Admin: can pick up midterms/sols from Rocco's office.

Last time: • finished VC01M \(\rightarrow (\text{CHF} \rightarrow \text{PAC})\) thm
  • Applic: PAC learning LTFS over \(\mathbb{R}\), efficiently!

\[ \#\text{LTFS over } 10^{15} \approx 2^{n^2} \Rightarrow n^2, \frac{1}{\varepsilon} \left(1, 1941 + 1941\right) \]
  gives \(n^2 SC\)

\[ \text{VC01M} = n+1 \rightarrow n n SC \text{ suffices.} \]

**BOOSTING:**
  • boosting confidence (easy)
  • started boosting accuracy

Today:
  "proof of concept": 3-step boosting
  (base of Schapire orig. recursive booster)

KV 4 - 4.32
  • boosting by filtering vs
  boosting by sampling; introd. AdaBoost

Spirit of boosting: exploiting the WC
  "guarantee:" ANY \(\mathcal{D}\): WC gives \(\frac{1}{2} + \gamma\) acc h.p.

Devise \(\mathcal{D}_1, \mathcal{D}_2, \ldots\) s.t. WC must provide "new info"
  to achieve \(\frac{1}{2} + \gamma\) acc.

3-stage boosting. Assume \(\gamma = \frac{1}{10}\), \(\gamma\)
  every run of WC on any dist \(\mathcal{D}\) gives \(\frac{1}{4}\)
\( \text{hyp } h' \not\models \Pr_{x \sim \mathcal{D}'} [h'(x) \neq c(x)] = \frac{4}{10} \).

1. Run \( A \) on \( \mathcal{D}' = \mathcal{D} \) to get \( h_1 \).

\[
\Pr_{x \sim \mathcal{D}'} [h_1(x) \neq c(x)] = \frac{4}{10}.
\]

2. What should \( \mathcal{D}_2 \) be, to "force \( A \) to give a hyp \( h_2 \) that has new useful info"?

The right \( \mathcal{D}_2 \) is "scales up" wt of each \( x \) s.t. \( h_1(x) \neq c(x) \) by \( \frac{5}{4} \).

- "scales down" wt \( h_1(x) = c(x) \) by \( \frac{5}{6} \).

This is \( \mathcal{D}_2 \).

\[
\Pr_{x \sim \mathcal{D}_2} [h_2(x) \neq c(x)] = \frac{1}{2}.
\]

So \( h_2 \), which sat.

\[
\Pr_{x \sim \mathcal{D}_2} [h_2(x) \neq c(x)] = \frac{4}{10},
\]

"must have new info about \( c \)."
Q: How to sim. $\text{EX}(c, \mathcal{D}_2)$ given $\text{EX}(c, \mathcal{D}_1)$?

To draw from $\text{EX}(c, \mathcal{D}_2)$:

fair $\mathcal{H}$: draw from \text{until get one where $h_i(x) = c(x)$; use that $x, c(x)$}

T: $h_i(x) \neq c(x)$

Highly efficient (by prob. 4/10).

what if $\Pr[h_i(x) = c(x)] = 99.99\%$?

0.01\% of time?

Then $h_i$ already highly acc, + c

3. Have $h_1, h_2$. What's a good $\mathcal{D}_3$?

$\mathcal{D}_3$ is $\mathcal{D}$ restr. to \text{"dis"}:

\begin{align*}
\text{x s.t. } & h_1(x) \neq h_2(x) \\
\text{Run } A \text{ on } & \text{EX}(c, \mathcal{D}_3).
\end{align*}

To sim. $\text{EX}(c, \mathcal{D}_3)$:

check if $h_i(x) \neq h_2(x)$; if so, use it; if not disc and repeat.

Could be that $\Pr[h_i(x) \neq h_2(x)]$ very small, $\Pr[x \sim \mathcal{D}]$.

Then very ineff (??). But $h_3$ "matters" only on $n$ fracs of $\mathcal{D}$.

So in very small, pick $h_3$ arbitrarily, + OK.
Final $h$ is $\text{MAJ}(h_1, h_2, h_3)$.

**Claim:** $\Pr[h(x) \neq c(x)] = 35.2\%$.

**PF:** Divide $X$ into 4 regions based on $h_1, h_2$'s agree/disagree w/ $c$:

<table>
<thead>
<tr>
<th>Region</th>
<th>Condition 1</th>
<th>Condition 2</th>
<th>$\Pr[h_1(x) \neq c(x)]$</th>
<th>$\Pr[h_2(x) \neq c(x)]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_1$</td>
<td>$h_1(x) = c(x)$, $h_2(x) = c(x)$</td>
<td>$h_1(x) = c(x)$, $h_2(x) \neq c(x)$</td>
<td>$\frac{6}{5} \cdot 0.5 \cdot \rho$</td>
<td>$\frac{6}{5} \cdot \rho$</td>
</tr>
<tr>
<td>$R_2$</td>
<td>$h_1(x) = c(x)$, $h_2(x) \neq c(x)$</td>
<td>$h_1(x) \neq c(x)$, $h_2(x) = c(x)$</td>
<td>$\frac{6}{5} \cdot \rho$</td>
<td>$\frac{6}{5} \cdot (0.4 - \rho)$</td>
</tr>
<tr>
<td>$R_3$</td>
<td>$h_1(x) \neq c(x)$, $h_2(x) \neq c(x)$</td>
<td>$h_1(x) \neq c(x)$, $h_2(x) \neq c(x)$</td>
<td>$\frac{4}{5} \cdot (0.4 - \rho)$</td>
<td>$\frac{4}{5} \cdot (0.4 - \rho)$</td>
</tr>
<tr>
<td>$R_4$</td>
<td>$h_1(x) \neq c(x)$, $h_2(x) = c(x)$</td>
<td>$h_1(x) = c(x)$, $h_2(x) = c(x)$</td>
<td>$\frac{4}{5} \cdot (0.4 - \rho)$</td>
<td>$\frac{4}{5} \cdot (0.1 + \rho)$</td>
</tr>
</tbody>
</table>

$D_2: 0.5 \cdot \rho$, $D_3: \frac{6}{5} \cdot \rho$.

Hypothesis: $h_2$ wrong on $R_2, R_3$. Let $\rho$ denote $\Pr[h(x) \neq c(x)] = 0.4$, so $R_3$.

Recall $D_2$ obtained by $\downarrow R_1, R_2$ by $\frac{5}{6}$, $\frac{5}{4}$ on $R_3, R_4$.

So invert this to recover $D_3$'s mass on each $R_i$. Get $R_1 - R_4$.

Final $h$'s error? $h$ wrong on all of $R_3$; right on all of $R_1$. $D_3$ puts all of it on $R_2, R_4$; $h_3$ wrong on 40% of $D_3$.

So final $h$'s error on 40% of $R_2 \cup R_4$.

$h$'s error under $D = D_1$:

$$\Pr[h(x) \neq c(x)] = \Pr[R_3] + \frac{4}{10} \Pr[R_2 \cup R_4]$$

$$= \frac{4}{5} \cdot (0.4 - \rho) + \frac{4}{10} \left( \frac{6}{5} \cdot \rho + \frac{4}{5} \cdot (0.1 + \rho) \right)$$
If \( WL \) has error \( \delta < \frac{1}{2} \), weak hyp \( h \) has error \( \leq 3\delta^2 - 2\delta^3 \) under \( D = D_i \).

Check: \( 3\delta^2 - 2\delta^3 < \delta \) for \( \delta < \frac{1}{2} \).

Next time: AdaBoost: single weighted Maj over \( h_1, h_2, \ldots \)