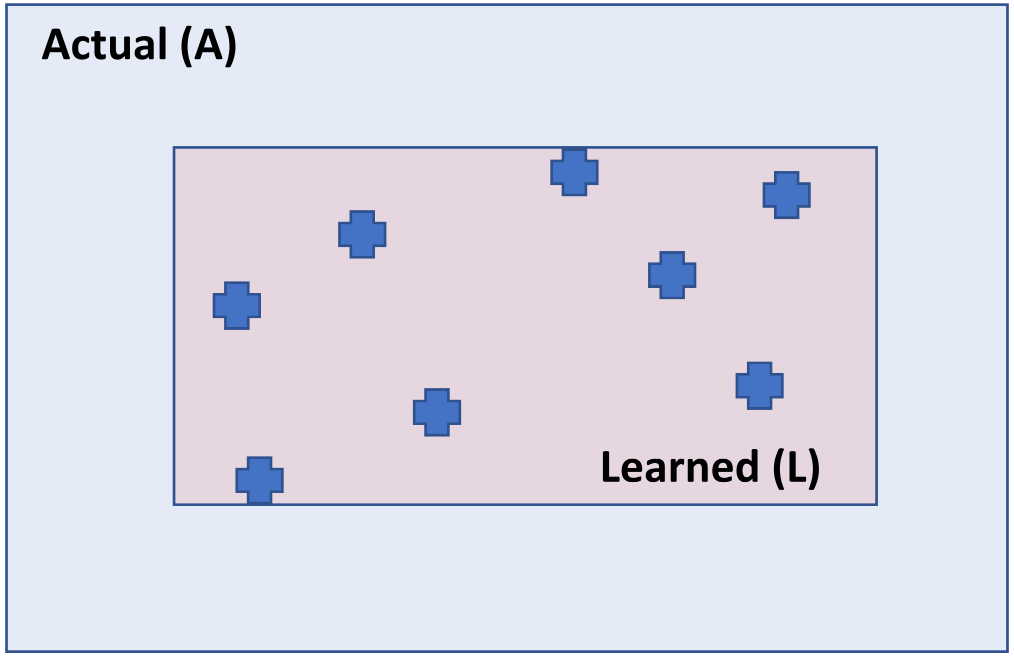
PAC Learning

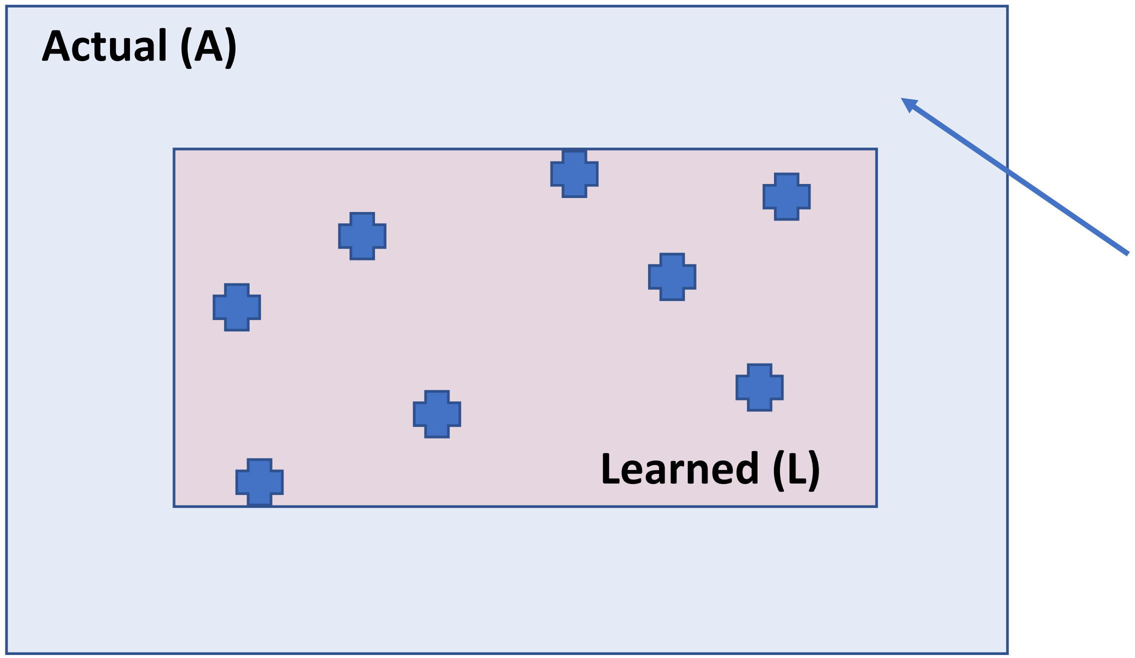
Recap: Rectangle Learning

- Given a set of *n* i.i.d. examples, find the best hypothesis rectangle
  - Agnostic setting – target concept is also a rectangle



Recap: Rectangle Learning

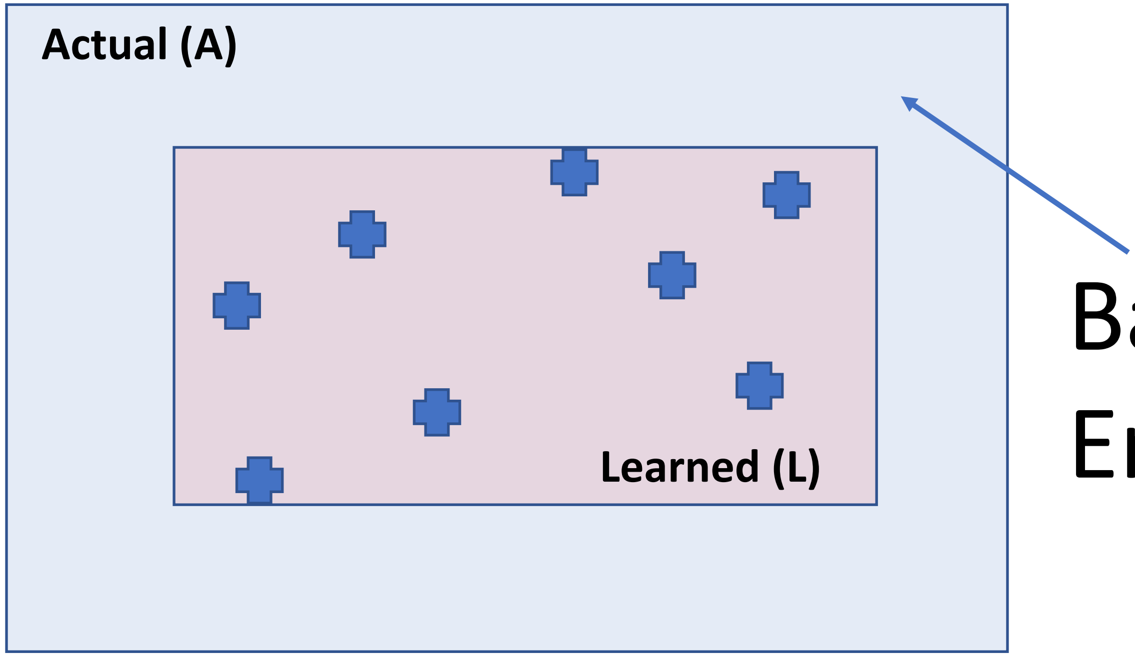
- Given a set of *n* i.i.d. examples, find the best hypothesis rectangle
  - Agnostic setting – target concept is also a rectangle



Error

Recap: Rectangle Learning

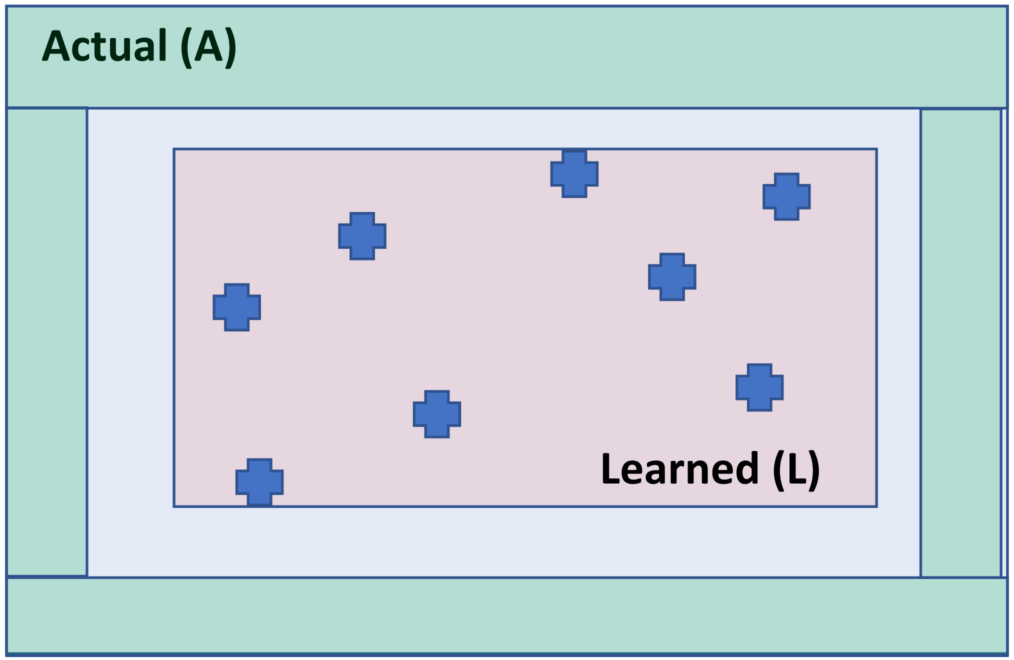
- Given a set of *n* i.i.d. examples, find the best hypothesis rectangle
  - Agnostic setting – target concept is also a rectangle



Bad Hypothesis:

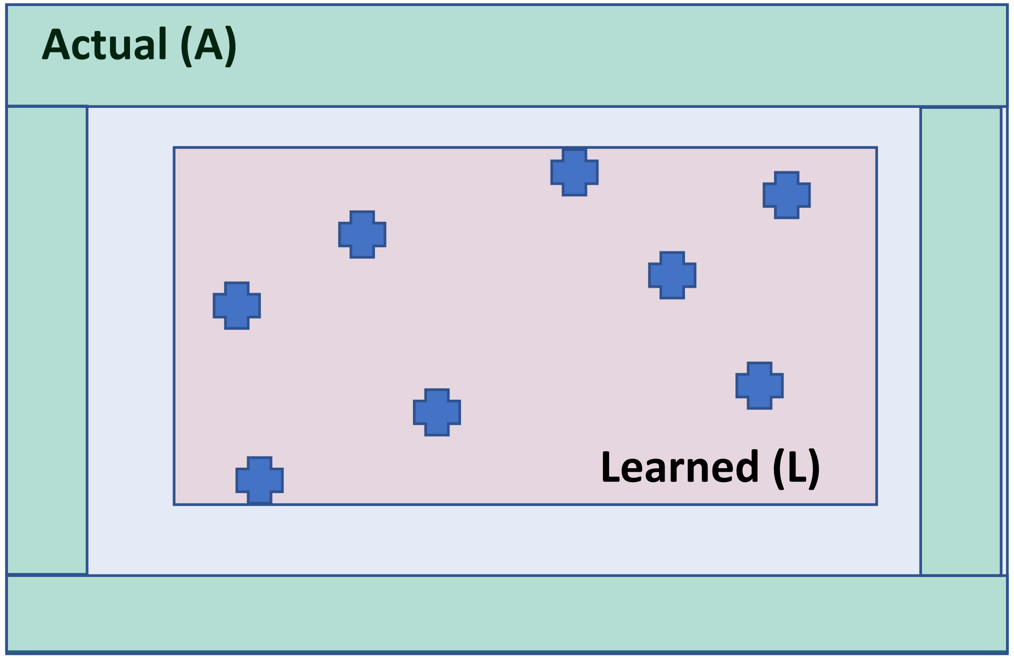
Error > ε

Recap: Rectangle Learning



Each strip has *area (prob.) = ε/4*

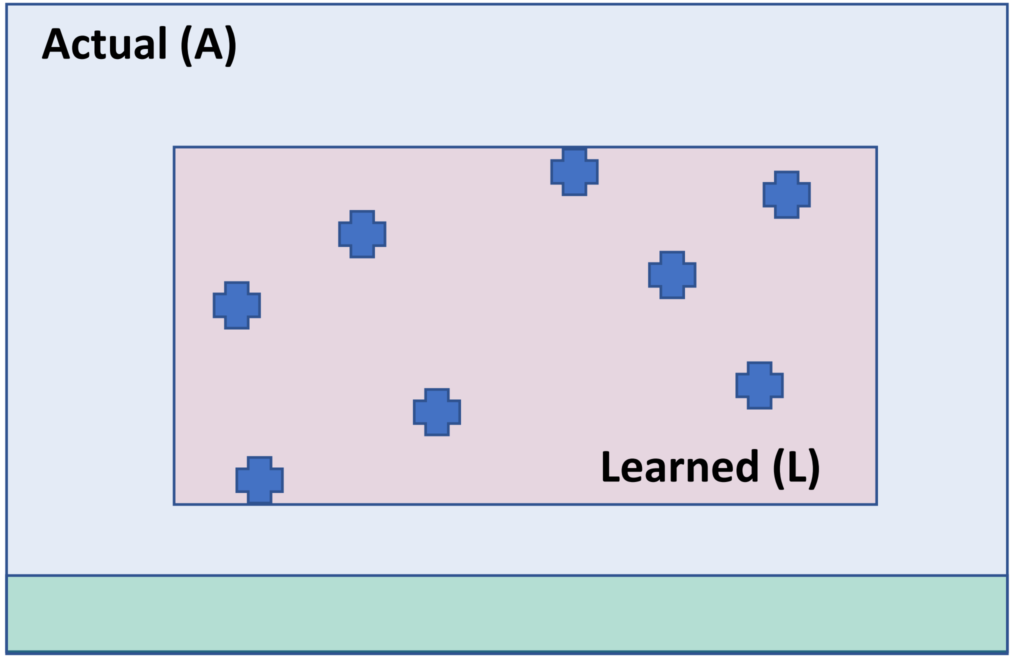
Recap: Rectangle Learning



How to get a bad hypothesis?

None of the *n* examples lie
In any of the strip

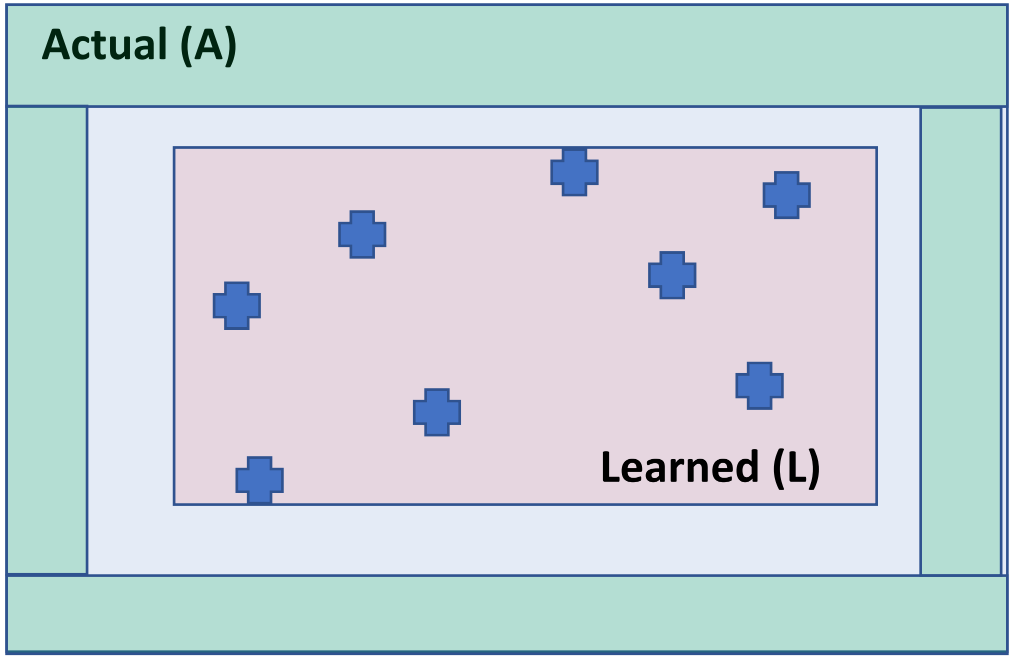
Recap: Rectangle Learning



What is the probability one example do not lie in one of the strip?

$(1 - \varepsilon/4)$

Recap: Rectangle Learning

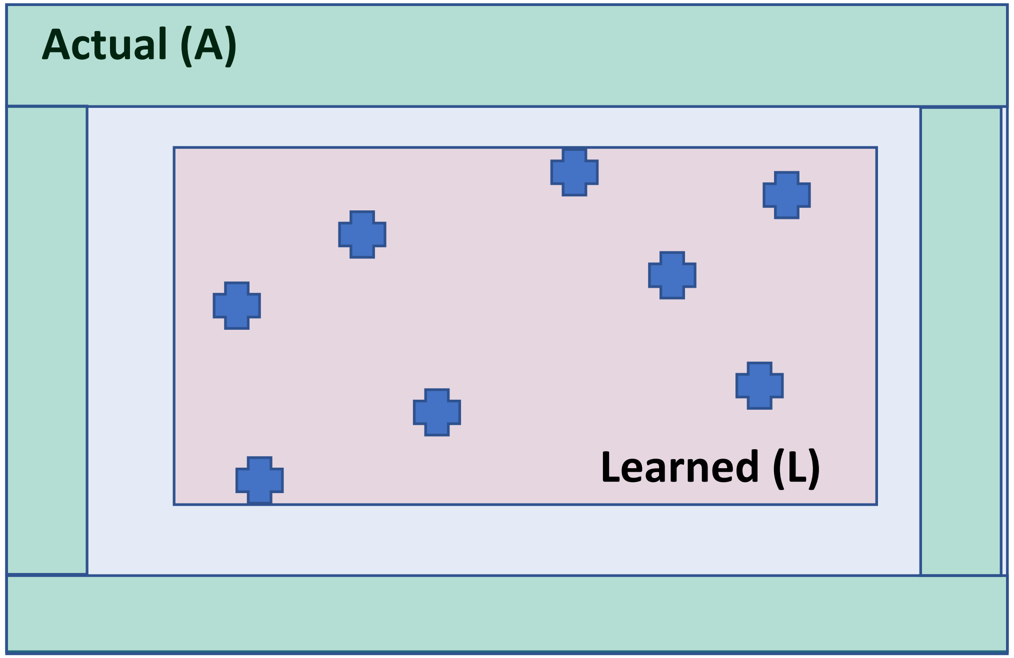


What is the probability none of the example do not lie in one of the strip? (i.i.d. examples)

$(1 - \varepsilon/4)^n$

(probability of obtaining a "bad hypothesis" after using n examples)

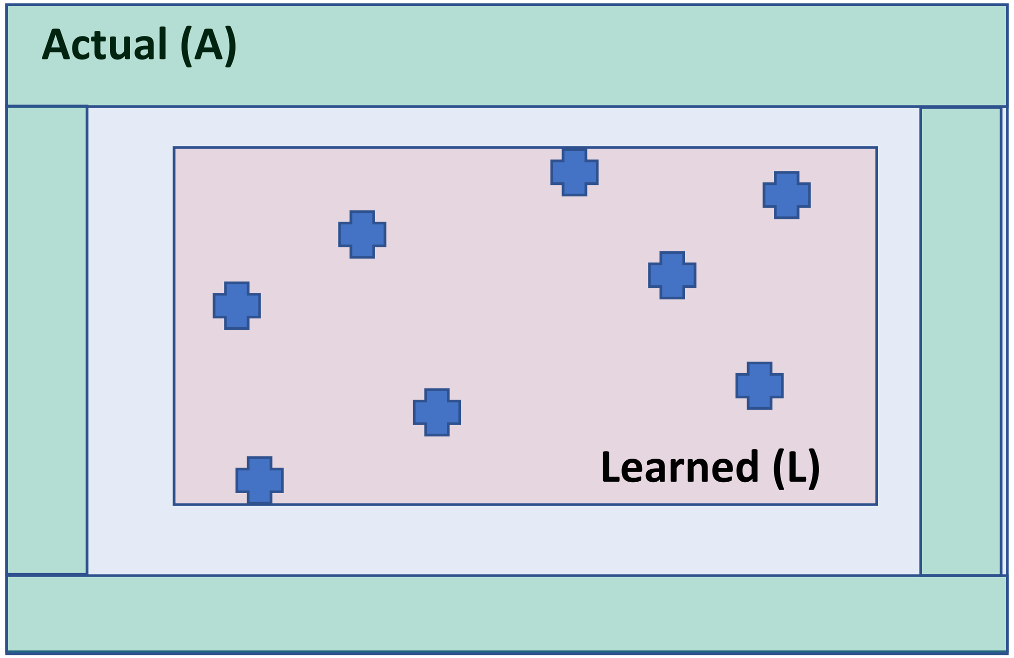
Recap: Rectangle Learning



Suppose we want the probability of obtaining a bad hypothesis to be <= δ :-

$4(1 - \varepsilon/4)^n <= \delta$

Recap: Rectangle Learning

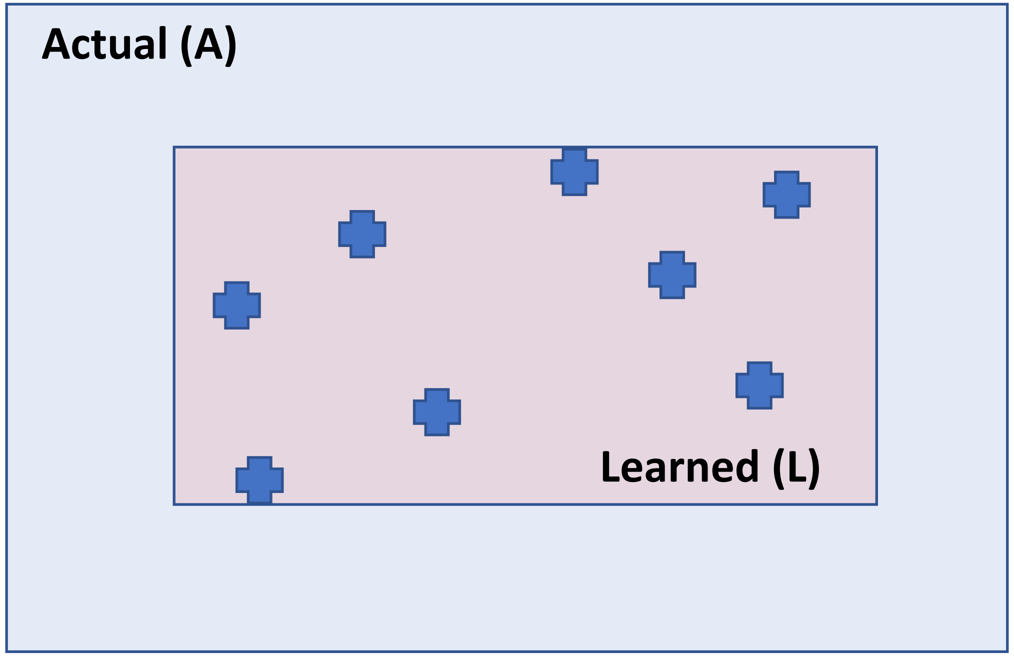


How many examples needed such that the probability of obtaining a bad hypothesis is <= δ :-

$$n \geq (4/\varepsilon)\ln(4/\delta)$$

$$(1-x) \leq e^{-x}$$

Correctness of Learned Hypothesis



The learned hypothesis is approximately correct with error <= *ε*

- *Always (for any training set)? No*

With a high probability *>= 1 – δ (Yes)*

*(*for any iid training set of size *n,* having same distribution*)*

Probably "Approximately Correct" (PAC)

A student is a good student if she gets more than 95% *(1- ε)* marks in 90% *(1 – δ)* of the courses.

$\varepsilon, \delta \rightarrow 0$, as $n \rightarrow \infty$ (asymptotically)

n should be finite for finite *ε, δ (sample complexity)*

Notations

- x – input, $x \in \mathbb{R}^d$
- y – output, $y \in \{0, 1\}$

- (Training) Sample – $S^m = \{(x_i, y_i), i = 1,\ldots,m\}$
  - *(independent and identically distributed)*

- Fixed but unknown data distribution: $\mathcal{D}$

Notations

- Function c, $f \in \mathcal{C}$, $\mathcal{F}$ (Function class, $x \in \mathbb{R}^d \rightarrow y \in \{0, 1\}$)
- Hypothesis $h \in \mathcal{H}$ (Hypothesis class)
- Target function: c, f
- Agnostic learning: $f \in \mathcal{H}$

Notations

- True Error: $Error_D = Pr_{(x,y) \in D}[h(x) \neg = y]$
- Empirical Error: $Error_S = Pr_{(x,y) \in S}[h(x) \neg = y] = \sum_{1,m} [h(x_i) \neg = y_i]$

PAC Learnability

Definition 1 *(PAC learnability) Let* $\mathcal{C}$ *be a class of boolean functions* $f : \{0,1\}^n \rightarrow 0,1$. *We say that* $\mathcal{C}$ *is PAC-learnable if there exists an algorithm* $\mathcal{L}$ *such that*
for every $f \in \mathcal{C}$
for any probability distribution $\mathcal{D}$
for any ϵ $(where 0 \leq \epsilon < \frac{1}{2})$
for any δ $(where 0 \leq \delta < 1$
algorithm $\mathcal{L}$ *on input* ϵ *and* δ *and a set of random examples picked from any probability distribution* $\mathcal{D}$ *outputs at least with a probability* $1 - \delta$, *concept* h *such that* $error(h, f) \leq \epsilon$.