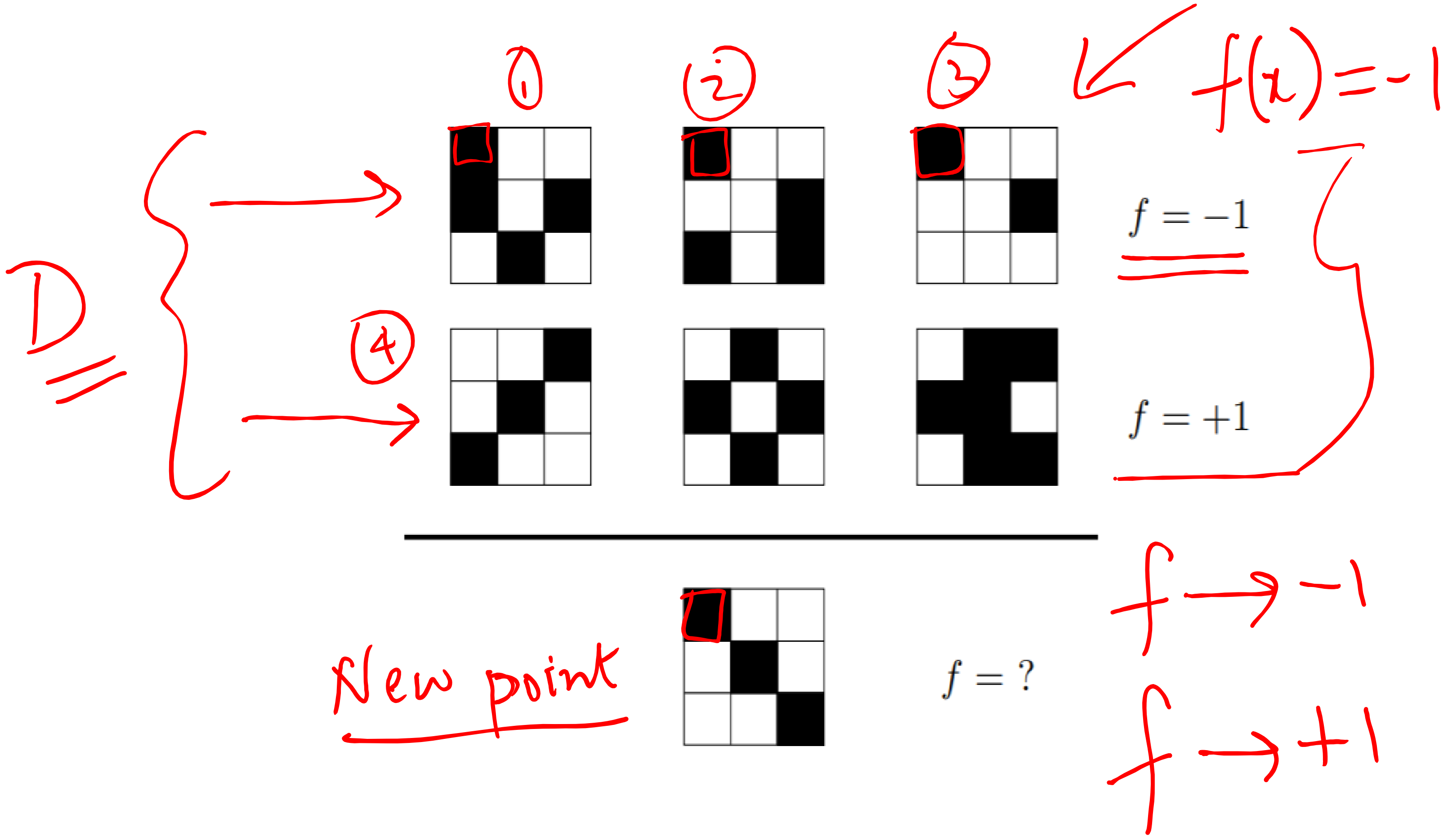
Machine Learning from Data

Lecture 3: Spring 2021

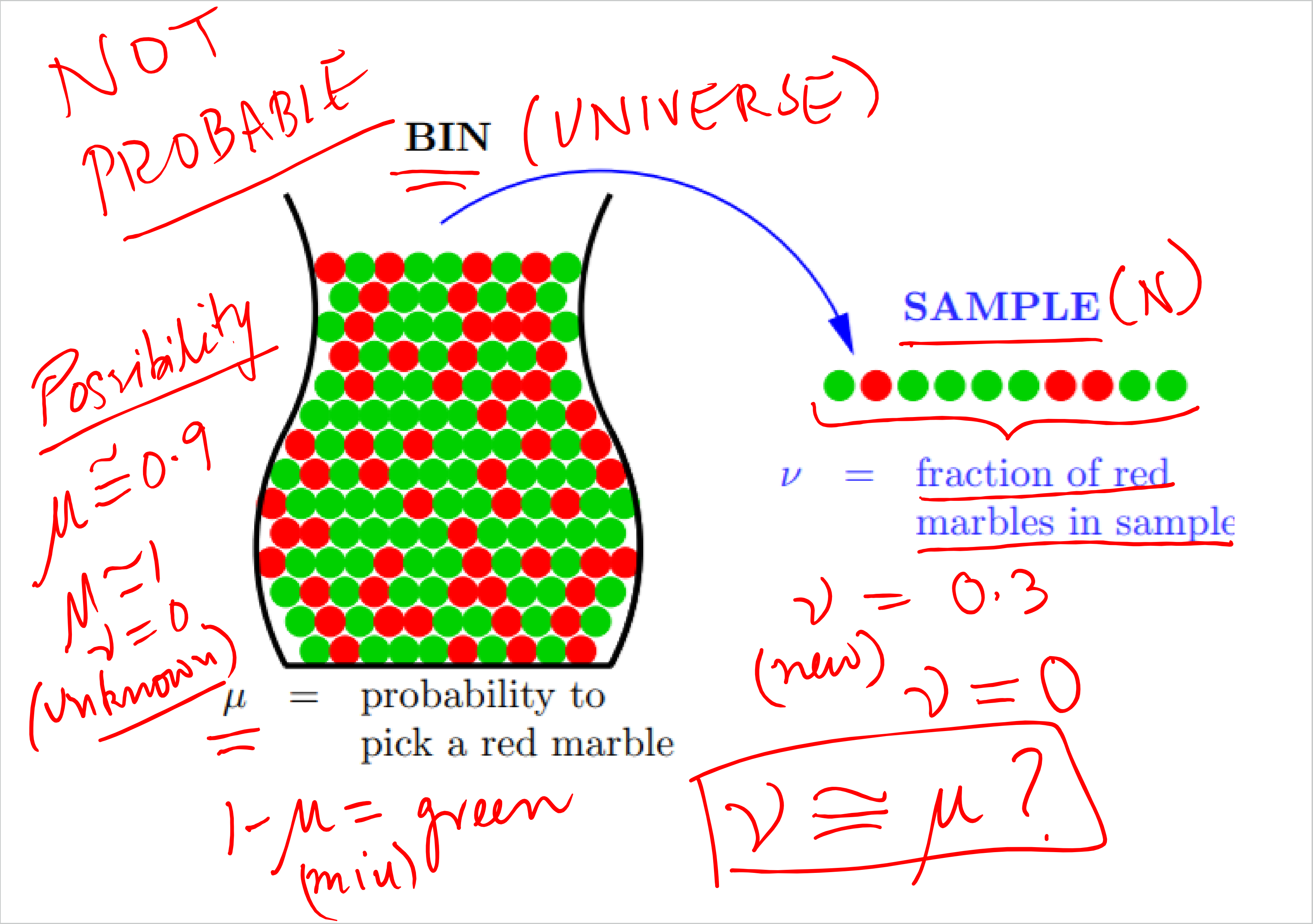
Today’s Lecture

- Is Learning Feasible?
- Infer outside the $\boxed{\text{dataset D}}$
- Hoeffding’s Inequality

Outside Data



Population Mean from the Sample Mean



Probability to the Rescue

- Draw 1000 (N) marbles repeatedly. ($\mu = 0.4$)

  Calculating ν (*)

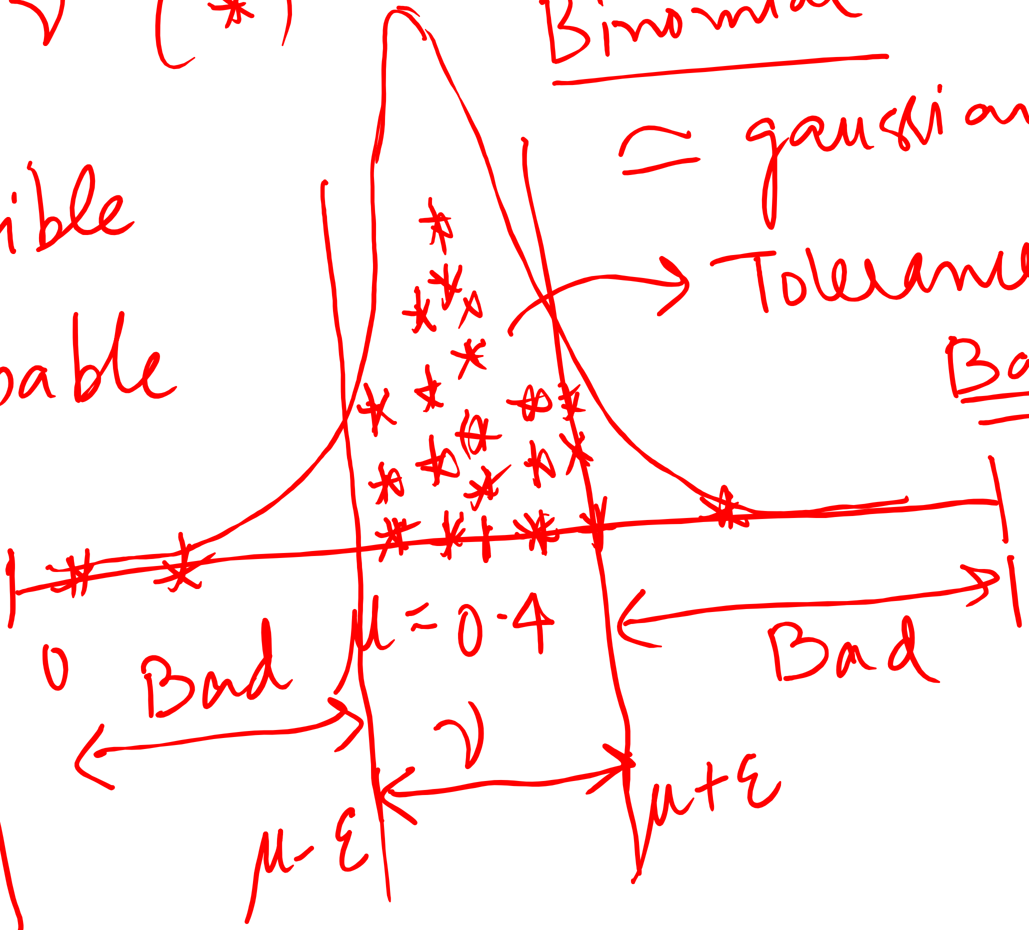
Binomial

$\simeq$ gaussian

$\boxed{\nu \simeq \mu}$

→ Bad is possible

Good is probable



→ Tolerance ϵ

Bad $\boxed{|\nu - \mu| > \epsilon}$

Good $\boxed{|\nu - \mu| \le \epsilon}$

→ As N grows

$\boxed{\nu \to \mu}$

Hoeffding's Inequality

- $\boxed{\nu \rightarrow \mu}$

$P[\text{Bad}] \longrightarrow$ small

$P[|\nu - \mu| > \epsilon]$ tolerance

$\text{Prob}[\text{Bad}] \leq$ Something small

$N \longrightarrow$ bigger $\qquad \nu$ closer to μ

Hoeffding's Inequality continued…

- $$P[|\nu - \mu| > \epsilon] \leq \underbrace{2e^{-2\epsilon^2 N}}$$ ✓

$\underbrace{\nu \quad \mu}_{\nu \to \mu}$

$\nu \to \underbrace{\mu}_{\text{unknown}}$

$N \to \infty \qquad \mu \simeq \nu \qquad$ high probability.

$\epsilon \to$ small

$$\to P(\text{good}) = \underbrace{\text{Prob}[|\nu - \mu| \leq \epsilon]} > \underbrace{1 - 2e^{-2\epsilon^2 N}}$$

Example

- $N = 1000$, observe ν
  - $\epsilon = 0.05 \rightarrow 1 - 2e^{-2(0.05)^2 1000} \simeq 99\%$ ✓

  99% of the time $\nu - 0.05 \le \mu \le \nu + 0.05$

  - $\epsilon = 0.1 \rightarrow 1 - 2e^{-2(0.1)^2 1000} \simeq 99.999\%$ ✓

  $\nu - 0.1 \le \mu \le \nu + 0.1$

Relationship with Learning

1) Samples must be random (IID)
2) High probability $\nu \to \mu$ (PAC)
3) N is large $\{ \epsilon^2 N$ must be large$\}$
4) $\nu \xrightarrow{\text{reaches out}} \mu$ for that particular bin.

ν = samples

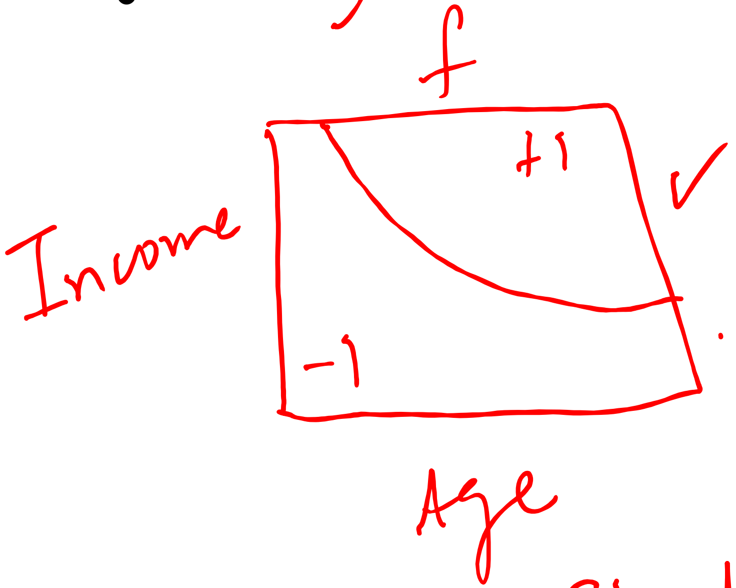
$\boxed{\nu \simeq \mu}$ ~~Hen~~ Is possible.

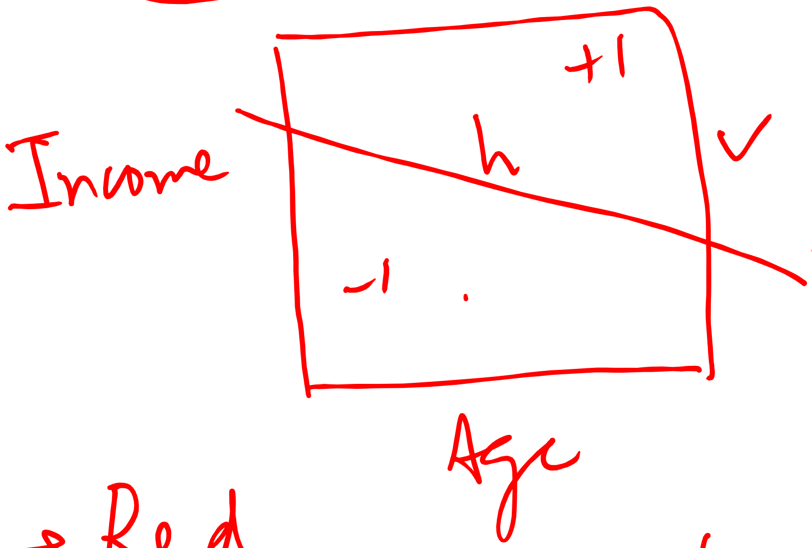
Relationship with Learning

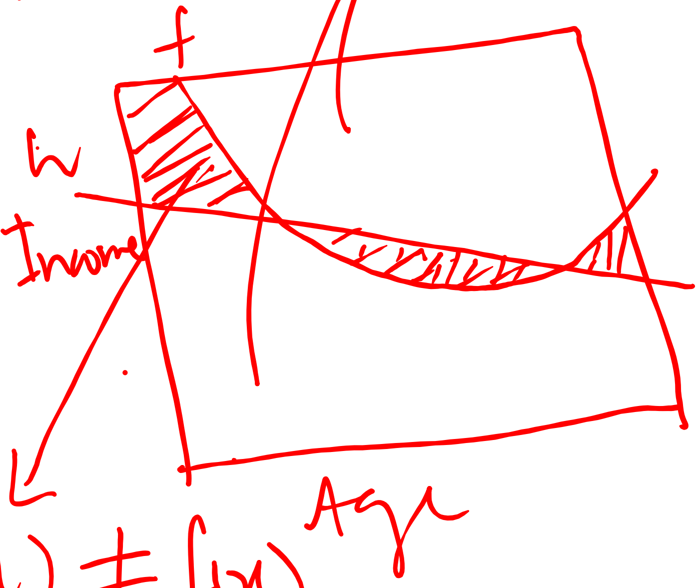
- $\mathcal{H}$

$f : \mathcal{X} \longrightarrow \mathcal{Y}$

$h \in \mathcal{H}$ $\qquad$ $h(x) = f(x)$ $\qquad$ $\mathcal{D}$







$h(x) \neq f(x) \leftarrow$ Shaded $\rightarrow$ Red

$h(x) = f(x) \leftarrow$ Unshaded $\rightarrow$ Green

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

$\text{Prob}\left[h(x) \neq f(x)\right] \longleftrightarrow \mu$

P , $\underline{\underline{f(x_n)}} = y_n$ (known) Bin $\rightarrow$ Learning problem.

Picked a sample

$$E_{in}(h) = \frac{1}{N} \sum_{n=1}^{N} [\![h(\underline{x}) \neq \underline{y_n}]\!]$$

(ν)

$$(\mu) \rightarrow E_{out}(h) = P[\, h(x) \neq f(x) \,]$$

Hoeffding's Summarized (Learning)

$$\mathbb{P}\left[|E_{\text{in}}(h) - E_{\text{out}}(h)| > \epsilon\right] \leq \underline{2e^{-2\epsilon^2 N}}, \qquad \text{for any } \underline{\epsilon > 0}.$$

$$\mathbb{P}\left[|E_{\text{in}}(h) - E_{\text{out}}(h)| \leq \epsilon\right] \geq 1 - 2e^{-2\epsilon^2 N}, \qquad \text{for any } \epsilon > 0.$$

ν μ tolerance

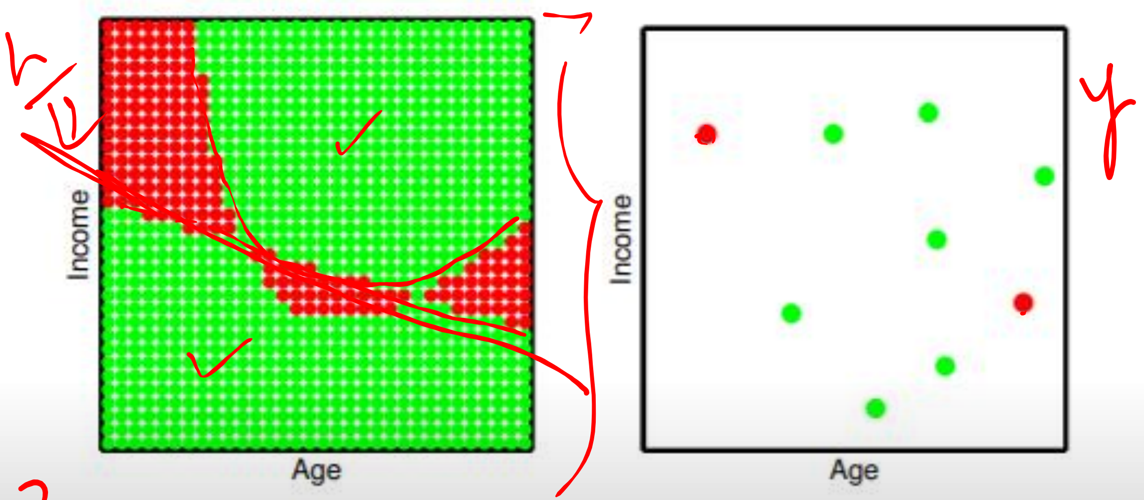
$E_{in}(h)$

fixed but unknown

If $E_{in}(h) \simeq 0 \Rightarrow E_{out}(h) \simeq 0$ with high probability

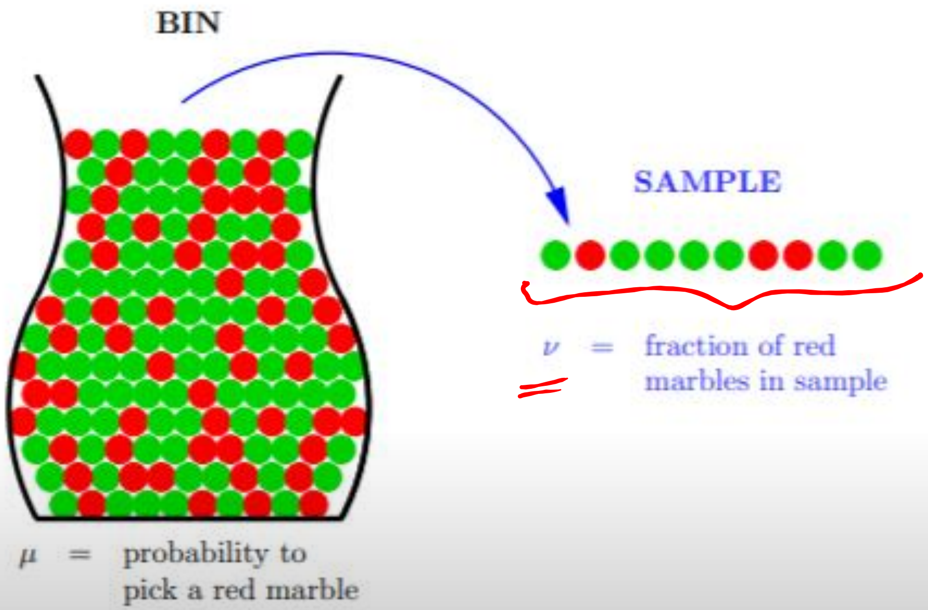
If $\underbrace{E_{in}(h) >>> 0}$ $f \neq h$

Function Vs Bin



Bin

Unknown f and $P(\mathbf{x})$, **fixed** h



Learning	**Bin Model**
input space $\mathcal{X}$	Bin
$\mathbf{x}$ for which $h(\mathbf{x}) = f(\mathbf{x})$	green marble
$\mathbf{x}$ for which $h(\mathbf{x}) \neq f(\mathbf{x})$	red marble
$P(\mathbf{x})$	randomly picking a marble
data set $\mathcal{D}$	sample of N marbles
Out-of-sample Error: $E_{\text{out}}(h) = \mathbb{P}_{\mathbf{x}}[h(\mathbf{x}) \neq f(\mathbf{x})]$	μ = probability of picking a red marble
In-sample Error: $E_{\text{in}}(h) = \frac{1}{N}\sum_{n=1}^{N} [\![h(\mathbf{x}) \neq f(\mathbf{x})]\!]$	ν = fraction of red marbles in the sample

unknown

Verification Vs. Real Learning

Verification

- fixed h (hypothesis)

h is certify

h does not depend on the data

$\boxed{E_{in}(h) \simeq E_{out}(h)}$

Can apply HB

Learning

fix a hypothesis set $\mathcal{H}$

Pick $g \in \mathcal{H}$ (using the data)

certify g (looks best on D)

$g \rightarrow$ picked from data.

$E_{in} \rightarrow$ small, pick

$E_{in}(g) \simeq E_{out}(g)$

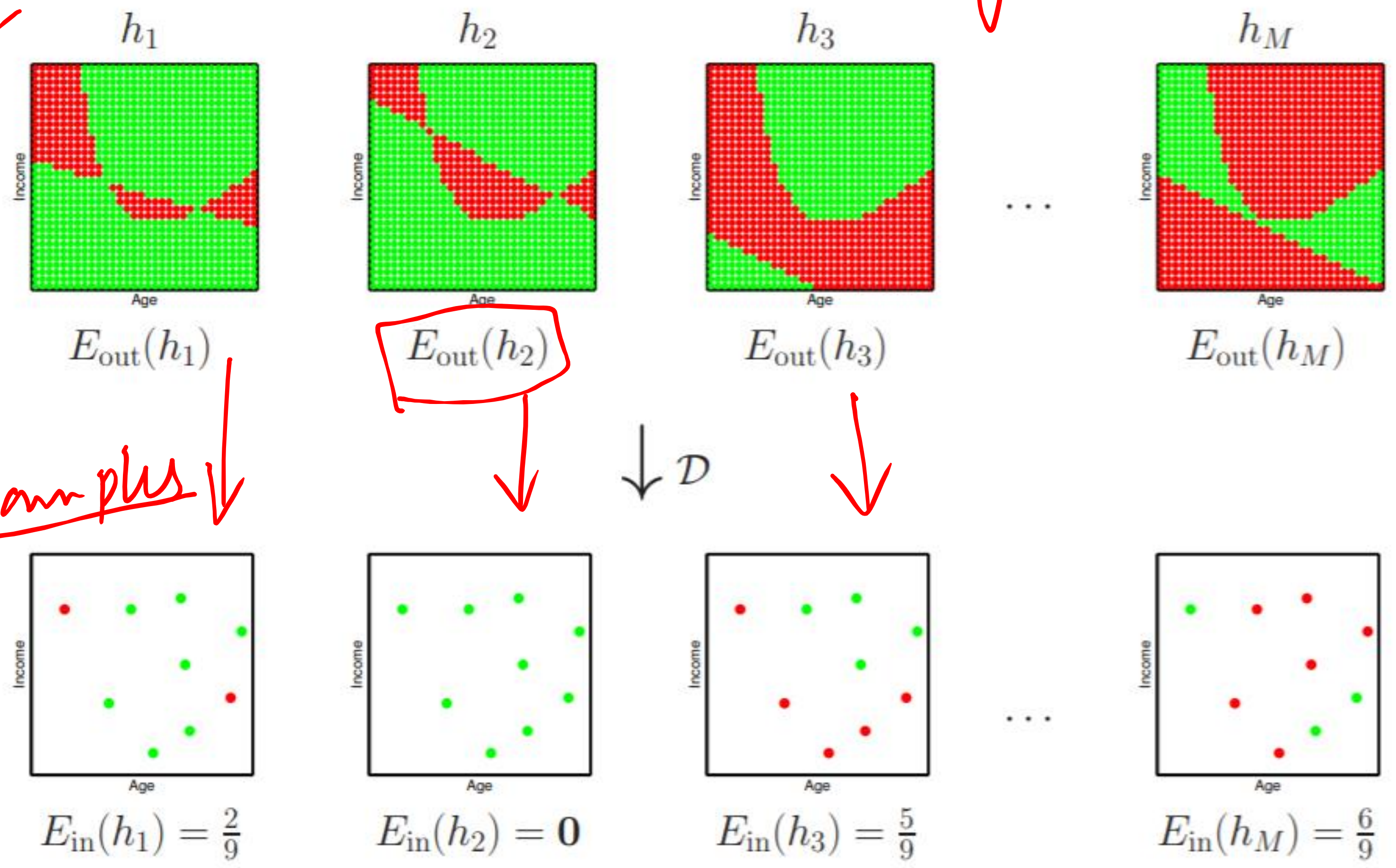
HB ✗

Bins

$g = h_2$

$g \simeq f$

$\underline{E_{in}} \rightarrow$ small



Pick the hypothesis with minimum E_{in}; will E_{out} be small?

Finite Learning Model

Selection Bias

Notion of choice when selecting g gave us the ability to

$$E_{in}(g) \simeq 0$$

Cannot say $E_{in}(g) \simeq E_{out}(g)$ ✗