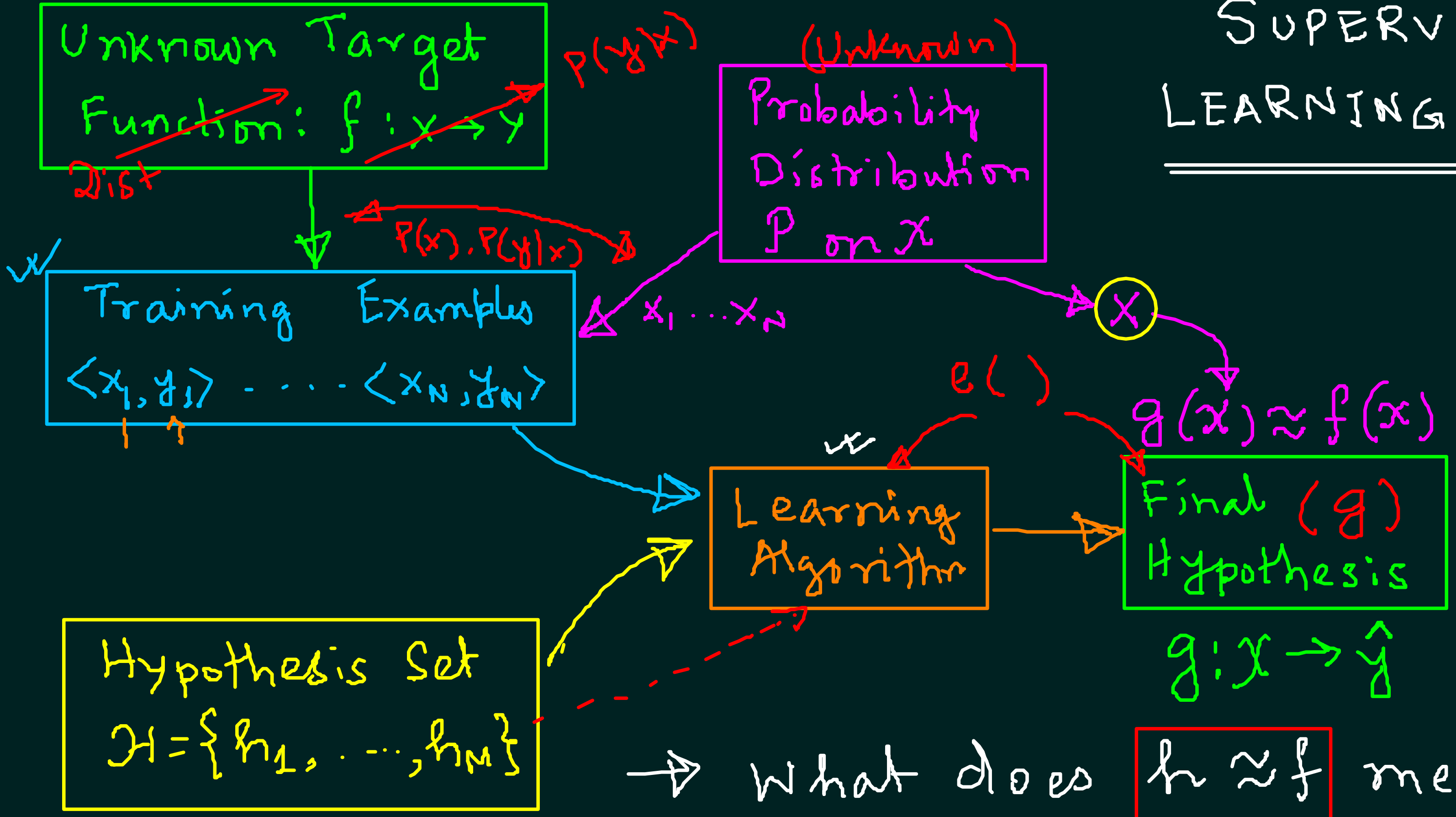
SUPERVISED LEARNING DIAGRAM



- In-Sample $\equiv$ Train
- Out-of-Sample $\equiv$ Test

$\rightarrow$ What does $\boxed{h \approx f}$ mean?

- Out-of-Sample

$$E_{out}(h) = \mathbb{E}_{x}\left[e\left(h(x), f(x)\right)\right]$$

$$E(h, f) = \text{error} = e\left(h(x), \underset{y}{\cancel{f(x)}}\right) \rightarrow \text{pointwise}$$

$\rightarrow$ $\boxed{\text{Squared}}$

$$\left(h(x) - \underset{}{\cancel{f(x)}}\ y\right)^2$$

$\rightarrow$ $\boxed{\text{Binary}}$ $\frac{+1}{-1}$

$$[\![\, h(x) \neq f(x) \,]\!] \quad \begin{matrix} 1 \\ 0 \end{matrix}$$

- In-Sample Error

$$E_{in}(h) = \frac{1}{N} \sum_{n=1}^{N} e\left(h(x_n), \underset{y_n}{\cancel{f(x_n)}}\right)$$

Ex: fingerprint recog.

 $\rightarrow \boxed{f} \rightarrow \begin{cases} +1 \\ -1 \end{cases}$

g

h \ f	+1	−1
+1	—	false accept
−1	false reject	—

$\boxed{\text{error}}$

h \ f	+1	−1
+1	0	1000 5
−1	1 10	0

wisely choose the penalty

- Intruder is entering IB office
- Supermarket Discount

$\boxed{\text{Choice of error function}}$

Choice of error function → plausible Measure, friendly error

plausible Measure → Squared error → one step learning

plausible Measure → Gaussian Error → (SGD)

friendly error → Square, Bell Curve

$f : X \to y$ → Noise

↳ unknown function

NO!

Credit Approval

Age, Sal, # Res. → $x_i \to$ Yes

$x_j \to$ No

} Train

< 1000 → NO

> 1000 → Yes

998 → Yes

J_1: 1000 → Yes

J_2: 1000 → NO

Prob(y | X) ← target distribution

unknown

w

Prob dist of X

$x_1 \ldots x_N$

TE:

$\langle x_1, y_1 \rangle, \ldots \ldots \langle x_N, y_N \rangle$

by Joint distribution

$P(x) \cdot P(y|X) = P(x,y)$

Noisy Target

deterministic target + Noise

$P(y|x)$

Special case of noisy target

$y = f(x) \dashrightarrow P(y|x) = 0$ except $y = f(x)$

$\boxed{f(X) = E(y|X)} + (y - f(x))$

$\boxed{E_{out}(g) \approx E_{in}(g)}$ ← Inductive Learning Principle

Lecture 2

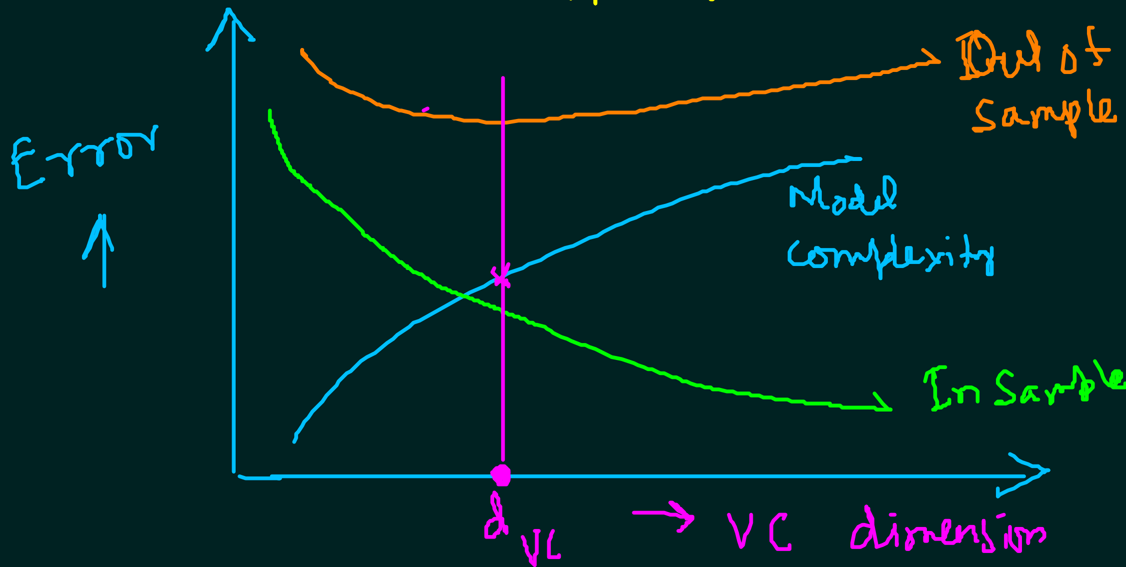
Learning is Feasible $\} \Rightarrow$ $\text{Prob}\left[\,|E_{out} - E_{in}| > \epsilon\,\right] \leq \boxed{2Me^{-\epsilon^2 N}}$

∞

$\boxed{E_{out}(g) \approx 0}$ — final aim ---→ all other Lecture

→ $E_{in}(g) \approx E_{out}(g)$ — is close enough????

→ $E_{in}(g) \approx 0$ ✓ covered



$\boxed{x_1 + 2x_2}$

$\left[x_1 + 2x_2 + 3x_1x_2 + \frac{1}{2}x_2^2\right]$

$\boxed{g}$ ✓

Testing: $P\left[\,|E_{out} - E_{in}| > \epsilon\,\right] \le 2\, e^{-2\epsilon^2 N}$

how you learn

how you gave final exam

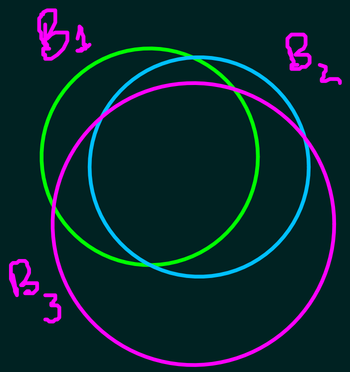
Training: $P\left[\,|E_{in} - E_{out}| > \epsilon\,\right] \le \boxed{2 M e^{-2\epsilon^2 N}}$

solve Mock Test

final exam

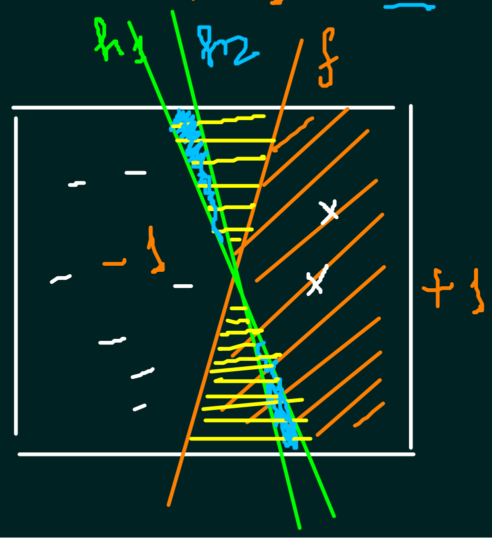
theory of generalization

$M = \boxed{(\infty)}$ → $\boxed{f(N)}$



$P(B_1 \text{ or } B_2 \text{ or } B_3)$

$\le P(B_1) + P(B_2) + \cdots$



$|E_{in}(h_1) - E_{out}(h_1)|$

$\approx |E_{in}(h_2) - E_{out}(h_2)|$

hope!