Today's Lecture

- Feasibility of learning (part 2)
  - Hoeffding inequality and multiple hypotheses
  - Noisy targets
- Generalization error (revisited)
- Toward an effective number of hypotheses
  - Dichotomies

next time

FEAS. OF LEARNING, part 2

FROM LAST TIME:

$$(1)\quad \boxed{P\left[\,\left|E_{\mathcal{D}}(h)-E_{out}(h)\right|>\epsilon\,\right]\leq 2e^{-2\epsilon^2 N} \quad \text{FOR ANY } \epsilon>0.}$$

HOEFFDING INEQ. REQUIRES:

1. SPECIFY h
2. DRAW $\mathcal{D}$
3. USE (1) TO GET $E_{OUT}(h)$ BOUNDS

ML PARADIGM $\Rightarrow$

1. COLLECT $\mathcal{D}$.
2. CONSTRUCT $\mathcal{H}$.
3. TRAIN TO FIND h_g.
4. CALCULATE $E_{in}(h_g)$ AND/OR $E_{Test}(h_g)$.
5. WANT TO FIND OR ESTIMATE $E_{out}(h_g)$.

HOW TO GET BOUND ON $E_{OUT}(h_g)$ WHEN WE NEED $\mathcal{D}$ TO CHOOSE h_g?

LET $\mathcal{H} = \{h_1, h_2, \cdots, h_M\}$.
LET h_g = FINAL CHOSEN h_m.

AML SHOWS THAT:

$$P\left[\left|E_{\mathcal{D}}(h_g) - E_{OUT}(h_g)\right| > \epsilon\right]$$

$$\leq P\left[\bigcup_{m=1}^{M}\left(\left|E_{\mathcal{D}}(h_m) - E_{OUT}(h_m)\right| > \epsilon\right)\right]$$

$$\leq \sum_{m=1}^{M} \underbrace{P\left[\left|E_{\mathcal{D}}(h_m) - E_{OUT}(h_m)\right| > \epsilon\right]}_{\leq 2e^{-2\epsilon^2 N}}$$

$$\leq 2Me^{-2\epsilon^2 N}.$$

$$(2) \therefore \boxed{P\left[\left|E_{\mathcal{D}}(h_g) - E_{OUT}(h_g)\right| > \epsilon\right] \leq 2Me^{-2\epsilon^2 N}}$$

$M = |\mathcal{H}|$ = # OF HYPOTHESES IN $\mathcal{H}$.

$\longrightarrow$ A LOOSE UPPER BOUND.

Ex: PLA: $h(\underline{x}) = \text{sgn}\{\underline{w}^T \underline{x}\}$,

$$\underline{w} \in \mathbb{R}^{D+1}$$

$$\Rightarrow M = \infty.$$

2 ASPECTS ~~of~~ OF LEARNING FEASIBILITY

FEASIBILITY OF LEARNING:
can we get $E_{out}(h_g)$ small enough?

Can we get $|E_{\mathcal{D}}(h_g) - E_{out}(h_g)|$ small enough?

$\Uparrow$

HOEFFDING INEQUALITY

Can we make $E_{\mathcal{D}}(h_g)$ small enough?

$\Uparrow$

ERROR OF h_g ON $\mathcal{D}$.

$\mathcal{D} = \mathcal{D}_{Tr}, \mathcal{D}_{val},$ or $\mathcal{D}_{Test}$.

EFFECT OF:

COMPLEXITY OF $\mathcal{H}$	$\lvert\cdot\rvert$ BOUND $\uparrow$,	$E_{\mathcal{D}} \downarrow$.
NUMBER OF DATA POINTS, N.	$\lvert\cdot\rvert$ BOUND $\downarrow$,	$E_{\mathcal{D}} \uparrow$.

Noisy Target Functions

INSTEAD OF: $y = f(\underline{x})$, AND DATA POINTS

y ↑ o/p class; f ← target function.

ARE FROM $p(\underline{x})$.

WE HAVE: $p(y|\underline{x})$, DATA POINTS COME FROM:

$$p(\underline{x}, y) = p(y|\underline{x})\, p(\underline{x})$$

e.g.: $y = f(\underline{x}) + \text{noise}$ (REGR)

OR $y = \text{sgn}\{g(\underline{x}) + \text{noise}\}$ (CLASS'N).

y_i's AR DRAWN FROM $p(y|\underline{x})$, WITH

$\underline{x}_i$ DRAWN FROM $p(\underline{x})$

USE $p(y|\underline{x})$ AS THE MORE GENERAL MODEL.

e.g., $y = f(\underline{x})$ BY LETTING:

$$p(y|\underline{x}) = \begin{cases} 1, & y = f(\underline{x}) \\ 0, & \text{otherwise} \end{cases}, \text{ OR } p(y|\underline{x}) = \delta[y - f(\underline{x})].$$

-- [FIG. 1.11 IN AML] --

GENERALIZATION ERROR [AML 2.1]

HOEFFDING INEQ:

$$P\left[\,|E_{\mathcal{D}}(h_{\text{[redacted]}}) - E_{out}(h_{\text{[redacted]}})| > \epsilon\,\right] \le \overbrace{2Me^{-2\epsilon^2 N}}^{\delta} \quad \forall \epsilon > 0.$$

RE-ARRANGE:

$$P\left[\,\underbrace{|E_{out}(h_{\text{[redacted]}}) - E_{\mathcal{D}}(h_{\text{[redacted]}})|}_{|B|} \le \epsilon\,\right] > 1-\delta$$

$$\delta \triangleq 2Me^{-2\epsilon^2 N}$$

$$\Rightarrow \quad \epsilon = \sqrt{\frac{1}{2N}\ln\frac{2M}{\delta}}$$

$B > 0 : \Rightarrow E_{out} \le E_{\mathcal{D}} + \epsilon$

$B < 0 : \Rightarrow E_{out} \ge E_{\mathcal{D}} - \epsilon$

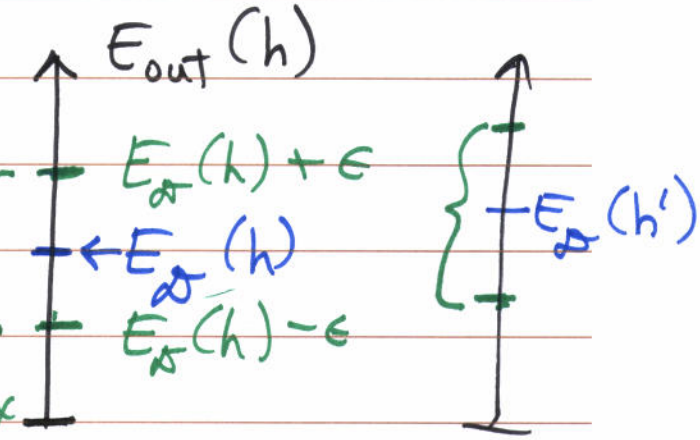
(2.1) $\therefore$ $$\boxed{P\left[E_{out}(h_g) \le E_{\mathcal{D}}(h_g) + \sqrt{\frac{1}{2N}\ln\frac{2M}{\delta}}\right] > 1-\delta}$$

(2.1') $$P\left[E_{out}(h_{\text{[redacted]}}) \ge E_{\mathcal{D}}(h_{\text{[redacted]}}) - \sqrt{\frac{1}{2N}\ln\frac{2M}{\delta}}\right] > 1-\delta$$

δ IS OUR TOLERANCE.

(2.1) GIVES UPPER BOUND ON $E_{out}(h_g)$ IN TERMS OF OUR TOLERANCE δ AND N, M.

(2.1′) GIVES LOWER BOUND ON $E_{out}(h)$, FOR ANY $h \in \mathcal{H}$.



$E_{out}(h)$ is in this range with probability $p > 1 - \delta$.