Lec 30 Why is learning Possible?

(E_{in} VS E_{out}) ch 1.3

1. Why is learning Possible?

↳ <u>e.g. Binary classification</u>

$\underline{x} \in \{0, 1\}^2$, $y = \{+1, -1\}$

$\underline{x}$	$y = f(x)$
(0,0)	+1
(0,1)	-1
(1,0)	-1
(1,1)	?

} given

← Can we learn? NO. Can be anything

Impossible to learn ⇒ Need randomness

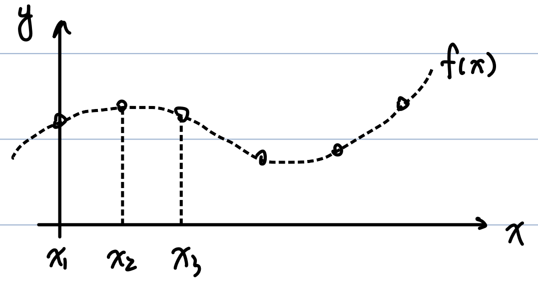
↳ e.g. A run with balls, each either red or green.

Want to estimate the % that are red.

· Randomly pick 50 balls and observe colors : G = 10, R = 40

∴ We expect the % red $\approx \frac{40}{50} = 0.8$

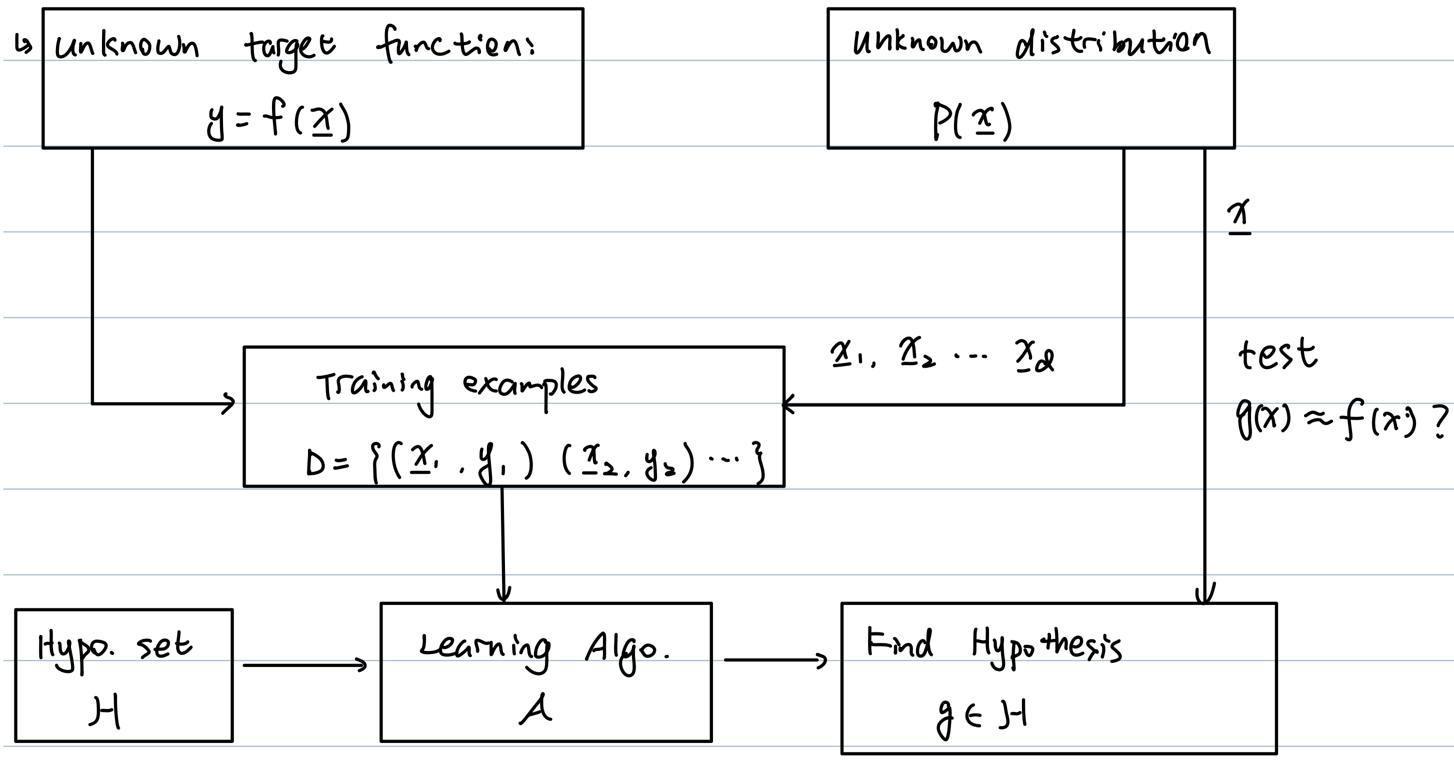
↳ e.g. Nyquist sampling theorem for interpolation:



· How many plot do we have ?

· How complex is $f(\cdot)$?

2. Probability Approximatly correct (PAC) learing



↳ e.g. Binary classification:

(1) Unknow function $f : \mathbb{R}^d \rightarrow \{+1, -1\}$

d: input dimension

(2) Unknown distribution: $P(\underline{x})$

(3) Training example:

$\underline{x}_1, \underline{x}_2 \cdots \underline{x}_N \overset{iid}{\sim} P(\underline{x})$ // iid: identical independent distribution

$\therefore P(\underline{x}_1, \underline{x}_2, \cdots \underline{x}_N) = P(\underline{x}_1)\ P(\underline{x}_2) \cdots P(\underline{x}_N)$

(4) Hypothesis set $\mathcal{H}$:

$\mathcal{H} = \{h_1, h_2, \cdots h_M\}$

$h_i : \mathbb{R}^d \rightarrow \{+1, -1\}$ (candidate for estimation of f)

e.g: $M=2$, $\mathcal{H} = \{h_1, h_2\}$

$h_1(\underline{x}) = +1, \ \forall \underline{x}$

$h_2(\underline{x}) = -1, \ \forall \underline{x}$

learning Algo $\mathcal{A}$: majority rule e.g: Binary learning classification:

$$\hat{y} = \text{sign}(\underline{w}^T \underline{x} + b)$$

then $\mathcal{H}$ = all possible $(\underline{w}, b) \in \mathbb{R}^{d+1}$

$M = +\infty$

$A = PLA$

3. Performace Matrix (for given hypothesis $g \in \mathcal{H}$)

① In sample (training) error,

$$E_{in}(g) = \frac{1}{N} \sum_{n=1}^{N} e(y_n, g(\underline{x}_n)) \qquad // \text{"}e_n\text{"}$$

$$e(\underbrace{y_n}_{\text{true label}}, \underbrace{g(\underline{x}_n)}_{\text{estimated}}) = \mathbb{1}(y_n \neq g(\underline{x}_n)) = \begin{cases} 1, & y_n \neq g(\underline{x}_n) \\ 0, & y_n = g(\underline{x}_n) \end{cases}$$

② Out-of-sample (test) error

$$E_{out}(g) = \underbrace{\mathbb{E}}_{\text{expectation over } p(\underline{x})}[e(y, g(\underline{x}))]$$

$\underline{x} \sim P(x)$, iid

$$= \mathbb{E}[\mathbb{1}(y_n \neq g(x_n))]$$

$$= 1 \cdot P_r\{y_n \neq g(x_n)\} + 0 \cdot P_r\{y_n = g(x_n)\}$$

$$= P_r\{y_n \neq g(x_n)\}$$

$E_{in}(g)$	$E_{out}(g)$
computable	unknown
random	deterministic

$P_r\{E_{out}(g)\} \approx E_{in}(g) \approx 1$ P.A.C