9 Feb 2026

Hoeffding's Bound & Applications

Hoeffding Bound: $X_1, \dots, X_n$ independent

$X = X_1 + \dots + X_n$

$X_i \in [a_i, b_i] \quad \forall i$

For any $\lambda > 0$

$$\Pr(X \geq \mathbb{E}X + \lambda) \leq \exp\left(-\frac{2\lambda^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

$$\Pr(X \leq \mathbb{E}X - \lambda) \leq \exp\left(-\frac{2\lambda^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

Applications

1. Estimating the average of a distribution.

Say we draw N rand samples

$Y_1, \dots, Y_N$

from distribution D on $[0, M]$.

Population average: $y = \mathbb{E}_{Y \sim D}[Y]$

Empirical average: $\hat{y} = \frac{1}{N}(Y_1 + \dots + Y_N)$.

Q: How large must N be so that

$$\Pr(|\hat{y}-y|>\varepsilon) < \delta \,?$$

Answer is a function of ε, δ, M and is called the sample complexity of (ε,δ)-PAC mean estimation.

"probably approximately correct"

$$\hat{y} = \underbrace{\frac{Y_1}{N}}_{X_1} + \underbrace{\frac{Y_2}{N}}_{X_2} + \cdots + \underbrace{\frac{Y_N}{N}}_{X_N}$$

$$\mathbb{E}\hat{y} = \frac{\overbrace{\mathbb{E}Y_1}^{=y}}{N} + \frac{\overbrace{\mathbb{E}Y_2}^{=y}}{N} + \cdots + \frac{\overbrace{\mathbb{E}Y_N}^{=y}}{N} = y$$

$$X_i \in [\overbrace{0}^{a_i}, \overbrace{\tfrac{M}{N}}^{b_i}] \qquad (b_i-a_i)^2 = \frac{M^2}{N^2}.$$

Hoeffding:

$$\Pr(\hat{y}-y \geq \varepsilon) \leq \exp\left(-\frac{2\varepsilon^2}{N\cdot(M^2/N^2)}\right)$$

$$= \exp\left(-\frac{2\varepsilon^2 N}{M^2}\right)$$

$$\Pr(\hat{y}-y \leq -\varepsilon) \leq \exp\left(-\frac{2\varepsilon^2 N}{M^2}\right)$$

$$\Pr(|\hat{y}-y| \geq \varepsilon) \leq 2\cdot\exp\left(-\frac{2\varepsilon^2 N}{M^2}\right)$$

want this $< \delta$ .... So make this $< \delta$

$$2\exp\left(-\frac{2\epsilon^2 N}{M^2}\right) < \delta$$

$$\exp\left(\frac{2\varepsilon^2 N}{M^2}\right) > \frac{2}{\delta}$$

$$\frac{2\varepsilon^2 N}{M^2} > \ln\left(\frac{2}{\delta}\right)$$

$$\boxed{N > \frac{1}{2}\left(\frac{M}{\varepsilon}\right)^2 \ln\left(\frac{2}{\delta}\right)}$$

(2) Sample complexity of (ε,δ)-PAC ERM.

Given N samples $z_1, \dots, z_N$ drawn i.i.d. from $\mathcal{D}$ on $\mathcal{Z}$.

Given m hypotheses $h_1, h_2, \dots, h_m \in \mathcal{H}$.

Given loss function

$$L : \mathcal{H} \times \mathcal{Z} \longrightarrow [0,1]$$

$L(h,z) =$ "how badly does h explain z?"

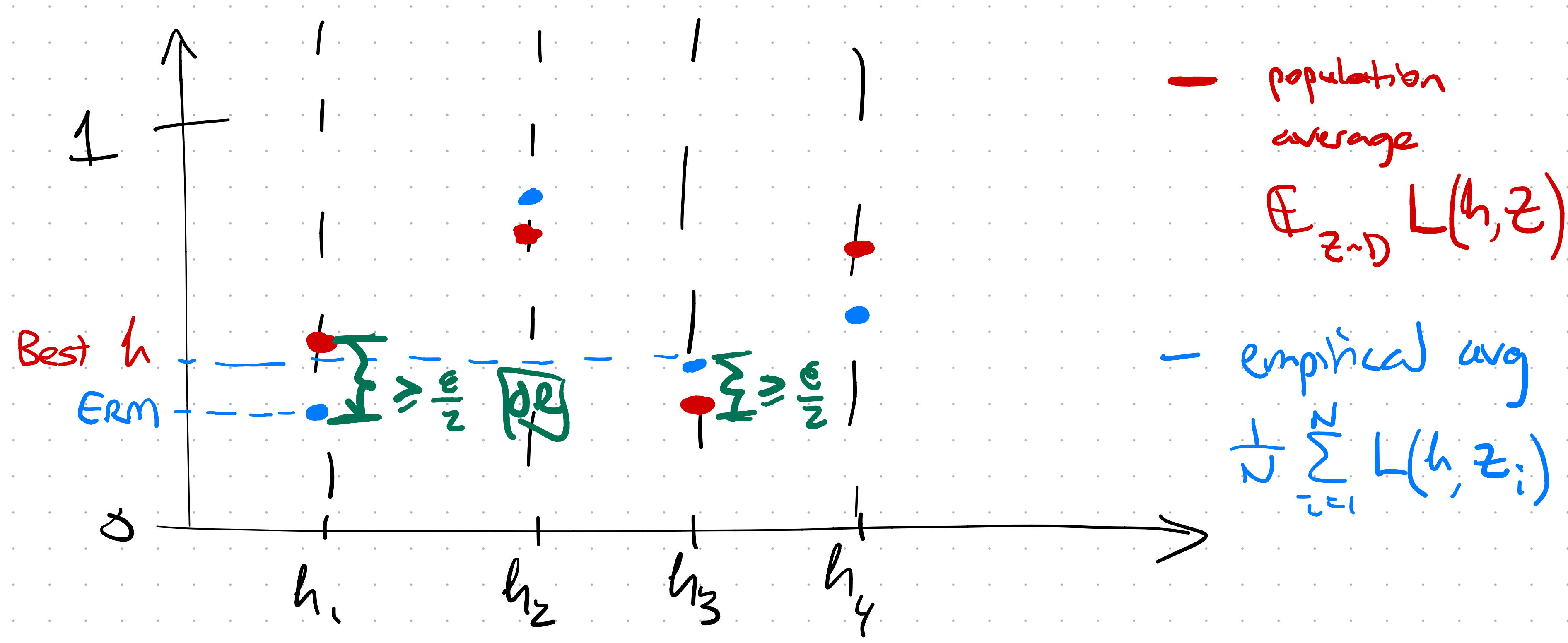
ERM is the algorithm that selects a hypothesis using the rule

$$h_{ERM} = \underset{h \in \{h_1, \dots, h_m\}}{\arg\min} \left[\frac{1}{N}\sum_{i=1}^{N} L(h, z_i)\right]$$

i.e. select from $\{h_1, \dots, h_m\}$ the one with least average loss on the data.

We say ERM is (ε, δ)-PAC if the following holds with probability $> 1-\delta$...

$$(*) \quad \mathbb{E}_{Z\sim D}\left[L(h_{ERM}, Z)\right] \le \varepsilon + \min_{i\in[m]} \mathbb{E}_{Z\sim D}\left[L(h_i, Z)\right]$$



<u>Upshot:</u> $(*)$ holds as long as we're sure that $\forall i \in [m]$

$$\left|\frac{1}{N}\sum_{j=1}^{N} L(h_i, z_j) - \mathbb{E}_{Z\sim D}\left(L(h_i, Z)\right)\right| \le \frac{\varepsilon}{2}$$

m "bad events", one per $i \in [m]$.

Bad event: average of N iid variables in $[0,1]$ differs from pop avg by $> \varepsilon/2$.

<u>Goal:</u> $\Pr(\text{none of these happen}) < \delta$.

$\Downarrow$ $\forall$ bad event $\Pr(\text{bad event}) < \delta/m$.

Plugging into earlier formula

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

$$\boxed{N \geq \frac{2}{\varepsilon^2} \ln\left(\frac{2m}{\delta}\right).}$$

sample cplx for (ε,δ)-PAC ERM with m hypotheses.

3) Reducing error rate of randomized algs.

<u>Decision problem</u>. Problem Π with $\{0,1\}$ answer.

We say Π belongs to BPP if

$\exists$ a poly-time algorithm

$A(x, r)$

x: problem input

r: random bits $r \in \{0,1\}^n$

s.t. $\forall x$ if $\Pi(x)=0$ $\Pr_{r\in\{0,1\}^n}[A(x,r)=0] \geq \frac{2}{3}$

if $\Pi(x)=1$ $\Pr_{r\in\{0,1\}^n}[A(x,r)=1] \geq \frac{2}{3}$.

<u>Goal</u>. Boost success prob from $2/3$ to $1-\delta$.

<u>Plan</u>. Use random string $R=(r_1, r_2, \ldots, r_N)$.

each unif. rand in $\{0,1\}^n$.

Algo $B(x,R) = \mathrm{MAJ}\left[A(x,r_1), A(x,r_2), \ldots, A(x,r_N)\right]$.

If $B(x,R)$ is wrong,

$\mathbb{E}[\#$ correct ans. among $A(x,r_1), \ldots, A(x,r_N)]$

$\mathbb{E}X \geq \frac{2N}{3}$

Realized # correct ans. $= X_1 + \cdots + X_N = X$

$\leq \frac{N}{2}$

$$X_i = \begin{cases} 1 & \text{if } A(x,r_i) \text{ correct} \\ 0 & \text{if not.} \end{cases}$$

$(b_i - a_i)^2 = (1-0)^2 = 1$

λ

$$\Pr(B(x,R) \text{ incorrect}) \leq \Pr\left(X \leq \mathbb{E}X - \frac{N}{6}\right)$$

$$\leq \exp\left(-\frac{2(N/6)^2}{N}\right)$$

$$= \exp\left(-\frac{N}{18}\right) < \delta$$

$$N > 18 \ln(1/\delta).$$

So if rand algo can solve Π with $1/3$ prob(err) using n rand bits, then $18n\ln(1/\delta)$ rand bits suffice for $\leq \delta$ prob(err). $\Longrightarrow$ $\mathrm{BPP} \subseteq \mathrm{P/poly}$