PAC Model

- Domain X
- set of labels $Y = \{0,1\}$
- Hypothesis class $H \subseteq Y^X$
- Algorithm $A : (X \times Y)^* \to Y^X$

Definition: H is PAC-learnable iff there exists an algorithm A such that for every $\varepsilon, \delta \in (0,1)$ there exists $m \geq 0$ such that such that for any distribution D over X and any target function $h^* \in H$

$$\Pr\left[err_{D,h^*}(A(S)) \leq \varepsilon \right] \geq 1-\delta$$

where $S = \left((x_1, h^*(x_1)), \ldots, (x_m, h^*(x_m)) \right)$

and $x_1, x_2, \ldots, x_m \sim_{iid} D$.

<u>Definition</u>: H is <u>agnostically PAC-learnable</u> iff there exists an algorithm A such that for every $\epsilon, \delta \in (0,1)$ there exist $m \geq 0$ such that for every distribution D over $X \times Y$

$$\Pr\left[err_D(A(S)) \leq \inf_{h \in H} err_D(h) + \varepsilon \right] \geq 1 - \delta$$

where $S = \left((x_1, y_1), \ldots, (x_m, y_m) \right) \sim D^m$.

- PAC = Probably Approximately Correct

- PAC model is due to Leslie Valiant (1984). He got Turing Award in 2010 for it.

- Agnostic PAC model David Haussler (1990)

Sample complexity

- If $\Pr[err_{D,h^*}(A(S)) \leq \varepsilon] \geq 1-\delta$

we say A <u>(ε,δ) - learns</u>

target h^* on distribution D.

- sample complexity, $m(A, h, D, \varepsilon, \delta)$, is the smallest m such that A (ε, δ)-learns h on D.

- Sample complexity of learning H on D using A:

$$m(A, H, D, \varepsilon, \delta) = \sup_{h \in H} m(A, h, D, \varepsilon, \delta)$$

- Sample complexity of learning H on D

$$m(H, D, \varepsilon, \delta) = \inf_{A} m(A, H, D, \varepsilon, \delta)$$

- Sample complexity of learning H

$$m(H, \varepsilon, \delta) = \sup_{D} m(H, D, \varepsilon, \delta)$$

- Other combinations of parameters are possible.
  [This appears in research literature. E.g. sample complexity of a specific algorithm, etc.]

- For agnostic PAC model the starting point is (ε, δ)-learning of H by A.

Then

$m(A, H, D, \varepsilon, \delta)$ is defined

since D already encodes the labels.

The remaining definitions are the same.

Note however the sample complexities are different.

- We have already seen that any finite H is PAC-learnable.

- We have seen that class of <u>all</u> functions on an infinite domain is <u>not</u> PAC-learnable.

- Foreshadowing:

<u>Theorem</u>:

- H is PAC-learnable iff $VC(H)$ is finite

- H is agnostically PAC-learnable iff $VC(H)$ is finite

- We will also characterize sample complexity

$$m \approx \frac{VC(H) + \ln(1/\delta)}{\varepsilon} \quad \text{(PAC)}$$

$$m \approx \frac{VC(H) + \ln(1/\delta)}{\varepsilon^2} \quad \text{(Agnostic PAC)}$$