16 Mar 2022 Chernoff Bound Applications, Continued

A common pattern for using Chernoff bound.

- There's a finite # of "bad events" we want to avoid

  E.g. in ERM analysis, one bad event is

  $$\left| \frac{1}{N} \sum_{j=1}^{N} L(h_i, z_j) - \mathbb{E}[L(h_i, z)] \right| > \frac{\varepsilon}{2}$$

  We have one such bad event for each $i \in [m]$.

- We want to show if N is large enough, with high probability none of the bad events happen.

- Game plan: Use Chernoff or Hoeffding to show $\Pr(\text{bad event } \# i) \ll 1$ for each i.

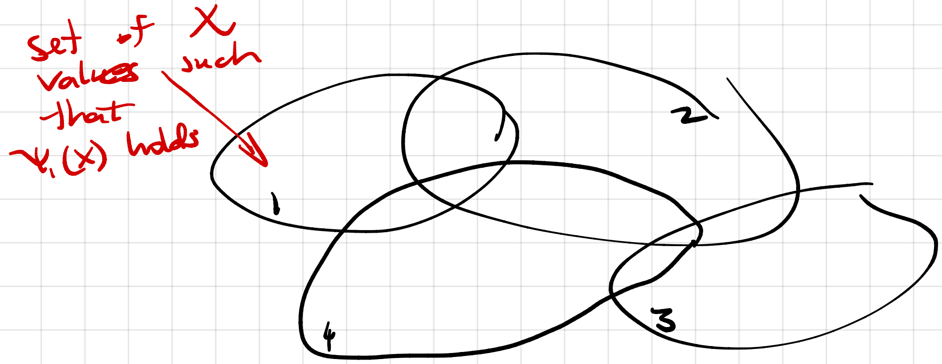
  Then, use "union bound" to show

  $$\Pr(\exists i \text{ s.t. bad event } \# i \text{ occurs}) < \delta.$$

<u>Union Bound:</u> If X is a random variable and $\psi_1, \psi_2, \ldots, \psi_m$ are Boolean predicates $\psi_i(X)$ then

$$\Pr\left(\bigvee_{i=1}^{m} \psi_i(X)\right) \leq \sum_{i=1}^{m} \Pr(\psi_i(X))$$

In plain English: the probability at least one of $\psi_1(X), \ldots, \psi_m(X)$ happens, is at most the sum of the probabilities of $\psi_1, \ldots, \psi_m$.



Proof. Let $\phi_i(X) = \Psi_i(X) \wedge \left(\bigwedge_{k=1}^{i-1} \overline{\Psi_k(X)} \right)$

The predicates $\phi_1, \phi_2, \ldots, \phi_m$ are mutually exclusive and $\bigvee_{i=1}^{m} \phi_i = \bigvee_{i=1}^{m} \Psi_i$.

Probability is finitely additive:

$$\Pr\left(\bigvee_{i=1}^{m} \phi_i \right) = \sum_{i=1}^{m} \Pr(\phi_i)$$

Also, $\Pr(\Psi_i) \geq \Pr(\phi_i)$ because $\Psi_i(X) = \text{TRUE}$ whenever $\phi_i(X) = \text{TRUE}$

$$\Pr\left(\bigvee_{i=1}^{m} \Psi_i \right) = \Pr\left(\bigvee_{i=1}^{m} \phi_i \right) = \sum_{i=1}^{m} \Pr(\phi_i) \leq \sum_{i=1}^{m} \Pr(\Psi_i)$$

Going back to ERM analysis we chose # samples, N, large enough that we could show using Hoeffding,

$$\forall h_i \quad \Pr\left(\left| \frac{1}{N} \sum_{j=1}^{N} L(h_i, z_j) - \mathbb{E}[L(h_i, z)] \right| > \frac{\varepsilon}{2} \right) < \frac{\delta}{m}$$

Union Bound:

$$\Pr\left(\exists i \ \left| \frac{1}{N} \sum_{j=1}^{N} L(h_i, z_j) - \mathbb{E}[L(h_i, z)] \right| > \frac{\varepsilon}{2} \right) < m \cdot \frac{\delta}{m} = \delta$$

Using Chernoff Bound in analysis of randomized algorithms.

Decision Problem: Problem that has $\{0,1\}$ answer.
(Equivalently a $\{\text{FALSE}, \text{TRUE}\}$ answer.)

P: Class of decision problems that have a deterministic, poly-time algorithm that always outputs correct answer.

BPP: Decision problems that have a randomized poly-time algorithm that always outputs a $\{0,1\}$ answer, and answers correctly with prob $\geq \frac{2}{3}$ on every possible input.

Reducing error probability of BPP algorithms.

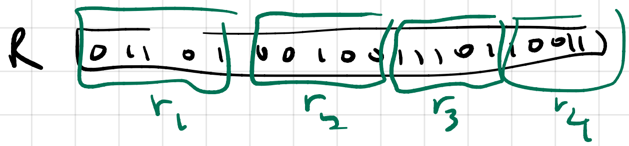
Suppose we have algo. $A(\overset{\text{input}}{x}, \underset{\text{random bits}}{r})$

s.t. $\forall x$ $\Pr(A(x,r) \text{ is correct}) \geq 2/3$.

We'd like to create algo $B(x, \overset{\text{more random bits}}{R})$

s.t. $\forall x$ $\Pr(B(x,R) \text{ is correct}) \geq 1-\delta$.

$B(x,R)$: partition random string R into $r_1, r_2, \ldots, r_m$ $\quad m = O(\log(\frac{1}{\delta}))$.



Run $A(x,r_1), A(x,r_2), \ldots, A(x,r_m)$.

Take majority vote: if $\sum_{i=1}^{m} A(x,r_i) \geq \frac{m}{2}$ output 1, else output 0.

$\forall x$ $\quad \mathbb{E}[A(x,r)] \geq 2/3$ if correct ans is 1.

$\qquad \mathbb{E}[A(x,r)] \leq \frac{1}{3}$ if correct ans. is 0

Estimating $\mathbb{E}(A(x,r))$ within additive error $\frac{1}{6}$ implies $B(x,R)$ outputs correct answer.

To get $\varepsilon = \frac{1}{6}$ accuracy with prob $\geq 1-\delta$, need

$$m \geq \frac{1}{2\varepsilon^2} \ln\left(\frac{2}{\delta}\right) = 18 \ln\left(\frac{2}{\delta}\right).$$

Suppose input x has length n bits.

And suppose we want $< 2^{-n}$ probability of error.

$$\frac{2}{\delta} = 2^{n+1}, \qquad 18 \ln\left(\frac{2}{\delta}\right) = 18(n+1)\ln(2) < 14(n+1).$$

We have shown when $m = 14(n+1)$,

$B(x,R)$ runs in time $14 \cdot (n+1) \cdot \mathrm{TIME}(A(x,r))$.

and $\forall x \in \{0,1\}^n$,

$$\Pr(B(x,R) \text{ is wrong on } x) < 2^{-n}.$$

Sum over $x \in \{0,1\}^n$

$$\Pr(\exists x \in \{0,1\}^n \text{ s.t. } B(x,R) \text{ is wrong on } x) < 1.$$

$\therefore$ $\exists$ a string R_n such that

$$\forall x \in \{0,1\}^n \quad B(x, R_n) \text{ is correct.}$$

P/poly: decision problems with a 2-variable algorithm $B(x,y)$ running in time poly$(|x|)$ s.t. $\forall n$ $\exists$ string y_n s.t. $\forall x \in \{0,1\}^n$ $B(x, y_n)$ is correct.

BPP $\subseteq$ P/poly.