Last time:

$\prod_{t=1}^{T} \sqrt{1-4\gamma_t^2}$

- analysis of AdaBoost ⇝ state & prove thm about its performance
- start unit on PAC learning in the presence of noise
  - general framework, partic. noise models
    - malicious noise
    - RCN

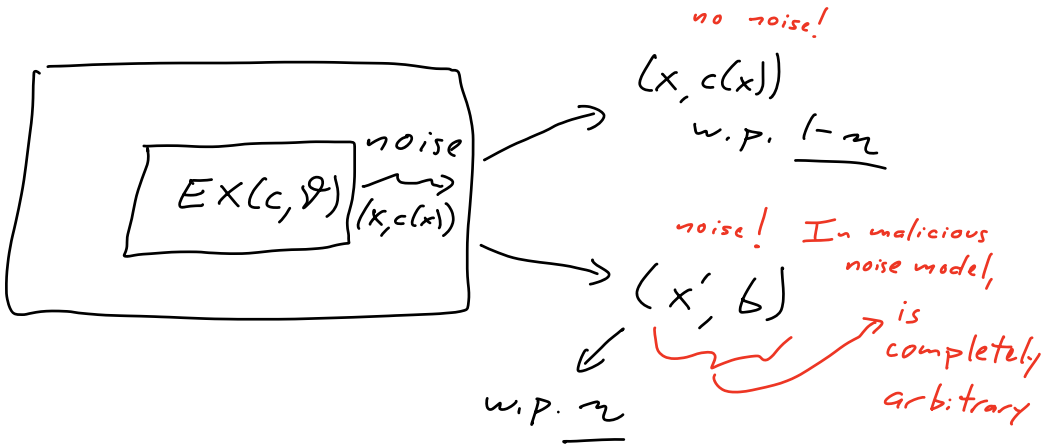
    } η = noise rate
  - toy scenario: can't handle mal. noise at rate $\frac{1}{5}$,
    can " RCN at any $\eta < \frac{1}{2}$.

Today:

- Lower bound: Learning with mal. noise is hard
- (sad) pos. result for) | | /
- start learning with RCN

Questions?

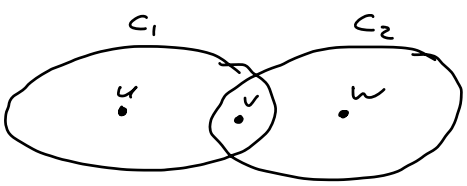
Recall malicious noise setup:



Let's show an information-theoretic lower bound on

PAC
error achievable by any learner in presence of mal. noise.

Def: Let's say $\mathcal{C}$ is *distinct* if
it contains c_1, c_2 $\forall$ X contains u, v, w s.t.



$c_1(u) = 1$
$c_1(v) = 1$
$c_1(w) = 0$

Thm: Let $\mathcal{C}$ be a distinct concept class.
Suppose mal noise rate η is $\geq \frac{2\tau}{1+2\tau}$.
Then no alg can PAC learn (with
to error $< \tau$ $\forall$ conf. $1-\delta > \frac{1}{2}$.

Pf: Consider only c_1, c_2.
Define dist $\mathcal{D}$, adversaries A_1, A_2.
Show if $\eta = \frac{2\tau}{1+2\tau}$, then
to learner,
$\mathcal{D}$ perfectly indistinguishable from $\mathcal{D}$
$c_1 \vee A_1$ $c_2 \vee A_2$.

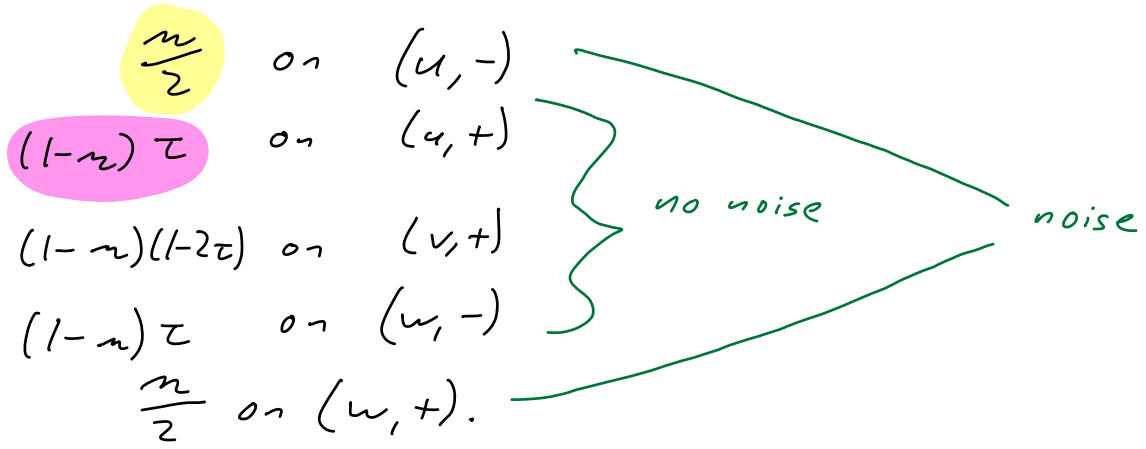
plan

Details: $\mathcal{D}$ puts τ on u
$1-2\tau$ on v.
τ on w

Adversary A_1: of the adv's η "noise budget":

$\frac{n}{2}$ on $(u,-)$, $\frac{n}{2}$ on $(w,+)$. } sum to n

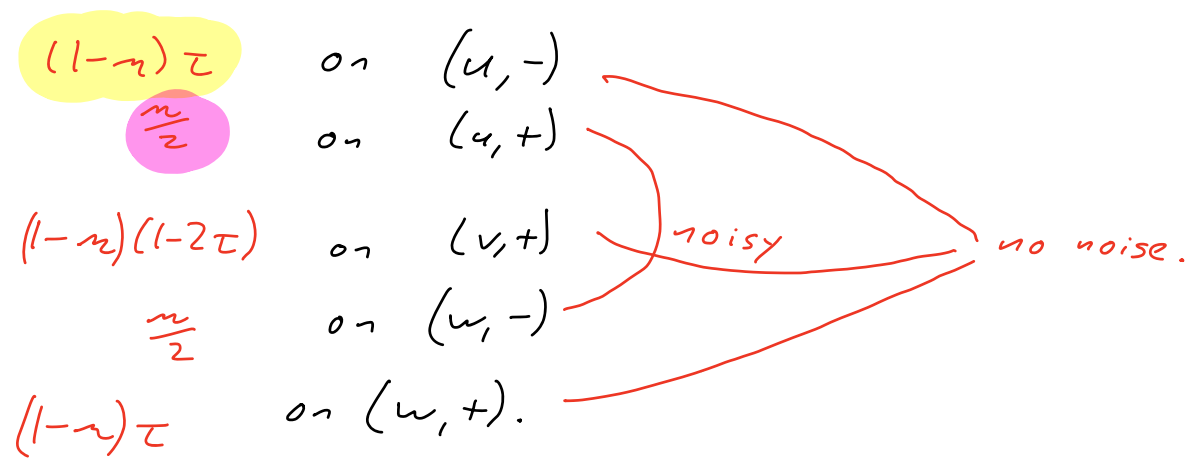
So view of learner on $\mathcal{D}, A_1, c_1$:



Adversary A_2: of the adv's n "noise budget":

$\frac{n}{2}$ on $(u,+)$, $\frac{n}{2}$ on $(w,-)$. } sum to n

So view of learner on $\mathcal{D}, A_2, c_2$:



If $\frac{\eta}{2} = (1-\eta)\tau$, then wt of $(u,-)$ same in both settings (& true for all 5 lab ex!)

$$\eta = 2\tau - 2\tau\eta$$

$$\eta(1+2\tau) = 2\tau$$

$$\eta = \frac{2\tau}{1+2\tau}.$$

So impossible to get any info about whether c_1 or c_2.

pretend world 1 or world 2 was chosen ($)

- τ on u
- $1-2\tau$ on v
- τ on w

learner • can say "$h = I$" error τ w.p. 1

- can say "$h = I$ on v, 0 on u,w" error τ w.p. 1
- can say "c_1" error 2τ w.p. $\frac{1}{2}$, 0 " $\frac{1}{2}$
- can say "c_2" error 2τ w.p. $\frac{1}{2}$, 0 " $\frac{1}{2}$

But no matter which of these learner does

$\Pr[\text{error} \geq \tau] \geq \frac{1}{2}$.

What can we do in presence of mal. noise?

Say have PAC learner A for $\mathcal{C}$, (works in noiseless setting) but there's mal. noise.

① Hope no noise in the ex. the alg. received.

② repeat multiple runs, each run (K times)

- hyp. test the K hyp's.

Sketch:

Say $m = \#$ ex. needed to $(\frac{\varepsilon}{2}, \frac{\delta}{4})$-PAC learn. ($\varepsilon, \delta$ goal.)

Run alg on m ex: $\Pr[\text{no noisy ex}] = (1-\eta)^m$

$\Pr[\text{one run has some noisy ex}] = 1-(1-\eta)^m$

$\Pr[\text{each of } K \text{ indep runs has some noisy ex}] = (1-(1-\eta)^m)^K$.

If $\eta = \frac{\ln m}{m}$, $K = m \cdot \ln(2/\delta)$:

$$\left(1-\left(1-\frac{\ln m}{m}\right)^m\right)^{m\ln(2/\delta)}$$

$$\approx \left(e^{-\frac{\ln m}{m}}\right)^m = e^{-\ln m} = \frac{1}{m}$$

$$\hookrightarrow \approx \left(1-\frac{1}{m}\right)^{m \ln 2/\delta}$$

$$\approx e^{-\ln 2/\delta} = \frac{\delta}{2}.$$

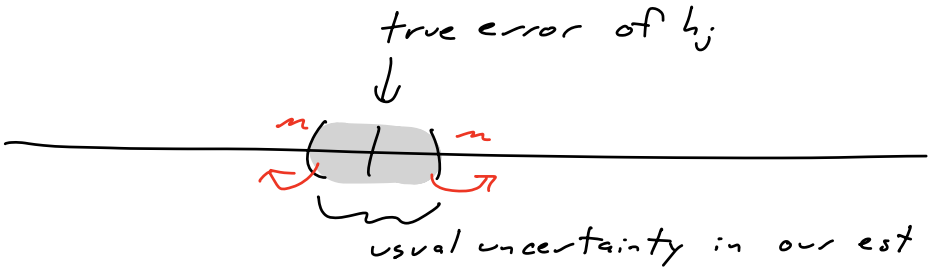
$\left(1-\frac{1}{n}\right)^n \approx \frac{1}{e}$ large n

So of hyps $h_1, \ldots, h_k$, w.p. $1-\frac{\delta}{2}$,

w.p. $1-\frac{\delta}{4}$ some h_i has error $< \varepsilon/2$.

So w.p. $\geq 1-\frac{3\delta}{4}$, some h_i has error $\leq \frac{\varepsilon}{2}$.

Last part: find a good $h \in \{h_1, \ldots, h_k\}$.



h_j from sampling even if no noise; addit. η error.

This enforces $\eta \leq \frac{\varepsilon}{8}$ as well.

Overall, can handle η up to

$$\approx \min\left(\frac{\varepsilon}{8}, \frac{\ln m}{m}\right)$$

Random Classification Noise

Recall: • $(x, c(x))$ clean from $EX(c, \mathcal{D})$

$EX^{\eta}(c, \mathcal{D})$ { • w.p. $1-\eta$, learner gets $(x, c(x))$
• w.p. η, learner gets $(x, \overline{c(x)})$.

Def: Alg A PAC learns $\mathcal{C}$ with RCN if:

- $\forall c \in \mathcal{C}$,
- $\forall$ dist $\mathcal{D}$,
- $\forall\ 0 \le \eta < \frac{1}{2}$, $\forall \varepsilon, \delta > 0$,

if A is given $\varepsilon, \delta, \eta$ & access to $EX^{\eta}(c, \mathcal{D})$,
w.p. $1-\delta$ A outputs h s.t.

$$\Pr_{x \sim \mathcal{D}}[h(x) \neq c(x)] \le \varepsilon$$

Efficient: runtime $\text{poly}\left(\frac{1}{\varepsilon}, \log\frac{1}{\delta}, \text{size}(c), n, \frac{1}{1-2\eta}\right)$

Note 1: goal is to do well on noiseless ex.

Note 2: Sps true error of h on c is τ,

i.e. $\Pr_{x \sim \mathcal{D}}[h(x) \neq c(x)] = \tau$.

View in $EX^m(c,\vartheta)$ world:

$h(x) \neq c(x)$ ← no noise; noise → $h(x) = c(x)$

$$\Pr_{(x,y)\sim EX^m(c,\vartheta)}[h(x) \neq y] = \tau(1-\eta) + \eta(1-\tau)$$

$$= \tau + \eta - 2\tau\eta = \tau + \eta(1-2\tau)$$

$$= \eta + \tau(1-2\eta)$$

$\eta < \frac{1}{2}$:

if $\tau_1 < \tau_2$

$$\eta + \tau_1(1-2\eta) < \eta + \tau_2(1-2\eta)$$

bad hyp's look worse, under $EX^m(c,\vartheta)$, than good ones

---

Note 3: Assumed learner knows η.

If don't know η: • run alg with different guesses for η $\quad \Delta, 2\Delta, 3\Delta, \ldots, \frac{1}{2}-\Delta$ $\quad h_1, h_2, \ldots, h_{\frac{1}{2\Delta}}$

• test $h_1, \ldots, h_{\frac{1}{2\Delta}}$ to find good one: ☺

Next time: • revisit old PAC learner for non conj., now in $EX^m(c,\vartheta)$.

↝ (elim.)

• modify it to only use estimates of noiseless probabilities

- show we can est. / we need

even given only $EX^n(c, D)$