$ time. (Elim alg, CHF).

Fact 2: Any 3-term DNF is logically $\equiv$ to a 3-CNF (each clause has ≤ 3 vars)

Pf: De Morgan:

$$(abcd) \vee (ef) \vee (gh) \equiv$$

$$(a \vee e \vee g) \wedge (a \vee e \vee h) \wedge (a \vee f \vee g) \wedge (a \vee f \vee h) \wedge$$
$$(b \vee e \vee g) \wedge \ldots$$

Next time: