Last time:

Expert $1, \dots, N$

$1/N \quad \dots \quad 1/N$

$\underline{w_1^1} \quad \dots \quad \underline{w_N^1}$

$\beta \in (0, 1)$

$t = 1, \dots, T$

$l_1^t \quad \dots \quad l_N^t \in [0, 1]$

$$w_i^{t+1} \leftarrow w_i^t \cdot \beta^{l_i^t}$$

$l_i^t = 0, \quad w_i^{t+1} = w_i^t$

$l_i^t = 1, \quad \leftarrow w_i^t \cdot \beta$

Goal:

$\underline{\text{Alg}} \approx$ best expert.

$$\boxed{L_i \triangleq \text{loss of expert } \underline{i} = \sum_{t=1}^{T} \underline{l_i^t}}$$

(if you ~~fool~~ follow expert i at the very beginning.

$\rightarrow$ loss of your alg).

Alg: $w_1^t \; \cdots \; w_N^t$

$$\rightarrow \left(\frac{w_1^t}{\sum_{i=1}^{N} w_i^t}, \; \cdots, \; \frac{w_N^t}{\sum_{i=1}^{N} w_i^t} \right)$$

loss of Alg at $\boxed{t\text{-iter}}$:

$$E_{i \sim (w_1^t \cdots w_N^t)} [\, l_i^t \,]$$

$$= \frac{w_1^t}{\sum w_i^t} \cdot l_1^t + \frac{w_2^t}{\sum w_i^t} \cdot l_2^t + \cdots + \frac{w_N^t}{\sum w_i^t} \cdot l_N^t$$

$$= \frac{1}{\sum_{i=1}^{N} w_i^t} \left[\boxed{\sum_{i=1}^{N} w_i^t \cdot l_i^t} \right]$$

Thm: (informal)

$$L_A \le L_i + \log N$$

$$\forall i \in \{1, \dots, N\}$$

$$\frac{1}{T} L_A \le \frac{1}{T} L_i + \frac{\log N}{T}$$

$T \to \infty$. $\qquad \to 0$

	A	i^*
1	100	5
2	90	6
	⋮	⋮
30.	4	4.1

Pf:

$$\frac{1}{T} \cdot L_A - \frac{1}{T} \cdot L_i \leq \frac{\log N}{T} \to 0$$

$$= \frac{1}{T} \sum_{t=1}^{T} \left(l_A^t - l_i^t \right) \to 0$$

as $T \to \infty$.

$A \qquad\qquad i$

$8 \longleftrightarrow 6$

$5 \longleftrightarrow 7$

$2 \longleftrightarrow 4$

$6 \longleftrightarrow \varphi$

$\vdots \qquad\qquad \vdots$

necessary

finite iters

$$|Alg - i^*| \geq \square$$

Pf:

$$\sum_{i=1}^{N} w_i^{t+1} \geq \square \leftarrow L_i$$

$$\sum_{i=1}^{N} w_i^{t+1} < \square \leftarrow L_A$$

Chaining: $\square \leftarrow L_A \geq \square \leftarrow L_i$

---

① $\sum_{i=1}^{N} w_i^{T+1} \geq w_i^{T+1}$

key: all $w_i^{T+1} > 0$.

$$w_i^{t+1} \leftarrow w_i^{t} \cdot \beta^{L_i^t} = w_i^{T} \cdot \beta^{L_i^T}$$

$$= w_i^{T-1} \cdot \beta^{L_i^{T-1}} \cdot \beta^{L_i^T}$$

$$= \cdots$$

$$= w_i^1 \cdot \beta^{L_i^1} \cdot \beta^{L_i^2} \cdot \ldots \cdot \beta^{L_i^T}$$

$$= \frac{1}{N} \cdot \beta^{\sum_{t=1}^{T} L_i^t}$$

$$= \frac{1}{N} \cdot \beta^{L_i}$$

$$\boxed{\sum_{i=1}^{N} w_i^{T+1} \geq \frac{1}{N} \cdot \beta^{L_i}}$$

$$\underline{\sum_{i=1}^{N} w_i^{T+1}} = \sum_{i=1}^{N} w_i^T \cdot \underline{\beta}^{L_i^T}$$

$$\boxed{\begin{array}{l}\text{Fact:} \\ \underline{\alpha^r} \le 1 - (1-\alpha)\cdot r \\ \forall\ \alpha > 0,\quad r \in [0,1]\end{array}}$$

$$\le \sum_{i=1}^{N} w_i^T \cdot \left[1 - (1-\beta)\cdot L_i^T\right]$$

$$= \sum_{i=1}^{N} w_i^T - (1-\beta)\cdot \sum_{i=1}^{N} \underline{w_i^T \cdot L_i^T}$$

$$= \sum w_i^T - (1-\beta)\left(\sum w_i^T\right)\cdot \frac{1}{\sum w_i^T} \cdot \frac{\sum w_i^T \cdot L_i^T}{}$$

$$= L_A^T$$

$$= \sum w_i^T - (1-\beta)\cdot \sum w_i^T \cdot L_A^T$$

$$= \sum w_i^T \left[1 - (1-\beta)\cdot L_A^T \right]$$

$$\Rightarrow \sum w_i^{T+1} \le \underline{\sum w_i^T} \cdot \left[1 - (1-\beta) L_A^T \right]$$

$$\boxed{\begin{array}{l} \forall x \in \mathbb{R} \\ 1 + x \le e^x \end{array}}$$

$$\le \underline{\sum w_i^{T-1}} \cdot \left[1 - (1-\beta)\cdot L_A^{T-1} \right] \cdot \left[1 - (1-\beta) L_A^T \right]$$

$$\underbrace{-(1-\beta)\cdot L_A^t}$$

$$\le \overbrace{\sum w_i^1}^{=1} \prod_{t=1}^{T} \left[1 - (1-\beta) L_A^t \right]$$

$$\le \prod_{t=1}^{T} e^{-(1-\beta)\cdot L_A^t}.$$

$$= e^{-(1-\beta)\cdot \sum_{t=1}^{T} L_A^t} \quad L_A$$

$$\boxed{\sum w_i^{T+1} \le e^{-(1-\beta)\cdot L_A}}$$

$$\sum w_i^{T+1} \ge \frac{1}{N}\cdot \beta^{L_i}$$

$$\Rightarrow \frac{1}{N}\cdot \beta^{L_i} \le e^{-(1-\beta)\cdot L_A}$$

$$\Rightarrow e^{(1-\beta)L_A} \le \frac{N}{\beta^{L_i}}$$

$$\Rightarrow (1-\beta)\, L_A \le \log N + L_i \cdot \log \frac{1}{\beta}$$

$$\Rightarrow L_A \leq \frac{\log \frac{1}{\beta}}{1-\beta} L_i + \frac{\log N}{1-\beta}$$

$$\beta \in (0,1) \longrightarrow \beta = 1/2.$$

$$\Rightarrow L_A \leq 2\log 2 \cdot L_i + 2\log N$$

$\square$

instance space.

$$X = \mathbb{R}^d$$

$$X = S^{d-1}$$

Label space

$$Y = \{-1, +1\}$$

$$\{0, 1\}$$

$$\cancel{\{1, 2, 3, 4, \ldots\}}$$

concept class (model)

$$h: X \to Y$$

$$H = \{h: X \to Y\}$$

$X = \mathbb{R}^2$. $Y = \{-1, 1\}$

$y = 1$ iff $x_1 > 0$.

$$h(x) = \begin{cases} 1, & \text{iff } x_1 > 0. \\ -1, & \text{otherwise.} \end{cases}$$

$$h(x) = \begin{cases} 1, & \text{iff } x_1 + x_2 > 0 \\ -1 & \text{else.} \end{cases}$$

....

$X \times Y$ $\quad 7 = 40$

D

all data are drawn from

D.

h^*: $\delta = 0.01$

draw of training

~~Goal~~

$\forall\, \epsilon \in (0,1),\ \delta \in (0,1)$

Goal: find $\hat{h}$, w.p. $1-\delta$,

s.t. $\underbrace{Pr_{(x,y)\sim D}}_{\text{draw of test}} \underbrace{(\hat{h}(x) \neq y)}_{\text{error rate of } h} \leq \epsilon$

Given: $(x_1, y_1) \ldots (x_n, y_n)$

$$\boxed{\frac{1}{n} \#\text{mistakes of } h = \frac{1}{n} \sum_{x} \mathbb{1}\{h(x_i) \neq y_i\}} \xrightarrow{n \to \infty} \Pr_{(x,y)\sim D} \left(h(x) \neq y \right).$$

100 pts.

$$y = \begin{cases} 1, & x > \underline{1/2}. \\ -1, & \text{else.} \end{cases}$$

Probably approx. correct learning

δ

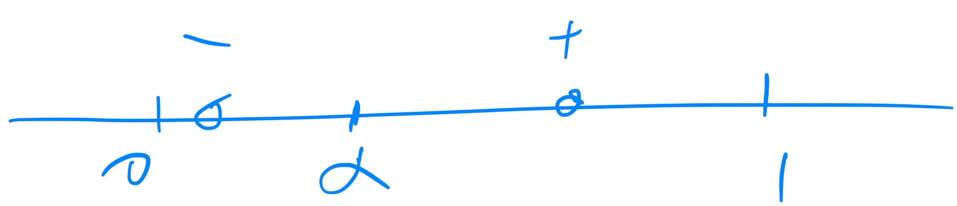
Valiant 1984.

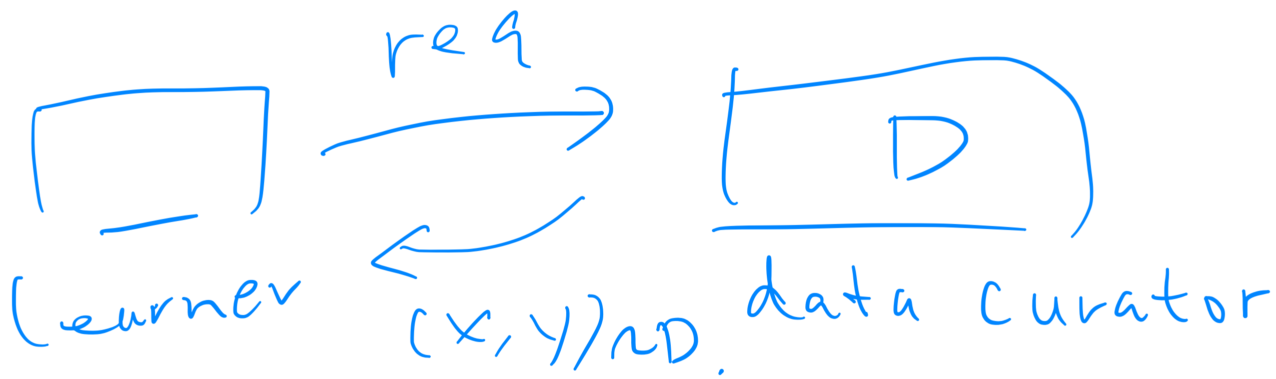
$X = [0, 1]$ $\quad D_X$ uniform

$Y = \{-1, +1\}$.

$$y = \begin{cases} +1, & x > \alpha. \\ -1, & x \leq \alpha. \end{cases}$$

$$|\hat{\alpha} - \alpha| \leq \epsilon$$





$S = \{(x_1, y_1), \ldots (x_n, y_n)\}$

$\epsilon = 10^{-6}$

$\hat{\alpha} = f(S)$

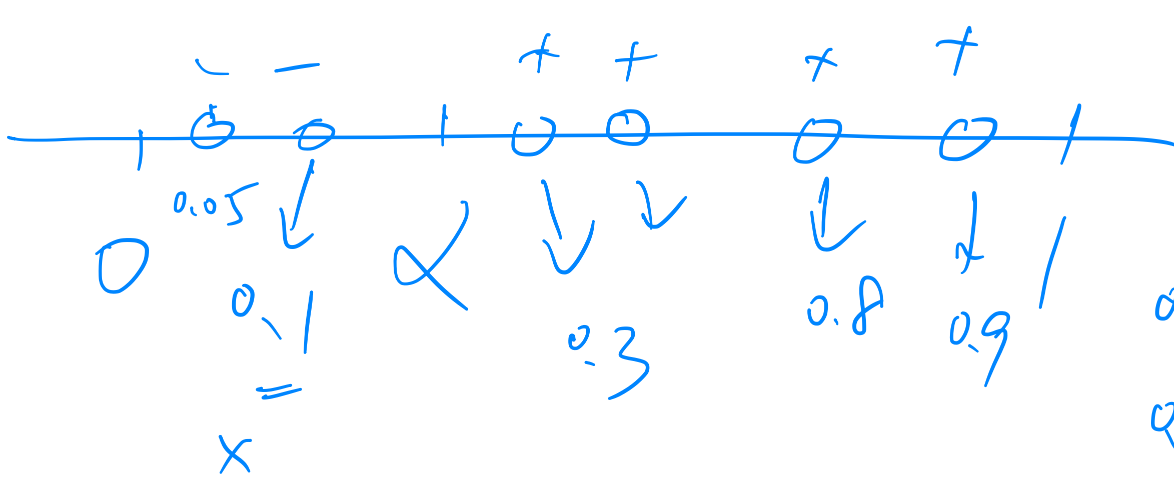
$|\hat{\alpha} - \alpha| \le \epsilon$

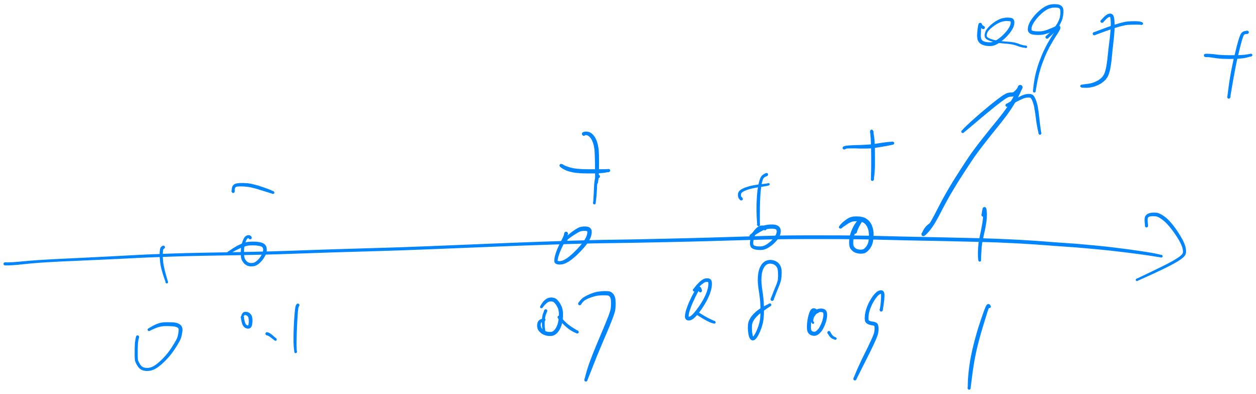
w.p. $1 - \delta$.

$\hat{\alpha} = 0.2$

$\alpha = 0.1$

$x = 0.9$ +
0.8 +
0.3 +
0.1 −
0.05 −
0.4 +





$$\alpha : [0, 0.9]$$

$$\Rightarrow [0.1, 0.9]$$

$$[0.1, 0.8]$$

$$\downarrow$$

$$[0.1, 0.7]$$

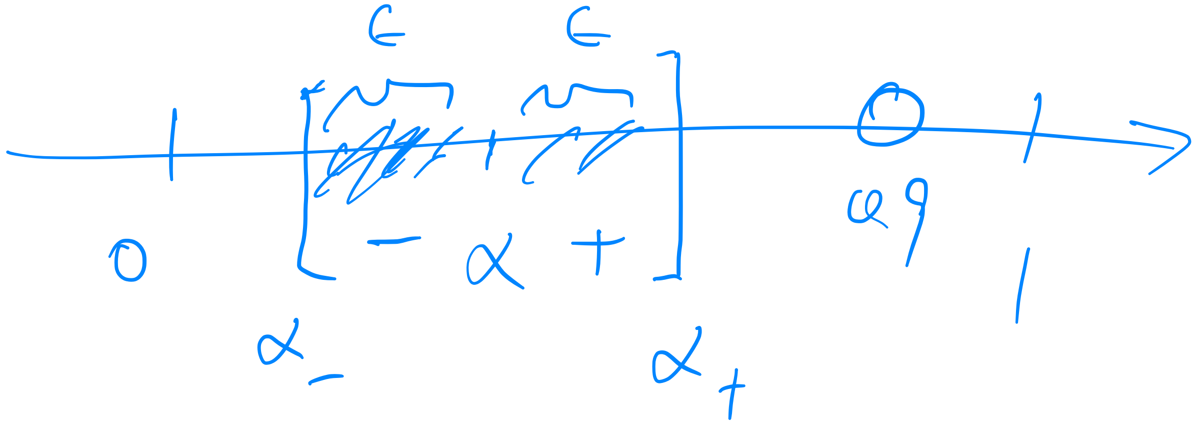
$\hat{\alpha} \rightarrow$ keep drawing pts.

$$\boxed{n = ?}$$

$\epsilon = 1/4$ $(x, y) \sim D$

w.p. $1/4$

$x \in [1/4, 1/2]$



can draw one pt $[\alpha_-, \alpha]$

w.p. ϵ

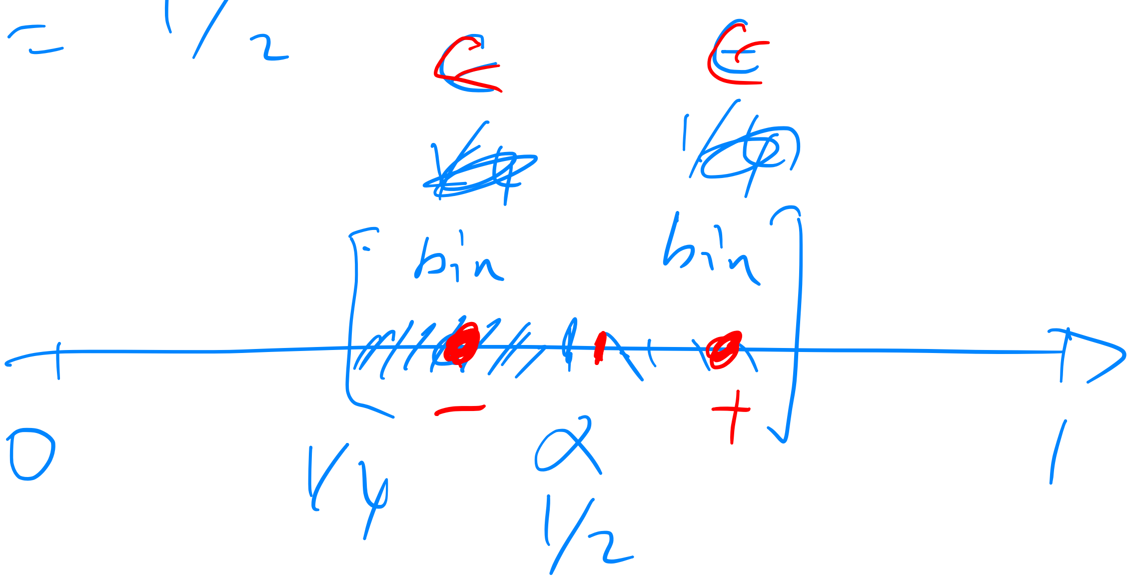
$x \in [\cdot - -]$

2nd pt $[\alpha, \alpha_+]$

$1/\epsilon$ pts

$\Theta(1/\epsilon)$ pts

$\alpha = 1/2$



(X_i, Y_i)

$$\boxed{Z_i = \begin{cases} 1, & \text{if } X_i \text{ 1st bin} \\ 0, & \text{otherwise.} \end{cases}}$$

$1, \ldots n.$

at least one. 1st bin.

$\Leftrightarrow$ $S = \sum_{i=1}^{n} Z_i$. find n.

$\underline{S \geq 1}$ wp. $1-\delta$.

Hoeffding's.

$$[a, b]$$

$$P\left(|S - E[S]| > t \right)$$

$$\leq 2 \cdot e^{-t^2 / n \cdot (b-a)^2}$$

$$a = 0 \quad b = 1, \quad E[S] = n \cdot \epsilon$$

$$P\left(|S - n\epsilon| > t \right) \qquad S \geq n\epsilon - t$$

$$\leq 2 \cdot e^{-t^2/n} \geq 1 - \delta$$

w.p. $\boxed{1 - 2 \cdot e^{-t^2/n}}$ $\quad |S - n\epsilon| \leq t$

$$S \geq \underbrace{n\epsilon - t}_{\geq 1} \quad \text{------} \quad 4$$

$$\begin{cases} n\epsilon - t \geq 1 & \text{------} \quad 1 \\ 2 \cdot e^{-t^2/n} \leq \delta & \text{------} \quad 2 \end{cases}$$

$$t = \frac{1}{2} n\epsilon$$

$$1 \implies \frac{n\epsilon}{2} \geq 1$$

$$\underline{n \geq \frac{2}{\epsilon}} \quad \text{--}$$

$$e^{t^2/n} \geq \frac{2}{\delta}$$

$$e^{\frac{1}{4} \cdot n\epsilon^2} \geq \frac{2}{\delta}$$

$$\frac{1}{4} \cdot n\epsilon^2 \geq \log\frac{2}{\delta}$$

$$n \geq \frac{4}{\epsilon^2} \cdot \log\frac{2}{\delta}$$

$$n \geq \frac{4}{\epsilon^2} \cdot \log \frac{2}{\delta} \cdot \boxed{\frac{8}{\epsilon^2} \cdot \log \frac{2}{\delta}}$$

if draw. $\underline{\frac{4}{\epsilon^2} \log \frac{2}{\delta}}$ pts

$\Rightarrow$ w.p. $1-\delta$. $\exists\ x_i$

$$\alpha - \epsilon \leq x_i \leq \alpha$$

---

draw $\underline{\frac{4}{\epsilon^2} \log \frac{2}{\delta}}$ pts.

w.p. $1-\delta$. $\exists\ x_j$

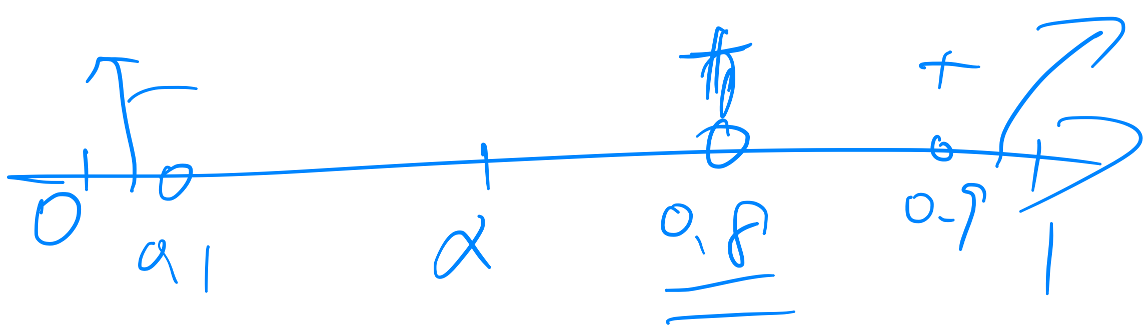
$$\alpha \leq x_j \leq \alpha + \epsilon$$

Chernoff's inequality.

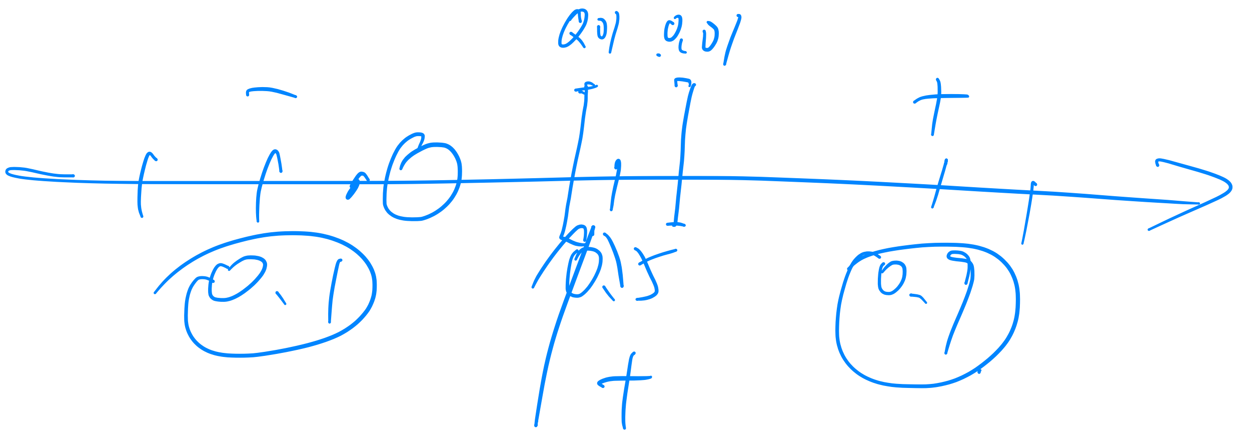
$$\frac{8}{\epsilon^2}\log\frac{2}{\delta} \to \frac{8}{\epsilon}\cdot\log\frac{2}{\delta}.$$

$$E[n] = \frac{1}{\epsilon}$$ # instances (x).

label-efficiency [0,1,0.8]



$$\#\text{labels} = \log n \simeq \log\frac{1}{\epsilon}$$



keep sending req → [ ]

until seeing $x_i \in [0.5-0.01, 0.5+0.01]$

(rejection sampling)

0.02

w.p. 0.02 1 pt

⟺ 50 pts.