Last time:  * WM alg.  
* RWM \xrightarrow{\text{"do almost as well as best expert"}}

\[ \text{Done } \neg \text{OLMB!} \]

Today: Probably Approximately Correct (PAC) Learning Model. \textcircled{Valiant '84}

KV Book 1.1-1.3, all Chap. 1, approx. prob. inex.

\[ \text{Drawbacks of OLMB:} \]

- Very worst-case: any ex seq. \( \text{ex. seq.} \)
- OLMB measures perf. from day 1. Hard
- \( \emptyset \) guarantee: can be unsat.

Motiv. for PAC model: realistic & helpful to assume statistical regularity in examples.

Hi-level idea of PAC setting:

* Assume the ex \( x \in X \) that learner gets are \textcircled{indep} drawn from some fixed (unknown)-underlying prob. dist. \( D \) \rightarrow model of the world

* "batch" model: get data set of indep. labeled \((x,c(x))\) pairs. Alg. ponders, outputs an \( h : X \rightarrow \{0,1\} \)\( \xrightarrow{\text{not online}} \)
Goal: construct a hyp $h$ that accurately pred. value of $c$ on ex. drawn from $\mathcal{D}$.

PAC framework for learning $c: \mathcal{C}$ using a hypothesis class $\mathcal{H}$:
- as before, there's a fixed unknown $c \in \mathcal{C}$
- Also a fixed unknown dist. $\mathcal{D}$ over $X$.
(Learner "knows $\mathcal{C}$").
- Learner is given a training set of $m$ labeled ex $(x, c(x))$ where each $x$ indep. drawn from $\mathcal{D}$.
  Equivalently: learning access an example oracle $EX(c, \mathcal{D})$
  
  $EX(c, \mathcal{D}) \rightarrow (x, c(x)) \sim \mathcal{D}$.

- Learner computes, $\star$ outputs a $h: X \rightarrow \{0,1\}$ where $h \in \mathcal{H}$.

**Def:** let $h, c : X \rightarrow \{0,1\}$; let $\mathcal{D}$ be dist. over $X$. The error of $h$ on $c$ w.r.t. $\mathcal{D}$ is

$$err_{\mathcal{D}}(h,c) := \Pr_{x \sim \mathcal{D}}[h(x) \neq c(x)].$$
What can we hope for?

- 0 error? No; \( \mathcal{D} \) may put tiny wt on some portion of \( X \). High acc (low error)...
- Definitely get low-error \( h \)? No; every one of \( m \) ex. from \( \mathcal{D} \) is.

CAN hope to, with high prob. achieve low-error hyp.

Def (Prelim. def.) "Alg. \( A \) PAC learns \( c \subset \mathcal{C} \) using hyp. class \( \mathcal{H} \) with \( m \) examples" means:

\[
\forall c \in \mathcal{C} \text{ (target)},
\forall \text{ dist } \mathcal{D} \text{ over domain } X,
\forall \text{ params } 0 < \delta, \epsilon < 1
\]

If \( A \) is given \( \epsilon, \delta \) + access to \( EX(c, \mathcal{D}) \)
\( A \) draws \( \leq m \) ex. from \( EX(c, \mathcal{D}) \) w.p. \( > 1 - \delta \)
(over \( m \) draws & any internal rand. of \( A \)), \( A \) outputs a \( h : \mathcal{H} \) s.t. \( \text{err}_g(h, c) \leq \epsilon \). "distribution-indep. learning".
Notes: \(\varepsilon = \text{"accuracy param."}\)

\(J = \text{"confidence param."}\)

\(m = \text{"sample complexity"}\)

Runtime? Call to \(\text{EX}\) one time step.

Efficient alg: \(\text{poly}\left(\frac{1}{\varepsilon}, \frac{1}{\delta}\right)\) \((\log \frac{1}{\delta})\)

Often \(X=\mathbb{R}^n, \{0,1\}^n:\) \(\text{poly}\left(\gamma, \frac{1}{\varepsilon}, \frac{1}{\delta}\right)\)

Note runtime always \(\geq (\#\text{samp. used})\)

\(m\).

\(\Rightarrow\) If \(\mathcal{H}=\mathcal{C}\), "proper" learning alg (by-the-book)

Example: PAC learning intervals.

\(X=[0,1], \mathcal{C} = \{[a,b]: a \leq b \in [0,1]\}\) = intervals.

\(c = [.42, .67]: c(.5) = 1, c(.8) = 0\)

There's some \(D\) over \([0,1]\); \(\mathcal{C}\) doesn't know it.

Target interval, say, is \(c = [a, b].\)

(Note \(h \leq c\)).
Alg. A is:

- draw \( \mathcal{D} \) ex from EX(c,D)
- Let \( a' = \text{leftmost} + (\text{value in } [0,0]) \)
  \( b' = \text{rightmost} + (\text{''''''}) \)

A outputs \( h = [a', b'] \).

\[
err_{\mathcal{D}}(h, c) = \Pr[ x \in [a, a'] \cup (b', b)] \sim \mathcal{D}
\]

\[
a_1 = \text{pt s.t. } \Pr[ x \in [a_1, a_1] ] = \frac{\varepsilon}{2} \sim \mathcal{D}
\]
\[
b_1 = \text{pt s.t. } \Pr[ x \in (b_1, b_1] ] = \frac{\varepsilon}{2} \sim \mathcal{D}
\]

As long as \( A \)’s \( \text{ex.} \) include both an \( \text{ex.} \) in \( L \) and an \( \text{ex.} \) in \( R \),

\[
err_{\mathcal{D}}(h, c) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon
\]

Consider a single rand. \( \text{ex.} \) \( x \sim \mathcal{D} \).

\[
Pr[ x \text{ misses } L ] = 1 - \frac{\varepsilon}{2}.
\]

So

\[
Pr[ \text{in indep. draws from } \mathcal{D} \text{ all miss } L ] = \left(1 - \frac{\varepsilon}{2}\right)^n
\]

\[
\leq e^{-\frac{\varepsilon}{2}}.
\]

Likewise for \( R \).

So (u.b.)

\[
Pr[ \text{in indep. ex. from } \mathcal{D} \text{ either all miss } L \text{ or all miss } R ]
\]
\( \leq \left( 2e^{-\frac{\epsilon m}{2}} \right) \) \[ \text{Want} \] \( \leq \delta. \)

Taking \( m = \frac{2}{\epsilon} \cdot \ln\left( \frac{2}{\delta} \right) \) suffices.

For this \( m \), w.p. \( \geq 1 - \delta \), an ex. hits \( \subseteq \), an ex. hits \( \subseteq \).

So w.p. \( \geq 1 - \delta \), \( \text{err}_E(h, c) \leq \epsilon. \)

So we showed \( A \) is a PAC alg. for \( C \).

(\text{efficient one}: \( m = \frac{2}{\epsilon} \cdot \ln\left( \frac{2}{\delta} \right) \))

\[ \text{A A R:} \quad \] controlled against getting bad sample of \( x \)'s.

\( \cdot \epsilon = \text{acc. param. It's ok to fail on } \epsilon \text{ frac of } \mathcal{D} \) ("rare ex" missed during training phase).

\( \cdot \)