Summary of Prev. Lecture:

What is Learning? $f: X \to Y$

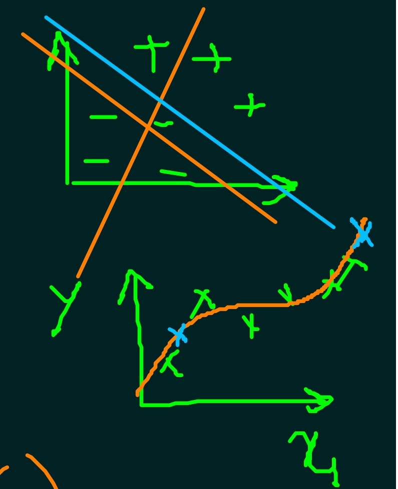
When can we learn? $g \approx f$

- ↳ Data
- ↳ Pattern
- ↳ Not Math.

Example Learning

- ↳ Classification
- ↳ Regression
- ↳ Association

$P(Y \mid X)$



Unknown Target $f: X \to Y$

↓

Training Ex.

$\langle x_1, y_1 \rangle \; \cdots \; \langle x_m, y_m \rangle$

→ DT NN SVM

Learn Algo

↓

h_i

$g \approx f$

hyp.

$w_1 - \checkmark$

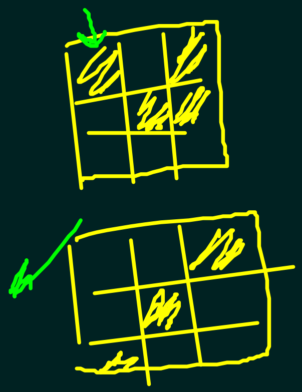
$w_2 \to \checkmark$

Hypothesis Set (H)

$h_1 h_2 \ldots\ldots$

Domain

Train Ex

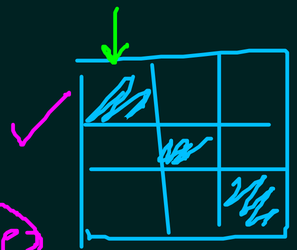


$f = +1$

$f = -1$

$f = +1$

Test



$f = ?$

$f = -1$

YES

$\Rightarrow$ Is learning feasible?

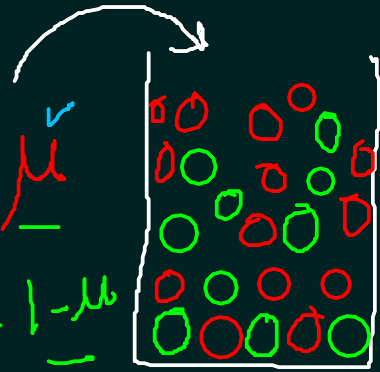
~~Can we learn?~~ (Probability)

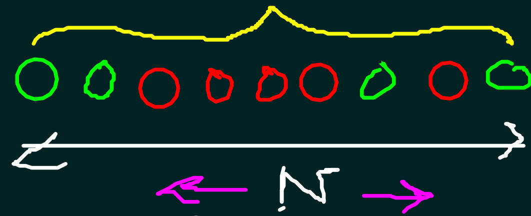
▷ What $\boxed{\nu}$ value says abt $\boxed{\mu}$?

↳ No! Possibility ✓ vs

↳ Yes! Probable!

unknow $\begin{cases} Prob(Red) = \mu \\ Prob(green) = 1-\mu \end{cases}$





$Prob^{S}(Red) = \nu$ ✓

$Prob[\text{Bad Approx}] \le$ xxxxxx less

$Prob[\underbrace{|\nu - \mu| > \epsilon}] \le$ $\boxed{\text{less}}$! $\frac{1}{100}$

$\boxed{\text{suff high } N \quad \boxed{\text{error}}\ \epsilon}$ ✓

Hoeffding InEq } ✓

$\Rightarrow \boxed{Prob[|\nu-\mu| > \epsilon] \le 2e^{-2\epsilon^2 N}}$

$\epsilon = 0.01$

H.I.E: $\text{Prob}[\,|\nu - \mu| > \epsilon\,] \le 2e^{-2\epsilon^2 N}$

$\mu \Rightarrow x$, α

$\nu \pm \epsilon$ ↳ error ↳ Sample Size

$\boxed{\text{P.A.C}}$ ← $(1-\alpha)$, $\pm\epsilon$, $\boxed{M}$

▷ What is its connection to learning?

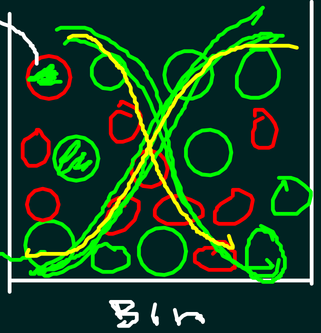
Input(x) → G, → R

$x \in X$

↪ $P[\,|E_{in}(h) - E_{out}(h)| > \epsilon\,] \le 2e^{-2\epsilon^2 N}$

$h(x) \neq f(x)$

$h(x) = f(x)$



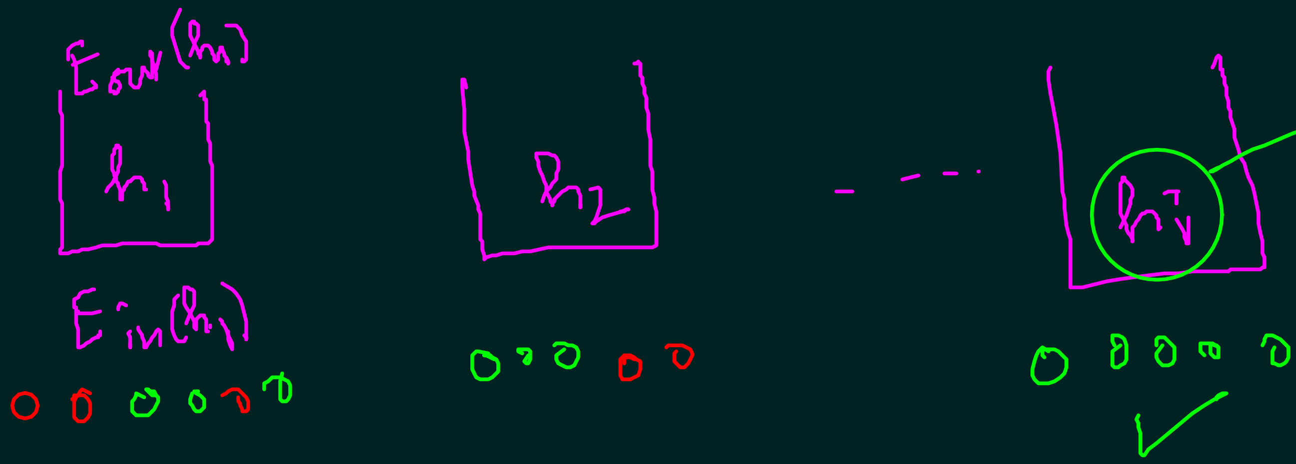
$\boxed{E_{in}(h) = \nu}$

$\boxed{E_{out}(h) = \mu}$

$\longleftarrow N \longrightarrow$

In Sample (See)

Not Really Learning, But verify "h"

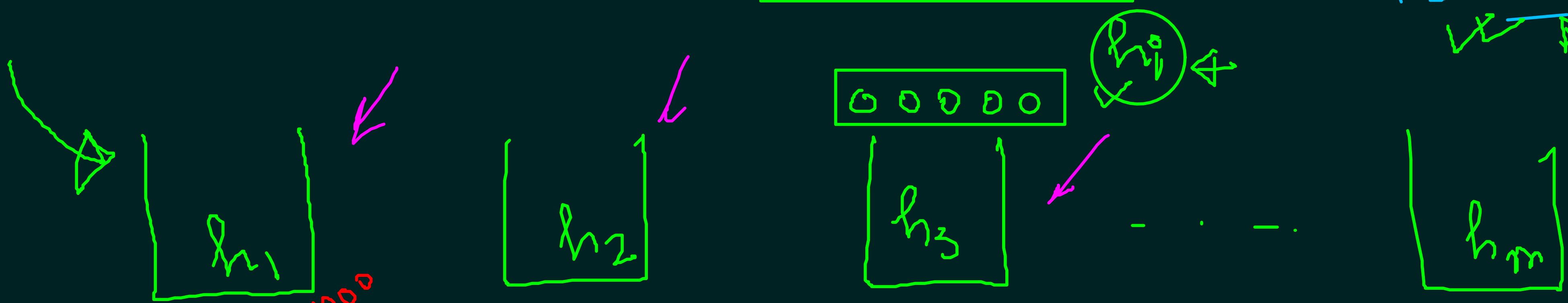


$h_i \approx f$??

$$\text{Prob}\left[\text{a fair coin give all heads in 10 flips}\right] = \frac{1}{2^{10}} \approx 0.1\%$$

$$P(H\,H\,H\,----\,H)$$

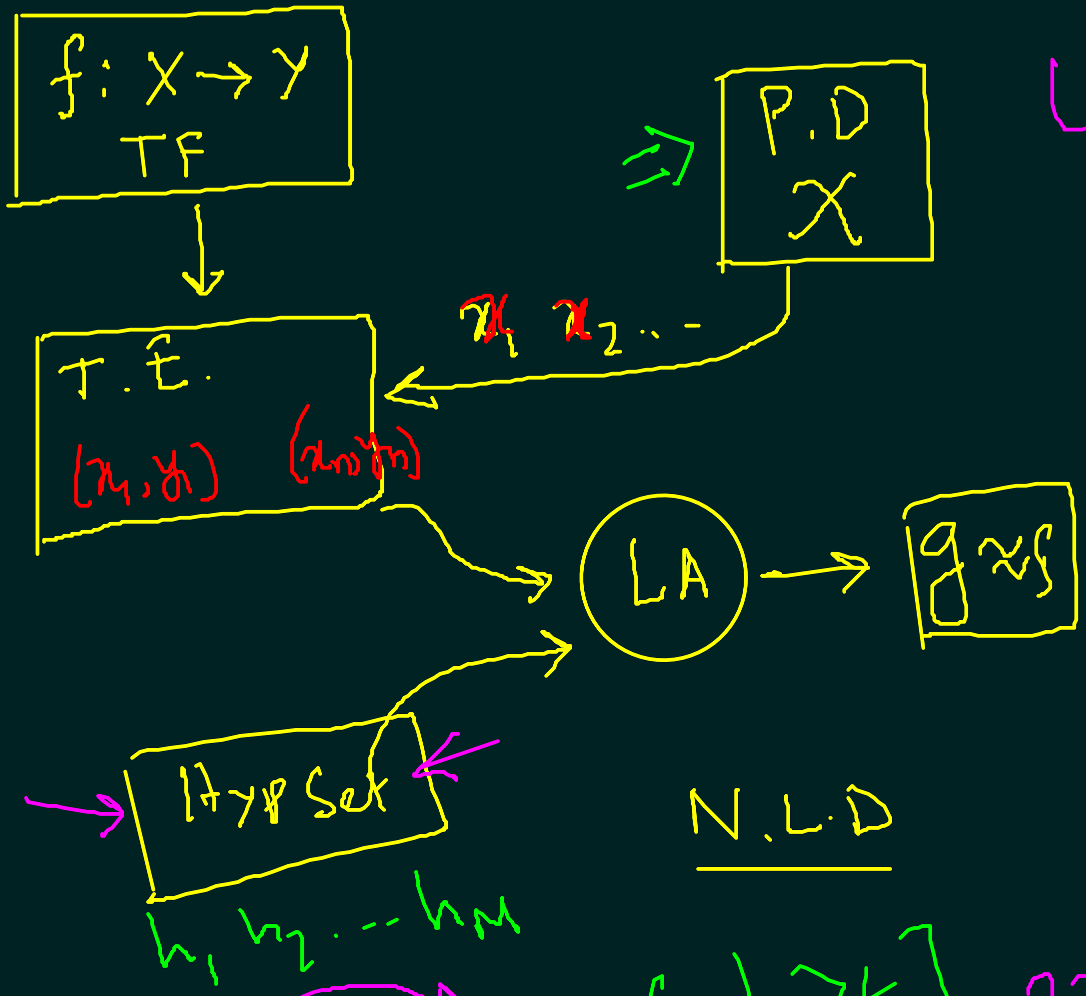
$$\text{Prob}\left[\text{1000 fair coins all heads 10 } \overset{\text{sep}}{\text{flip}}/\text{coin at least once}\right] = 1 - \left(1 - \frac{1}{2^{10}}\right)^{1000} \approx 63\%$$



$1000 \to 10 \to \left(1 - \frac{1}{2^{10}}\right)^{1000}$

$$P\left[\,|E_{in}(h) - E_{out}(h)| > \epsilon\,\right] \leq 2e^{-2\epsilon^2 N}$$

¿dilute?



$\hookrightarrow \text{Prob}\left[\, |E_{in}(g) - E_{out}(g=f)| > \epsilon \,\right] \leq \, ?$

h_1 or h_2 or h_3

$$P\left[\, |E_{in}(g) - E_{out}(g)| > \epsilon \,\right]$$

$$\leq P\left[\, |E_{in}(h_1) - E_{out}(h_1)| > \epsilon \text{ or } |E_{in}(h_2) - E_{out}(h_2)| > \epsilon \text{ or } \dots \,\right]$$

$$\leq \sum_{i=1}^{M} 2e^{-2\epsilon^2 N} = 2\boxed{M}e^{-2\epsilon^2 N}$$

[Generalization]

$$P\left[\, |E_{in}(h) - E_{out}(h)| > \epsilon \,\right] \leq \boxed{2Me^{-2\epsilon^2 N}} \; ??$$

$1+\epsilon$

PAC

~~feasible~~

$e = 2.81$

$$\text{Prob}\left[\, |E_{in}(g) - E_{out}(\overset{f}{g})| > \epsilon \,\right] \qquad h_i = \underline{\text{worst}}$$

$h_1 \quad h_2 \cdots h_M$

$$= \text{Prob}\Big[\, |E_{in}(h_1) - E_{out}(h_1)| > \epsilon \quad \boxed{\text{OR}}$$

$$|E_{in}(h_2) - E_{out}(h_2)| > \epsilon \quad \text{OR}$$

$$\vdots$$

$$|E_{in}(h_M) - E_{out}(h_M)| > \epsilon \,\Big]$$

$$\sum_{i=1}^{M} x = Mx$$

$$\le \sum_{i=1}^{M} 2e^{-2\epsilon^2 N} = 2\boxed{M}e^{-2\epsilon^2 N}$$

(upperbound)

$\boxed{\text{PAC}}$ ✗

$\boxed{\text{not}}$

Generalized

$$P(A \vee B) = P(A) + P(B) - \cancel{P(A \wedge B)}$$