Last time:
- "Chernoff bounds" (tail bounds on sums of indep. random variables) + applic. to hypothesis testing
- learning by finding consistent hyp. from a fixed class $\mathcal{H}$ "CHF"
- "Occam's Razor" ("cardinality version")

Today:
- Applic. of ↓: learning sparse disj., using few examples, with a hypothesis that is a pretty sparse disj, via "greedy set cover heuristic"
- start proper vs. improper learning:
  - can efficiently improperly learn 3-term DNF;
  - can't " properly " " ",
  unless NP=RP

Admin: PS 2 due, PS 3 out

Rocco OH this week:

Wed 4-6 CEPSR 620

Questions?

Let's learn (PAC) $\mathcal{C}$ = { all length-k mon disj.} over $X = \{0,1\}^n$!

Winnow: OLMB alg with $M = O(k \cdot \log n)$ mist.

Conversion $\Longrightarrow$ PAC alg. w/ $O\left(\frac{M}{\varepsilon} \cdot \ln\left(\frac{M}{\delta}\right)\right)$ ex. ☺

☹ Hyp. is weird; an LTF.

? Can we do this with mon. disj. hyp ?

(Y) via thm from last time:
we'll give an algorithmic way to find a consist. hyp., given m ex., from

$$\mathcal{H}_m = \{\text{all mon conj. over } x_1, \dots, x_n \text{ of length} \le k \cdot \ln m\}.$$

The alg. engine: "greedy heuristic" for Set Cover.

Here's a problem:

<u>SET COVER</u>

<u>Input:</u> family of r sets $S_1, S_2, \dots, S_r$
each $S_i \subseteq \underbrace{\{1, \dots, m\}}_{\text{"universe"}}$

<u>Goal:</u> Use min # of S_i's to <u>cover</u> all elts of universe.

<u>Ex:</u> $m = 7$

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

S_2, S_3 is optimal solution

<u>Culture:</u> This is an "NP-complete" problem: "¬"

almost certainly no poly-time alg., to find optimal sol.

☺: There's a highly eff. alg. to find a sol. that's provably not much worse than opt: "greedy heuristic".

GH:

First step: • Take the S_i s.t. $|S_i|$ largest.
• Delete all elts in S_i from each remaining S_j (update S_j's)

Next step: Repeat (using new S_j's)

Do this until all of $\{1, \ldots, m\}$ covered.

Ex: $m = 7$

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

GH: • first set it chooses is S_1.
• Update:
$S_2 = 7$
$S_3 = 6$
$S_4 = 7$
• Pick S_3; update
Pick S_4; done.

Found S_1, S_3, S_4 : sol. size 3, not 2.

Claim: Let OPT = size of optimal sol. (# sets).
GH always finds a sol. of $\leq \text{OPT} \cdot \ln(m)$ sets.

Pf:

Fact: Let $U \subseteq \{1, \ldots, m\}$ be a set of "uncovered" elts.
There must be some S_i (some $i \in [r]$) that covers $\geq |U|/OPT$ elts of U.

Pf: Some coll. of OPT many S_i's covers all of $[m]$, hence covers U; so one of them covers $\geq |U|/OPT$ elts of U, or else together couldn't cover U.

Consider execution of GH.
Let $U_i = \{$elts of $[m]$ that are uncovered after stages $1, \ldots, i\}$.

$U_0 = [m]$

Next ($(i+1)$st) step: covers $\geq |U_i|/OPT$ elts of U_i.
[By FACT above, some original S_j originally covered $\geq \frac{|U_i|}{OPT}$ elts of U_i -- + still does; updates didn't affect uncovered elts.]

So

$$|U_{i+1}| \leq |U_i| - \frac{|U_i|}{OPT} = |U_i| \cdot \left(1 - \frac{1}{OPT}\right).$$

$$|U_0| = m, \text{ so } |U_i| \leq m \cdot \left(1 - \frac{1}{OPT}\right)^i.$$

$$i = OPT \cdot \ln(m): \quad |U_i| \leq m\left(1 - \frac{1}{OPT}\right)^{OPT \ln m}$$

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

$x \geq 1$

$$< m \cdot \left(\frac{1}{e}\right)^{\ln m} = 1$$

so $U_{OPT \cdot \ln m} = 0$. :)

Back to Learning!

We'll give an alg which, given a set S of m ex. labeled by a length-k mon disj, outputs a "small" consist. mon. disj.

$\leq k \cdot \ln m$

(So $\mathcal{H}_m$ will $= \{$all $\}$; $|\mathcal{H}_m|$ will be $< \delta e^{\epsilon m}$ for suitable m, + done.)

The learning alg:

1) Run elim alg. on our m-ex sample:
(start w/ $h = x_1 \vee \dots \vee x_n$, for each neg ex z, if $z_i = 1$ remove x_i from h).
This gives a mon disj h_1 consist. w/ examples; wlog this h_1 is $x_1 \vee \dots \vee x_r$.
Note: any sub-disj. of $x_1 \vee \dots \vee x_r$ is cons. with the neg. ex. in the sample.

---

2) Goal: find small sub-disj. of $x_1 \vee \dots \vee x_r$ that's consist. w/ all the pos examples.
I.e. want small sub-collection of $x_1, \dots, x_r$ that "covers" all (at most m) pos ex in sample.
x_i covers j^{th} pos ex $\iff$ $x_i = 1$ in that example.
This is a set cover problem!

$x_i \longleftrightarrow$ set S_i played

pos ex $j \longleftrightarrow$ elt $j \in [m]$.

↳ Use GH to find a coll of $\leq \underline{OPT \cdot \ln(m)}$

x_i's to cover all pos ex.

$OPT \leq k$ b/c target is a mon disj of length $\leq k$.

So we have an eff alg. which, given a sample of m ex lab. by length-k mon disj, finds a consistent $h \in \mathcal{H}_m := \{$ length at most $k \cdot \ln m$ mon disj$\}$.

Have $|\mathcal{H}_m| \leq n^{k \ln m}$.

So as long as m satisfies

$$(\star) \quad \boxed{n^{k \ln m} \leq \delta e^{\varepsilon m},}$$

using a sample of size that m + running above alg. is an (ε, δ) PAC learner.

Verify:

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

suffices to ensure $\star$. Why?

Need $\star$, i.e. $(\ln)$

$k \cdot \ln(m) \cdot \ln(n) \leq \ln \delta + \varepsilon m$, i.e.

$$\underline{k \cdot \ln(m) \cdot \ln(n) + \ln \tfrac{1}{\delta} \leq \varepsilon m}$$

$$\boxed{\begin{array}{c} a, b \geq 2 \\ \Downarrow \\ a + b \leq ab \end{array}}$$

So suff for m to sat.

$$\underline{\underline{k \cdot \ln(m) \ln(n) \cdot \ln(1/\delta)}} \leq \varepsilon m, \text{ i.e.}$$

$$\boxed{\underbrace{K \cdot \ln(n) \cdot \ln\left(\tfrac{1}{\delta}\right)}_{A} \leq \frac{m}{\ln m}}\ .$$
ε

Want as small an m as poss, s.t.

$$\boxed{A \leq \frac{m}{\ln m}}$$

$m = e^A$?

$RHS = \frac{e^A}{A} \ggg A$

Try something:

would $m = A$ work ?

No: $A \overset{?}{\leq} \frac{A}{\ln A}$ false !!

would $m = A \ln A$ work?

no $A \overset{?}{\leq} \frac{A \ln A}{\ln(A \ln A)}$ <1 so didn't work...

would $m = 2A \ln A$ work ?

YES! $A \overset{?}{\leq} \frac{2A \ln A}{\ln(2A \ln A)} = A \frac{\ln(A^2)}{\ln(2A \ln A)}$

>1 + worked.

Proper vs Improper Learning

Take-home message: learning problems which are easy if you use the right $\mathcal{H}$ can become hard if " " " wrong $\mathcal{H}$.

$\mathcal{C}$ = all 3-term DNFs $T_1 \vee T_2 \vee T_3$ over $X = \{0,1\}^n$.

Q: Does there exist a poly-time alg to PAC learn $\mathcal{C}$?

① Y, if we use a different $\mathcal{H}$:

$\mathcal{H}$ = all 3-CNF formulas;

② N if we insist on proper learning.

---

①: Recall "de Morgan"'s law: for T/F vars a, b, c, d, have

$$\underbrace{(a \wedge b)}_{} \vee \underbrace{(c \wedge d)}_{} \equiv (a \vee c) \wedge (a \vee d) \wedge (b \vee c) \wedge (b \vee d).$$

LHS T: T, so c, d both T ✓

LHS F: F + F; say a, c both F: F.

Next time: use this for learning.