$D.$ $\qquad E[D] = \mu$

$X_1 \, - - - \, X_n \sim D.$

$X \sim D$ positive

Markov's.

$$\Pr(X > t \cdot \mu) < \frac{1}{t}$$

$t = 1000$

w.p. $.99$

$$X \leq 100 \cdot \mu.$$

$$\Leftrightarrow \quad \mu \geq \frac{X}{\underline{100}}$$

$\alpha = \mathrm{Var}(D)$ is given

$$\Pr(|X - \mu| > t) \leq \frac{\mathrm{Var}(X)}{t^2}$$

$$X = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\mathrm{Var}[X] = \frac{1}{n} \cdot \mathrm{Var}[X] = \frac{\alpha}{n}.$$

$$\Pr(|X - \mu| > t) \leq \frac{\alpha / n}{t^2}$$

$$\downarrow$$

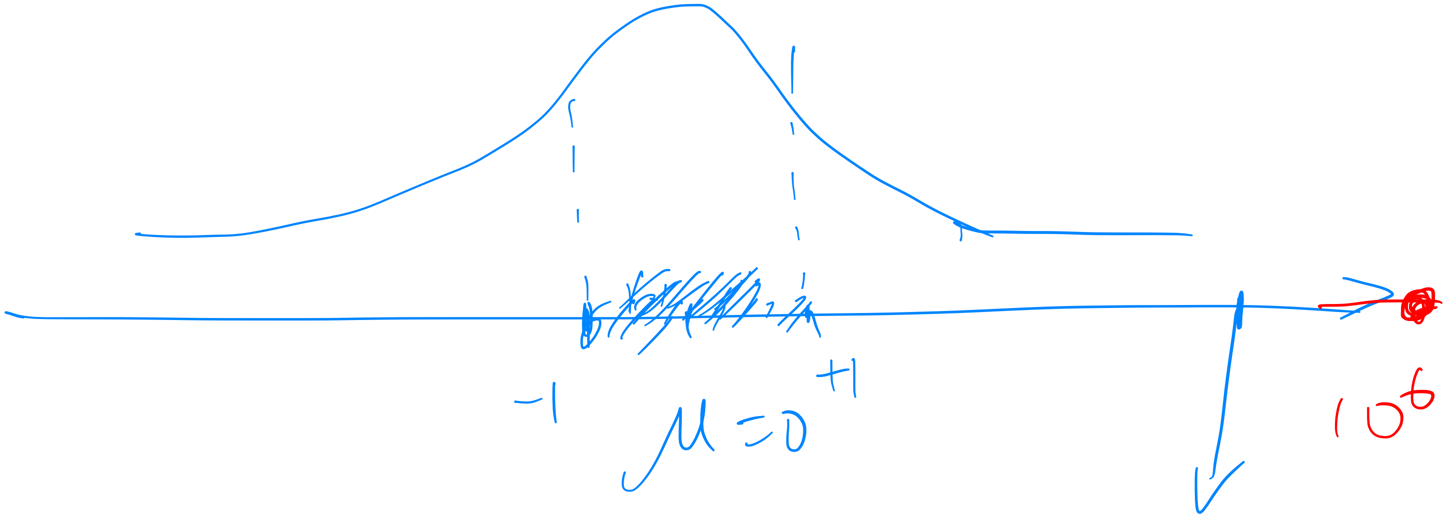
$t = \epsilon$ $\quad |X - \mu| > \epsilon \quad \frac{\alpha/n}{\epsilon^2}$

w.p. $1 - \frac{\alpha / n}{\epsilon^2}$. $|X - \mu| \leq \epsilon$

$$\boxed{n = \frac{100\alpha}{\epsilon^2}}$$

$= 0.99$

$$\frac{\alpha / n}{\epsilon^2} = 0.01$$

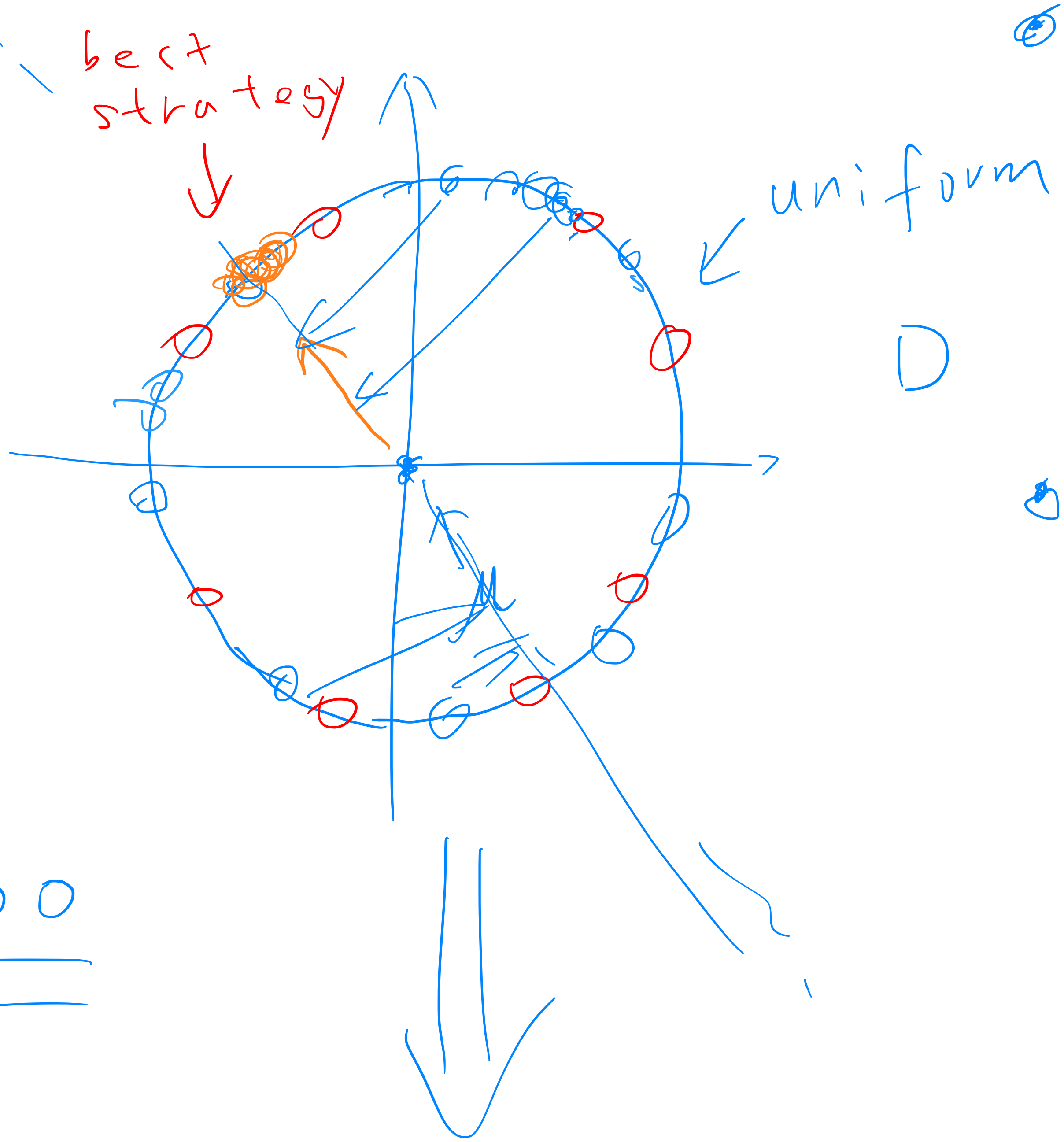


$$\frac{1}{n}\sum X_i$$

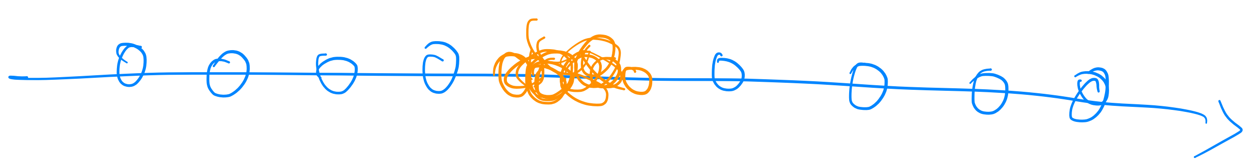
$$R^d$$

$$\sqrt{\epsilon / d}$$

$$\epsilon$$



1000



large variance.

a comp. w/ variance if all data are clean.

$$x_1 \quad \cdots \quad x_n$$

$$\downarrow \qquad\qquad \downarrow$$

$$q_1 \quad\qquad q_n$$

$$\boxed{0 \le q_i \le 1}$$

Goal: if x_i = clean,

$$q_i = 1$$

if x_i = corrupted,

$$q_i = 0$$

$$= \frac{1}{(1-\eta)n} \qquad \frac{\sum_{i=1}^{n} q_i x_i}{\sum x_i}$$

$$=$$

η = frac. of corrupted data.

(1) $0 \le q_i \le 1$

(2) $$\boxed{\frac{1}{(1-\eta)n} \cdot \sum q_i (x_i - \hat{\mu})(x_i - \hat{\mu})^T} \le \boxed{\quad} \Rightarrow I$$

cor. data have small weight.

given as far as we know clean data dist.

$$\hat{\mu} = \frac{1}{(1-\eta)n} \cdot \sum q_i x_i$$

LHS = cov. of D

$D = N(\mu, I)$.

$q_i = 1$. x_i corrupted.

(3) $\sum q_i \ge (1-\eta) n$.

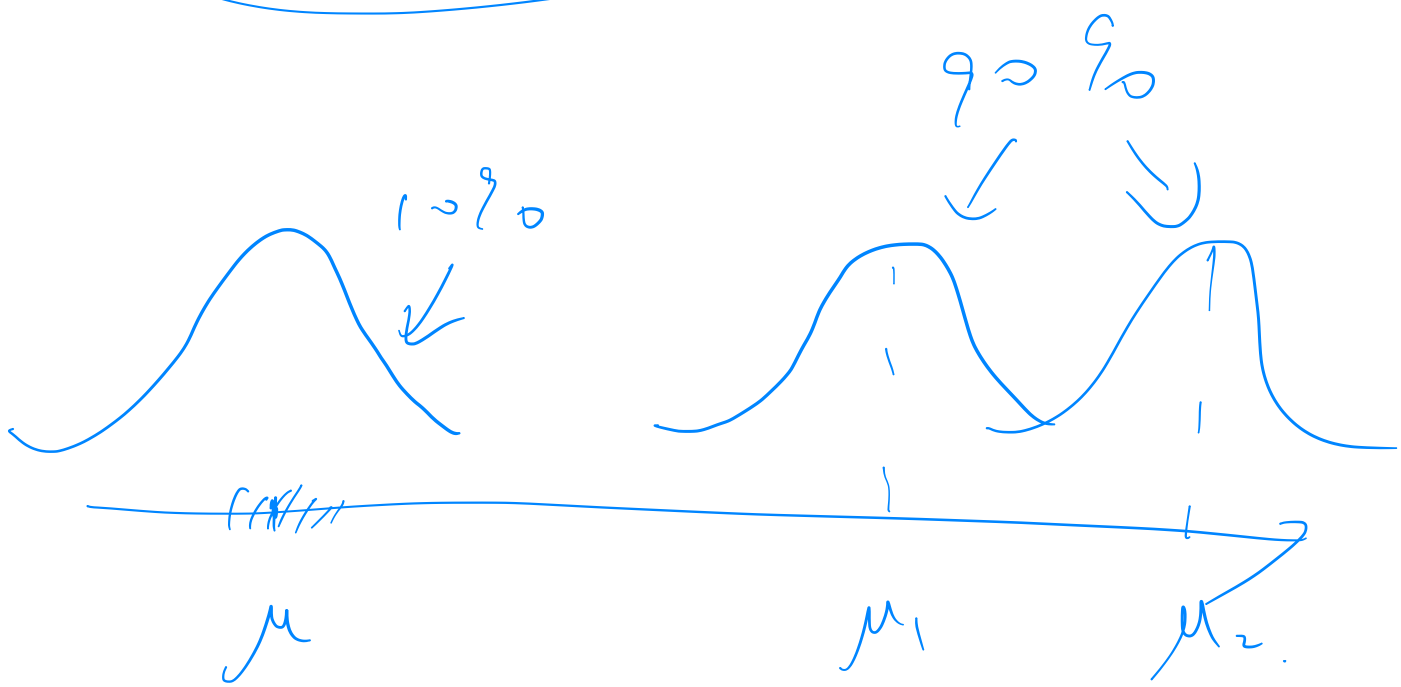
clean data have large weight

$\eta < 1/2$

99% false

– 4.9 (10 people)

– Warning: – –



$\rightarrow \{\hat{\mu}_1, \ldots, \hat{\mu}_k\}$

$$\min_{1 \le i \le k} \|\hat{\mu}_i - \mu\| \le \epsilon$$

true mean

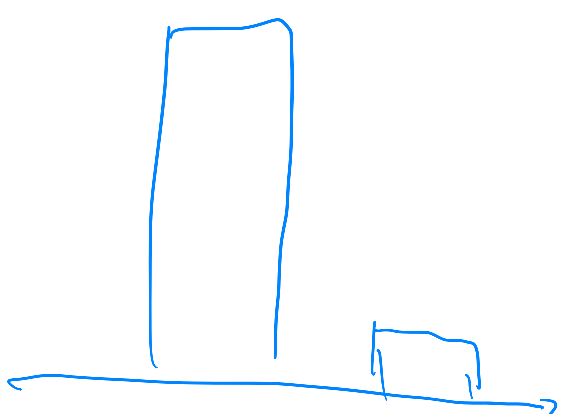
90% → 95%

90% → ph...

10% → Harisburg

10% → ≤10%

conf.



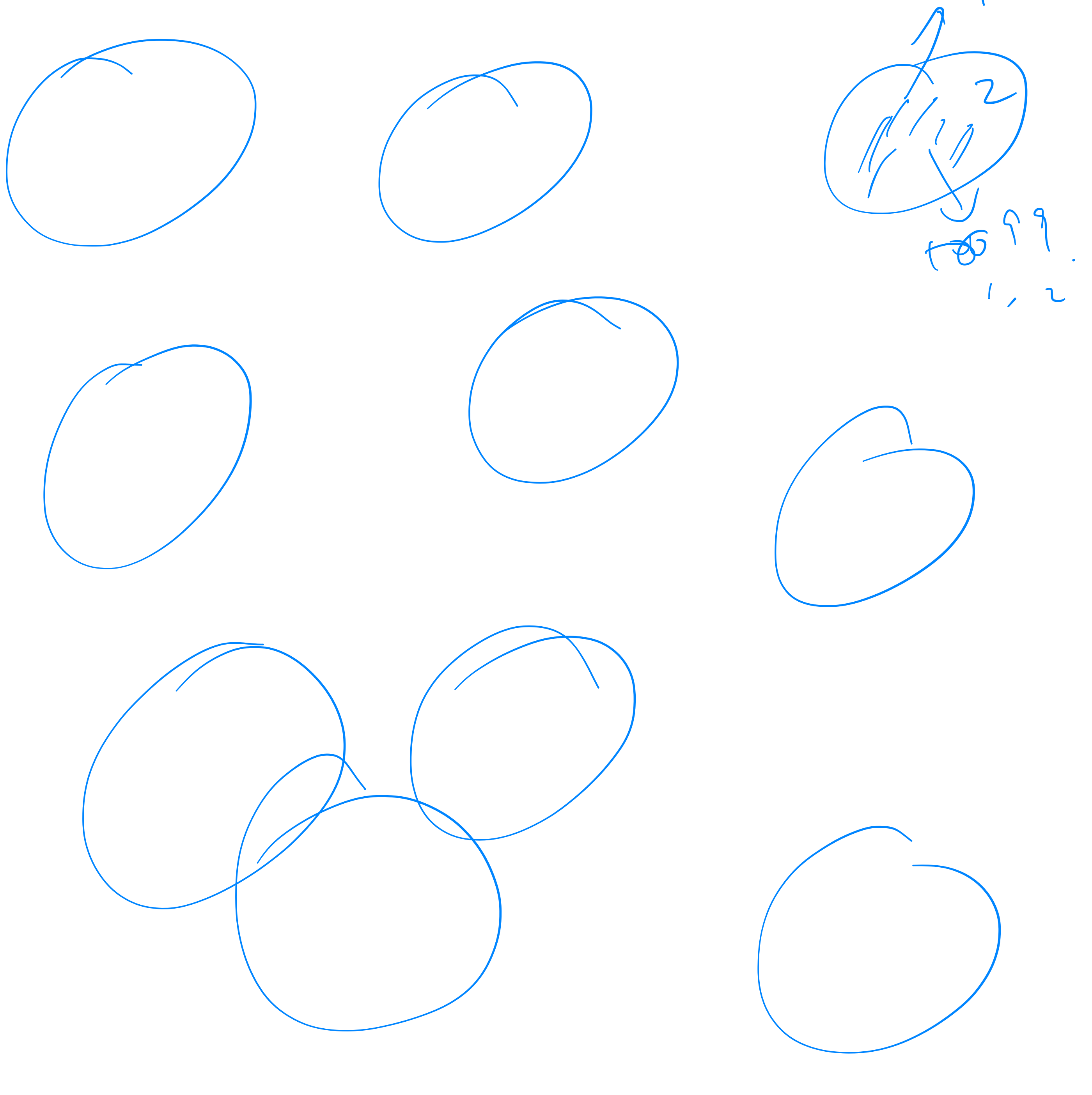
Con

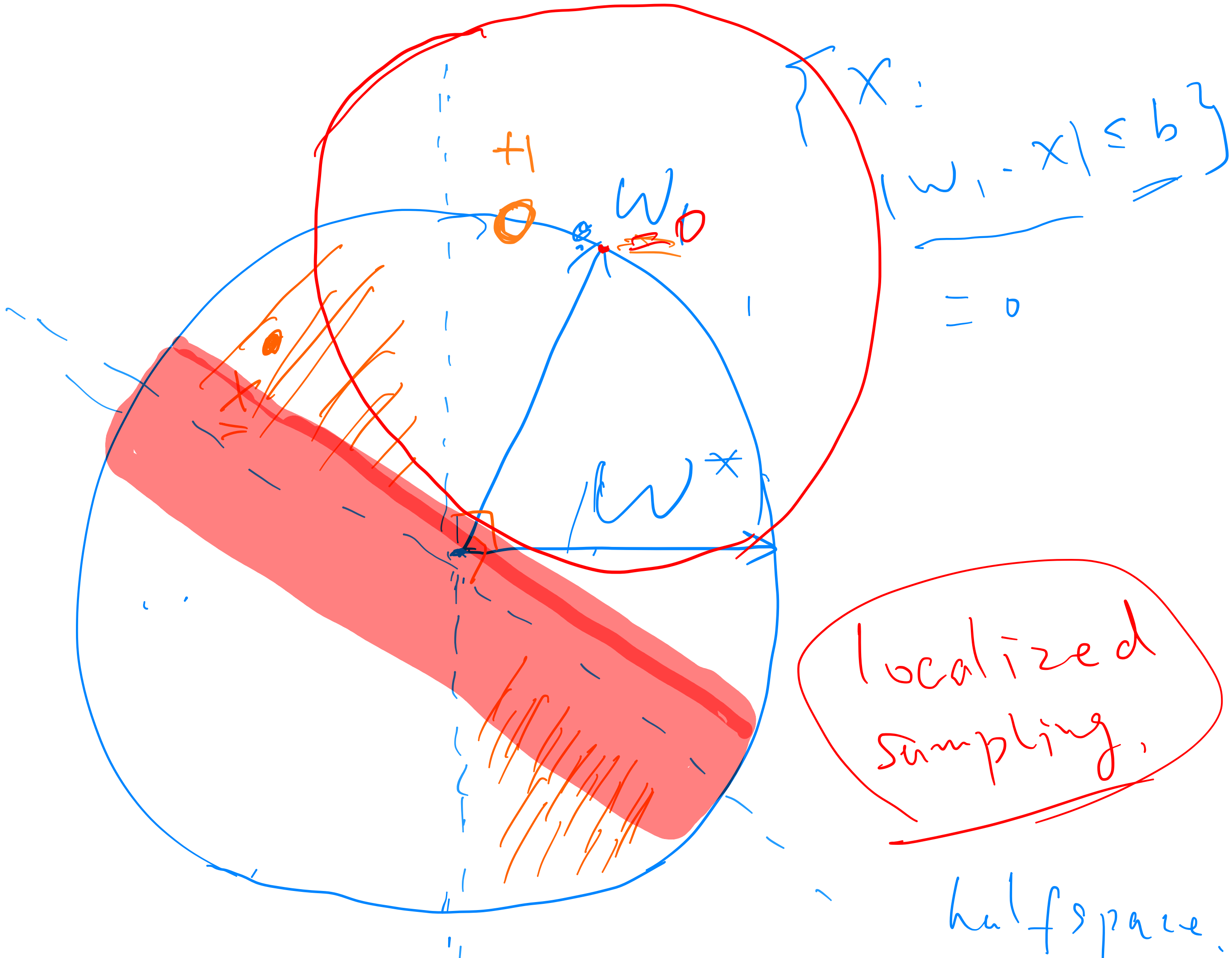
n X_{train} m X_{test}

M: $n < m$.

Active Learning

1000 pts.





eg x, $y = \text{sign}(W^* \cdot x)$

$$W^* \cdot x > 0$$

$$y = 1$$

$$w^* \cdot x < 0$$

$$y = -1$$

$\underline{\underline{w_0}} = (1, 0, 0, \ldots 0)\}$

$k = 1, \ldots T$.

$X_k = \{x : |w_{k-1} \cdot x| \le b_k\}$

. draw n_k pts from X_k.

label them

" strategy for labeling.

→ $\underline{\text{SVM}}$.

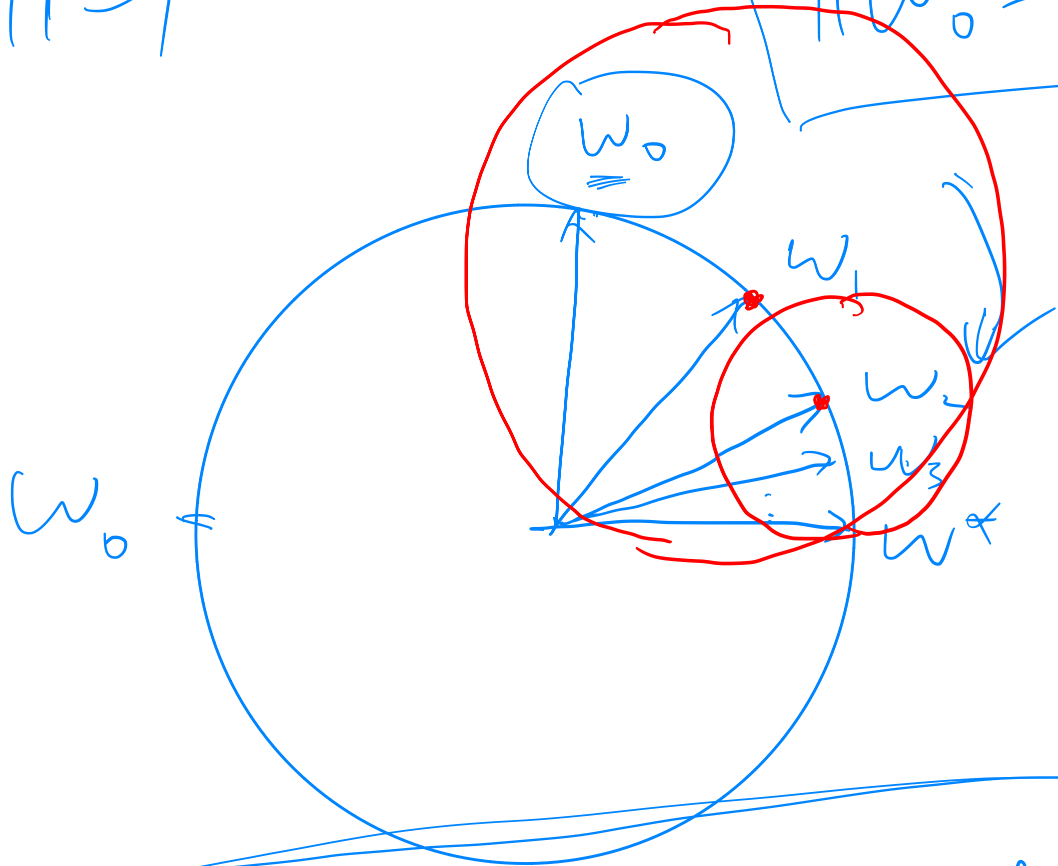
$\min_w \sum_{t=1}^{n_k} \max\{0, 1 - y_t \cdot w \cdot x_t\}$

$w_1^* \Leftarrow$ s.t. $\cancel{\|w\|_2 \le [\;]}$

$\|w - w_{k-1}\|_2 \le r_k$

$\|w\| = 1$ $\|w_0 - w^*\| \le 2$

$\Downarrow$



$\boxed{\#\text{labels/iter} = d.}$

init angle $= O(1)$.

$\#\text{iter} = \log\frac{1}{\epsilon}$

$\boxed{d \cdot \log\frac{1}{\epsilon}}$ $\leftarrow$

① $\Omega(d \cdot \log\frac{1}{\epsilon})$. ② $d \cdot \frac{1}{\epsilon}$

$$\|w_0 - w^*\| \le 2$$

$$\Downarrow$$

$$V_0 = \{ w : \|w - w_0\| \le 2 \}$$

$$w^* \in V_0$$

V_0 : trust region for w^*

$$\Downarrow$$

$$\|w_1 - w^*\| \le 1$$

$$\Downarrow$$

$$V_1 = \{ w : \|w - w_1\| \le 1 \}$$

$$\Downarrow$$

$$\vdots$$

Margin-based Active Learning

2007

condition on D.

2014 → y

2020

(x, y)

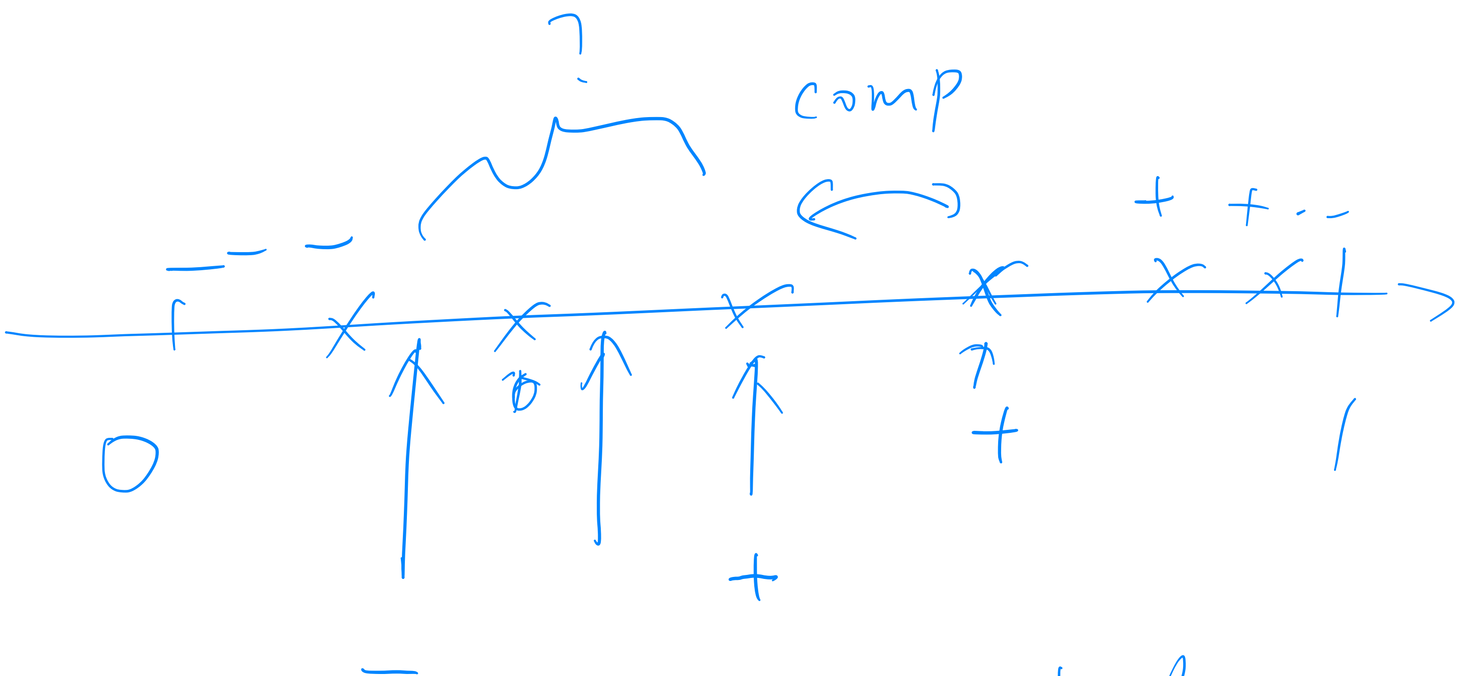
$d \cdot \log \frac{1}{\epsilon}$

$s \cdot \log d \cdot \log \frac{1}{\epsilon}$.

d/ϵ.

$\#\text{label} = d \cdot \log \frac{1}{\epsilon}$

$\#\text{comp} = d \cdot \frac{1}{\epsilon}$



log n   label.

n   comp