Geometric Sampling

Probability (and techniques)	Algorithms
Additive Chernoff Inequality	ε-samples and ε-nets
Double-sampling	
VC dimension	

Concentration phenomena

let $X_1, \cdots, X_n \in [0,1]$ be independent

$$\mu = \mathbb{E}[X_1 + \cdots + X_n]$$

Q: probability that

$X_1 + \cdots + X_n$ is close to μ?
(multiplicatively)

A: error probability exponentially small in μ

Multiplicative Chernoff bound

let $X_1, \dots, X_n \in [0,1]$ be independent

$\mu = E[X_1 + \dots + X_n]$

$\varepsilon \in (0,1)$

- $P[X_1 + \dots + X_n \geq (1+\varepsilon)\mu] \leq e^{-\varepsilon^2 \mu / 3}$

- $P[X_1 + \dots + X_n \leq (1-\varepsilon)\mu] \leq e^{-\varepsilon^2 \mu / 2}$

Additive Chernoff bound

let $X_1, \cdots, X_n \in [0,1]$ be independent

$\mu = E[X_1 + \cdots + X_n]/n$

$\varepsilon \in (0,1)$

- $P\left[\frac{X_1 + \cdots + X_n}{n} \geq \mu \boxed{+\varepsilon}\right] \leq e^{-2\varepsilon^2 n}$

- $P\left[\frac{X_1 + \cdots + X_n}{n} \leq \mu \boxed{-\varepsilon}\right] \leq e^{-2\varepsilon^2 n}$

(proof at end if time permits)

$X_1, \ldots, X_n \in [0,1]$ independent, $\varepsilon \in (0,1)$

Multiplicative Chernoff

$$\mu = E[X_1 + \cdots + X_n]$$

$$P[X_1 + \cdots + X_n \geq (1+\varepsilon)\mu] \leq e^{-\varepsilon^2 \mu / 3}$$

$$P[X_1 + \cdots + X_n \leq (1-\varepsilon)\mu] \leq e^{-\varepsilon^2 \mu / 2}$$

Additive Chernoff

$$\mu = E[X_1 + \cdots + X_n]/n$$

$$P\left[\frac{X_1 + \cdots + X_n}{n} \geq \mu + \varepsilon\right] \leq e^{-2\varepsilon^2 n}$$

$$P\left[\frac{X_1 + \cdots + X_n}{n} \leq \mu - \varepsilon\right] \leq e^{-2\varepsilon^2 n}$$

Lecture 12

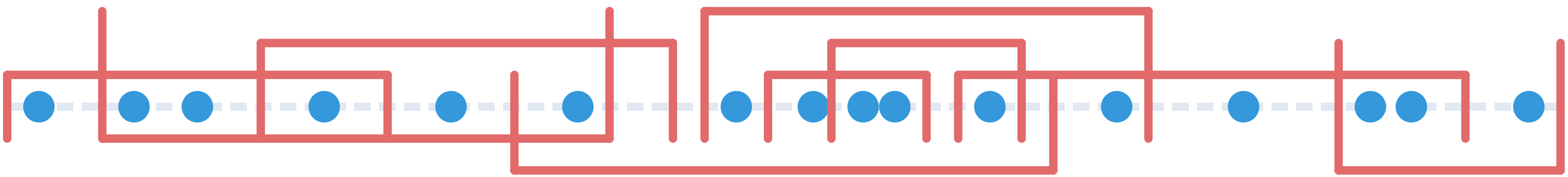
Probability (and techniques)	Algorithms
✓ Additive Chernoff Inequality	ε-samples and ε-nets
Double-sampling	
VC dimension	

Range Space

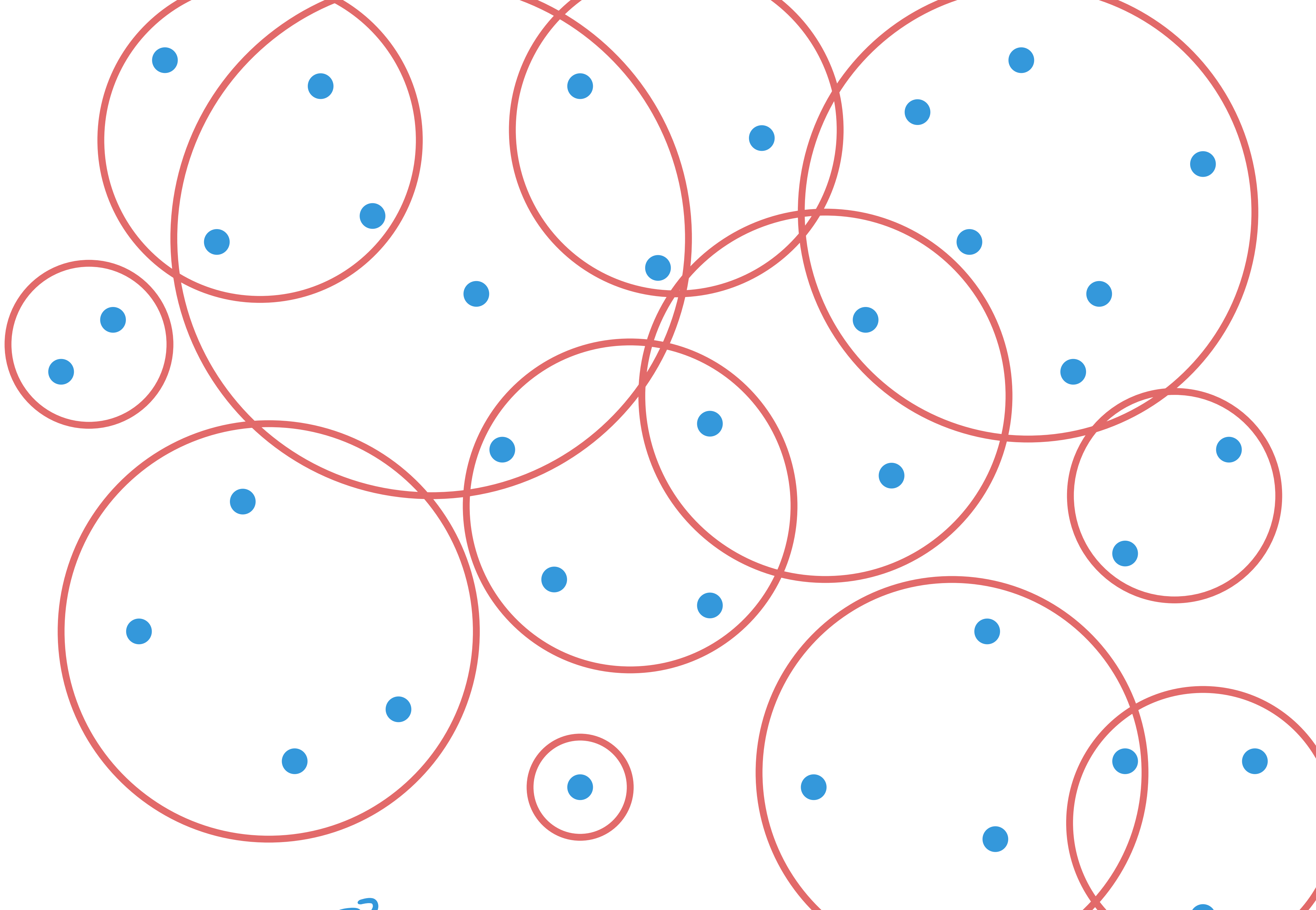
$$(P, R)$$

points

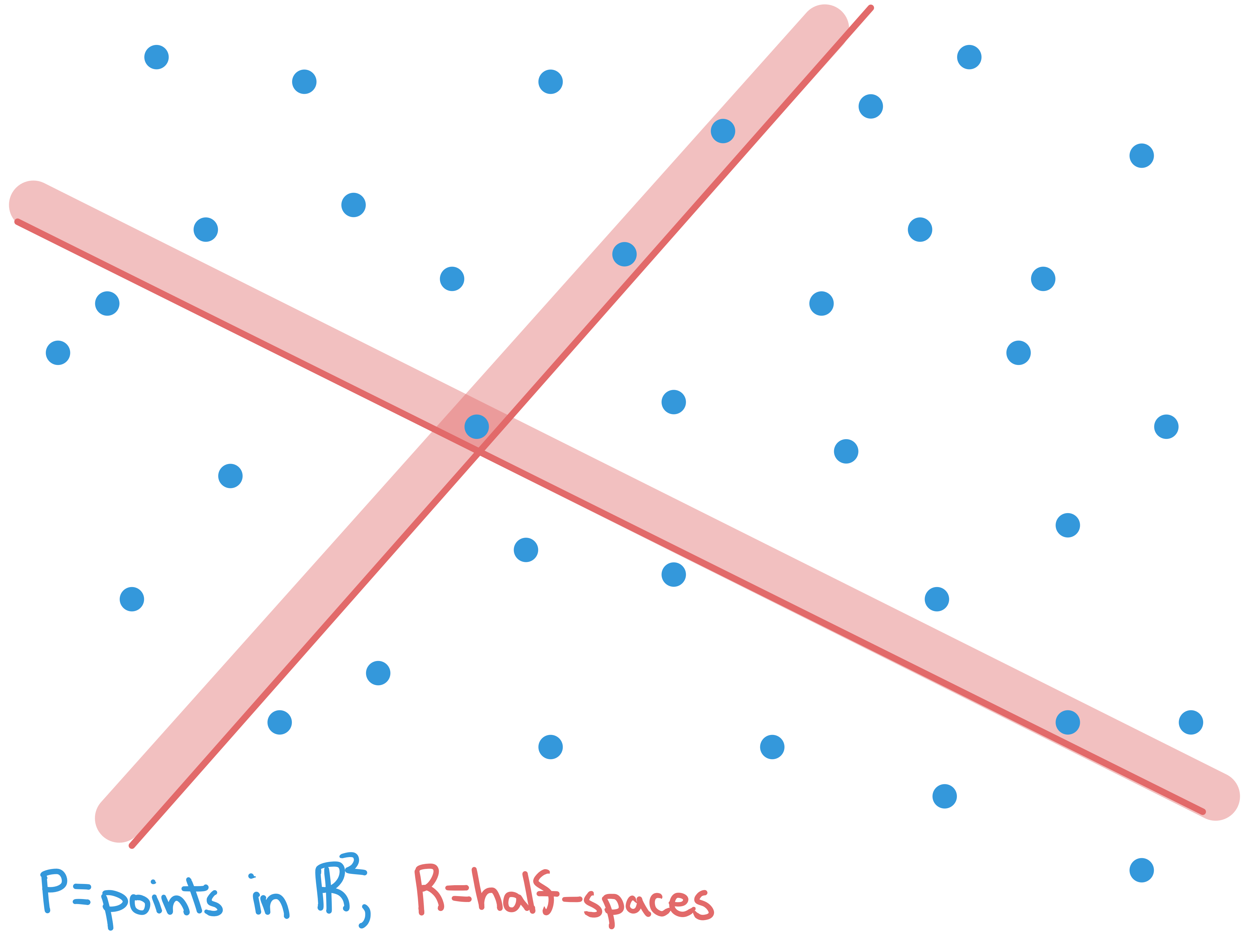
ranges
(subsets of P)



P = points in $\mathbb{R}$,   R = intervals



P = points in $\mathbb{R}^2$, R = all disks


P = points in $\mathbb{R}^2$, R = half-spaces

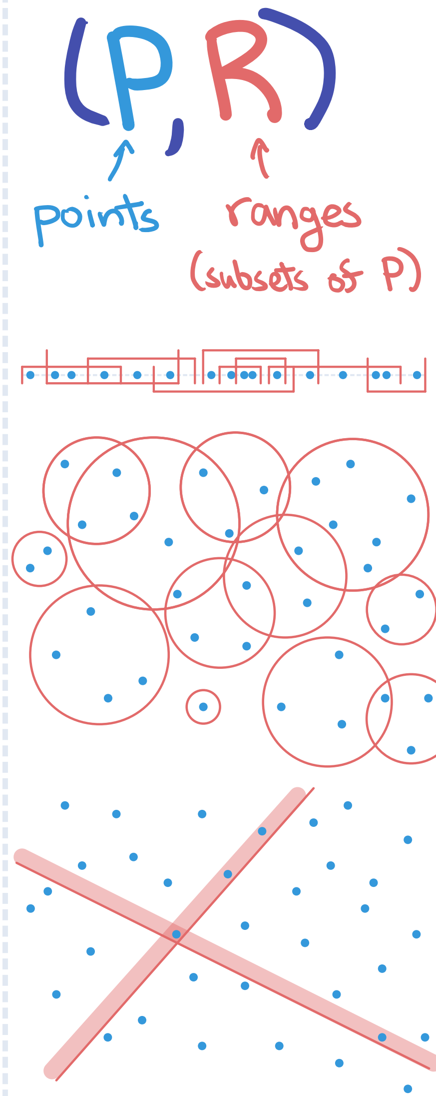
Typical queries: for $r \in R$:

Q: is r empty?

Q: how many points does r contain?

Q: does r contain $\geq \varepsilon$-fraction of P?

(for $\varepsilon \in (0,1)$)



measure $\mu(r) = \dfrac{|P \cap r|}{|P|}$

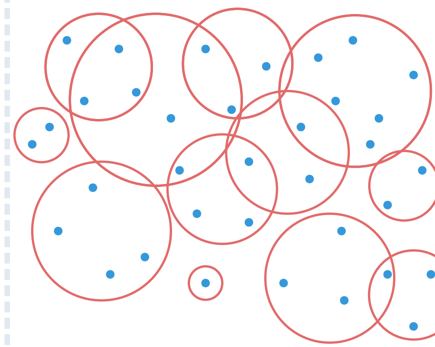
for $Q \subseteq P$, let $\mu_Q(r) = \dfrac{|Q \cap r|}{|Q|}$

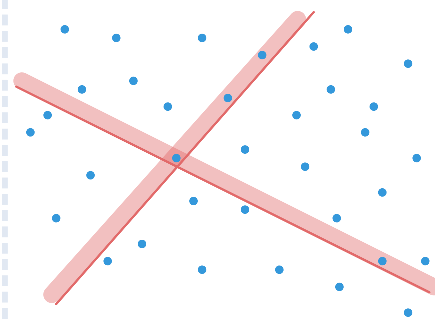
(high-level)
Question: how big should a random sample $Q \subseteq P$ be so that

$$\mu_Q \simeq \mu \ ?$$

(P, R)

points ranges (subsets of P)





Typical queries: for $r \in R$:

Q: is r empty?

Q: how many points does r contain?

Q: does r contain $\geq$ ε-fraction of P?

(for $\varepsilon \in (0,1)$)

ε-sample

Q is an ε-sample if

$$|\mu_Q(r) - \mu(r)| \le \varepsilon \quad \forall\, r \in R$$

ε-net

Q is an ε-net if

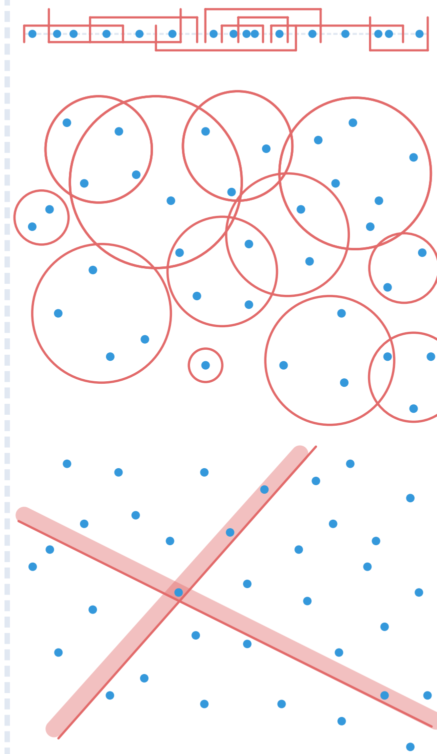
$$\mu_Q(r) > 0 \text{ whenever } \mu(r) \ge \varepsilon$$

(ε-sample $\Rightarrow$ ε-net)

(P, R)

P: points

R: ranges (subsets of P)



$$\mu(r) = \frac{|P \cap r|}{|P|}$$

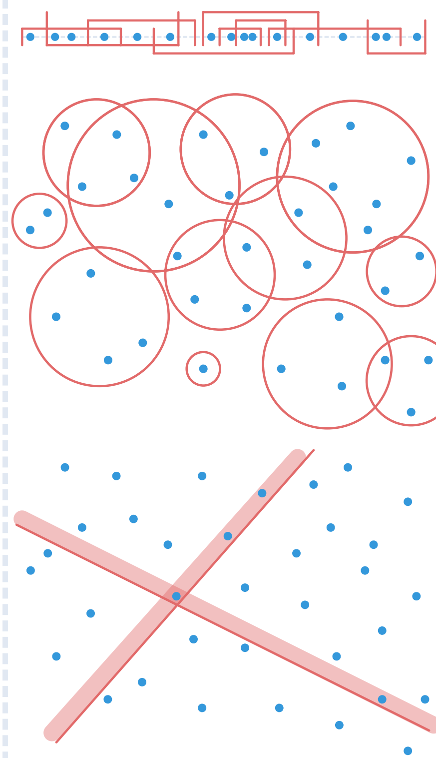
$$\mu_Q(r) = \frac{|Q \cap r|}{|Q|}$$

ε-sample

$$|\mu_Q(r) - \mu(r)| \le \varepsilon \quad \forall\, r \in R$$

(P, R)

points ranges (subsets of P)



$\mu(r) = \dfrac{|P \cap r|}{|P|}$

$\mu_Q(r) = \dfrac{|Q \cap r|}{|Q|}$

(!) $|Q| < (1-\varepsilon)|P|$ is impossible w/out any assumptions on R

e.g. $R = 2^P$ (= all subsets of P)

if $|Q| < (1-\varepsilon)|P|$, then $\bar{Q} = P \setminus Q$ has

$$\mu(\bar{Q}) > \varepsilon, \quad \mu_Q(\bar{Q}) = 0$$

(same for ε-nets)

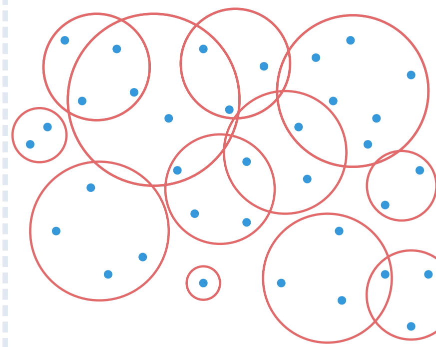
(!) $|Q| < (1-\varepsilon)|P|$ is impossible w/out any assumptions on R

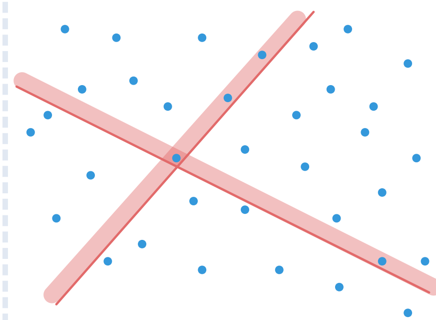
e.g. $R = 2^P$ (= all subsets of P)

(P, R)

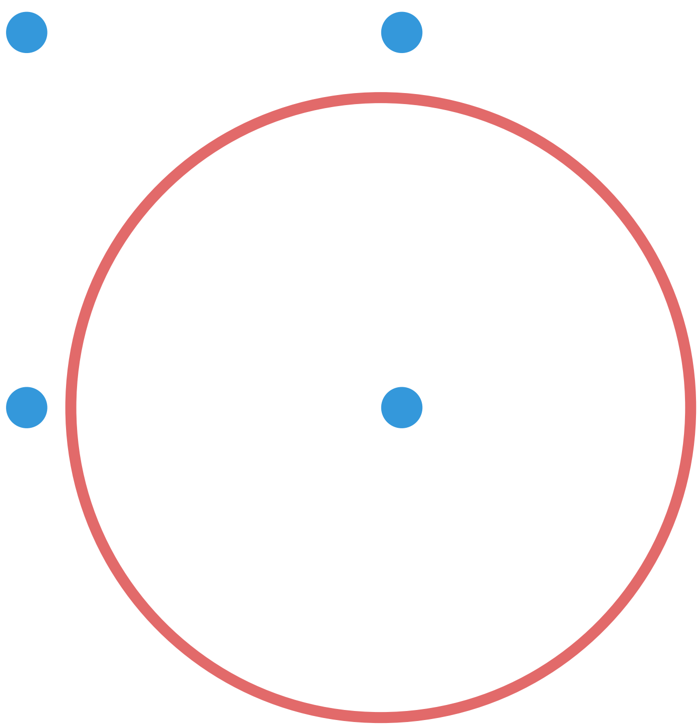
points

ranges (subsets of P)





many natural range spaces cannot induce (nearly) all subsets of P



e.g. cannot cover (only) opposite corners of a square w/ a disk

$\mu(r) = \frac{|P \cap r|}{|P|}$

$\mu_Q(r) = \frac{|Q \cap r|}{|Q|}$

(!) $|Q| < (1-\varepsilon)|P|$ is impossible w/out any assumptions on R

e.g. $R = 2^P$ (= all subsets of P)

many natural range spaces cannot induce (nearly) all subsets of P

let $R \wedge Q = \{ r \cap Q : r \in R \}$ for $Q \subseteq P$

growth function $g : \mathbb{N} \to \mathbb{N}$

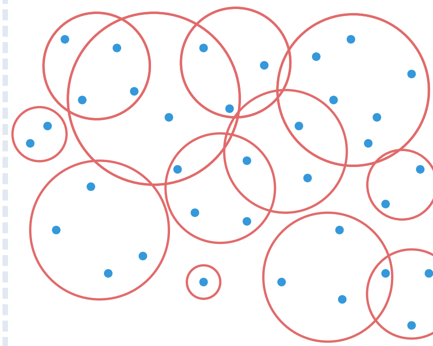
$$g(n) = \max \{ |R \wedge Q| : Q \subseteq P, |Q| \leq n \}$$

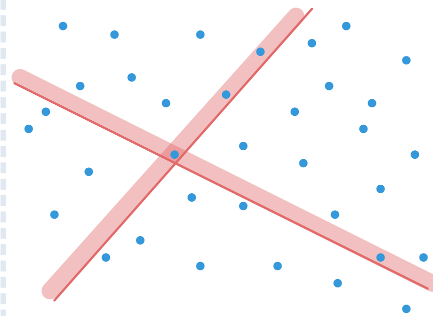
e.g. disks have $g(n) \leq O(n^3)$ (which we prove later)

(P, R)

points

ranges (subsets of P)





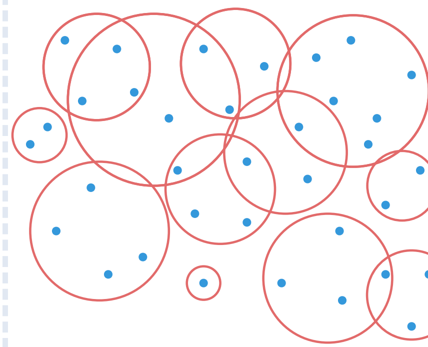
$$\mu(r) = \frac{|P \cap r|}{|P|}$$

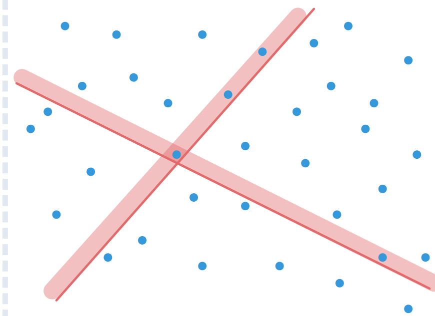
$$\mu_Q(r) = \frac{|Q \cap r|}{|Q|}$$

ε-sample $\quad |\mu_Q(r) - \mu(r)| \le \varepsilon \quad \forall\, r \in R$

(P, R)

points $\quad$ ranges (subsets of P)





let Q sample k points w/ repetition

for range $r \in R$:

$$\mu_Q(r) = \frac{1}{k}(X_1 + \cdots + X_k) \text{ w/ } X_i = \begin{cases} 1 & \text{if } i\text{th point in } r \\ 0 & \text{otherwise} \end{cases}$$

$$P[|\mu_Q(r) - \mu(r)| \ge \varepsilon]$$

$$= P[|\tfrac{1}{k}(X_1 + \cdots + X_k) - \mu(r)| \ge \varepsilon]$$

$$\le 2e^{-2\varepsilon^2 k}$$

$$\mu(r) = \frac{|P \cap r|}{|P|}$$

$$\mu_Q(r) = \frac{|Q \cap r|}{|Q|}$$

$$R \wedge Q = \{r \cap Q : r \in R\}$$

$$P\left[\frac{X_1 + \cdots + X_n}{n} \ge \mu + \varepsilon\right],\ P\left[\frac{X_1 + \cdots + X_n}{n} \le \mu - \varepsilon\right] \le e^{-2\varepsilon^2 n}$$

$$g(n) = \max\{|R \wedge Q| : Q \subseteq P,\ |Q| \le n\}$$

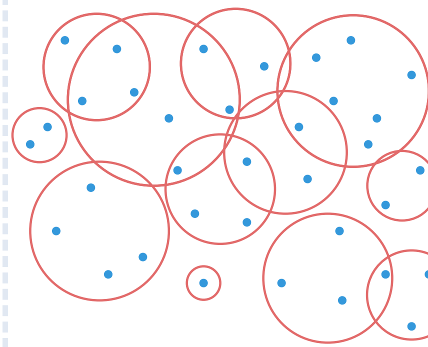
ε-sample $\quad |\mu_Q(r) - \mu(r)| \le \varepsilon \quad \forall\, r \in R$

(P, R)

points $\quad$ ranges (subsets of P)

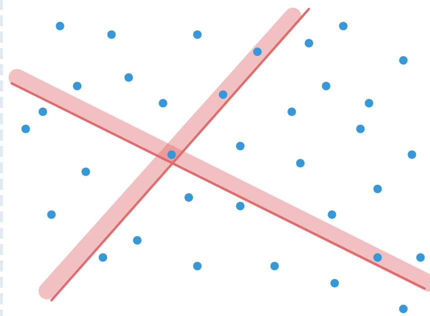
let Q sample k points w/ repetition

✓ $P[\,|\mu_Q(r) - \mu(r)| \ge \varepsilon\,] \le 2e^{-2\varepsilon^2 k}$ for $r \in R$



how many r's?

$|R \wedge P| \le g(|P|)$



$\mu(r) = \dfrac{|P \cap r|}{|P|}$

$\mu_Q(r) = \dfrac{|Q \cap r|}{|Q|}$

$R \wedge Q = \{ r \cap Q : r \in R \}$

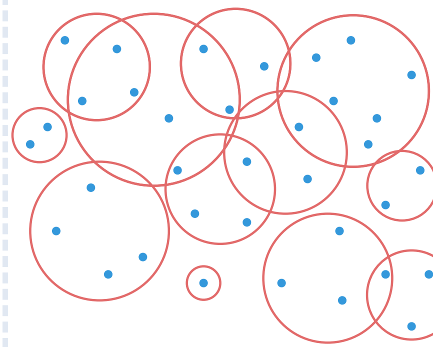
$P\left[\dfrac{X_1 + \cdots + X_n}{n} \ge \mu + \varepsilon\right], \; P\left[\dfrac{X_1 + \cdots + X_n}{n} \le \mu - \varepsilon\right] \le e^{-2\varepsilon^2 n}$

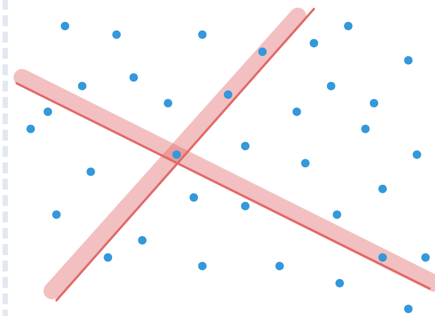
$g(n) = \max\{ |R \wedge Q| : Q \subseteq P,\ |Q| \le n \}$

ε-sample $\quad |\mu_Q(r) - \mu(r)| \le \varepsilon \quad \forall\, r \in R$

(P, R)

points $\quad$ ranges (subsets of P)





let Q sample k points w/ repetition

✓ $P[|\mu_Q(r) - \mu(r)| \ge \varepsilon] \le 2e^{-2\varepsilon^2 k}$ for $r \in R$

✓ $|R \wedge P| \le g(|P|)$

$k = O(\ln(g(|P|))/\varepsilon^2) \Rightarrow$

$\Pr[Q$ not ε-sample$]$

$$\underset{\text{(union bound)}}{\le} \sum_{r \in R \wedge P} P[|\mu_Q(r) - \mu(r)| \ge \varepsilon]$$

$$\le g(|P|)\, e^{-\varepsilon^2 k} \le \frac{1}{2}$$

$\mu(r) = \dfrac{|P \cap r|}{|P|}$

$\mu_Q(r) = \dfrac{|Q \cap r|}{|Q|}$

$R \wedge Q = \{ r \cap Q : r \in R \}$

$P\left[\dfrac{X_1 + \cdots + X_n}{n} \ge \mu + \varepsilon\right], \; P\left[\dfrac{X_1 + \cdots + X_n}{n} \le \mu - \varepsilon\right] \le e^{-2\varepsilon^2 n}$

$g(n) = \max\{ |R \wedge Q| : Q \subseteq P, |Q| \le n \}$

ε-sample $\quad |\mu_Q(r)-\mu(r)| \le \varepsilon \quad \forall r \in R$

(P, R)

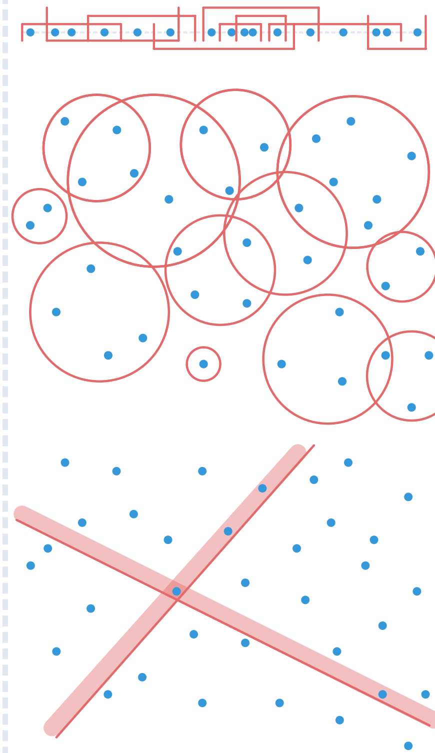
points $\quad$ ranges (subsets of P)

let Q sample k points w/ repetition

$$k = O\left(\frac{\ln(g(|P|))}{\varepsilon^2}\right) \Rightarrow \varepsilon\text{-sample}$$

$\left(= O\left(\frac{\ln(|P|)}{\varepsilon^2}\right) \text{ for } g(n) = n^{O(1)}\right)$

(w/ constant probability)



Also (exercise)

$$k = O\left(\frac{\ln(g(|P|))}{\varepsilon}\right) \Rightarrow \varepsilon\text{-net}$$

(w/ constant probability)

Better?

$$\mu(r) = \frac{|P \cap r|}{|P|}$$

$$\mu_Q(r) = \frac{|Q \cap r|}{|Q|}$$

$$R \wedge Q = \{ r \cap Q : r \in R \}$$

$$P\left[\frac{X_1 + \cdots + X_n}{n} \ge \mu + \varepsilon\right],\ P\left[\frac{X_1 + \cdots + X_n}{n} \le \mu - \varepsilon\right] \le e^{-2\varepsilon^2 n}$$

$$g(n) = \max \{ |R \wedge Q| : Q \subseteq P,\ |Q| \le n \}$$

ε-sample $\quad |\mu_Q(r) - \mu(r)| \le \varepsilon \quad \forall\, r \in R$

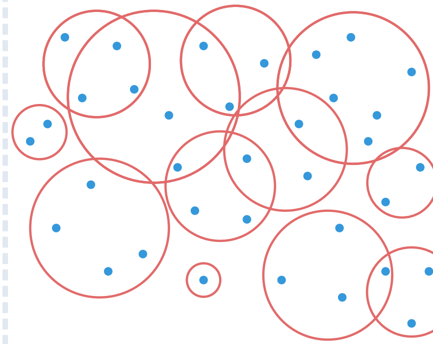
(P, R)

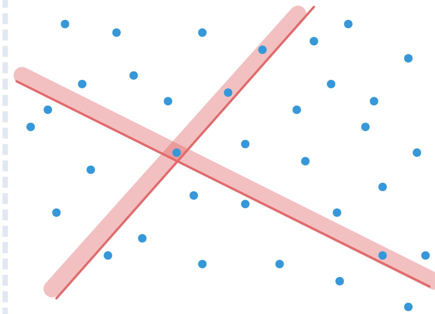
points $\quad$ ranges

(subsets of P)

let Q sample k points w/ repetition

$k = O\left(\frac{\ln(g(|P|))}{\varepsilon^2}\right) \Rightarrow$ ε-sample $\quad k = O\left(\frac{\ln(g(|P|))}{\varepsilon}\right) \Rightarrow$ ε-net (w/ constant probability)





ε-sample theorem [VC71]

$$k \ge O\left(\frac{\ln(g(2k))}{\varepsilon^2}\right) \Rightarrow \varepsilon\text{-sample}$$

(w/ constant probability)

e.g. $g(n) = O(n^d)$

$$k \ge O\left(\frac{d \ln(d/\varepsilon)}{\varepsilon^2}\right) \Rightarrow \varepsilon\text{-sample}$$

independent of $|P|$ (!)

$$\mu(r) = \frac{|P \cap r|}{|P|}$$

$$\mu_Q(r) = \frac{|Q \cap r|}{|Q|}$$

$$R \wedge Q = \{ r \cap Q : r \in R \}$$

$$P\left[\frac{X_1 + \cdots + X_n}{n} \ge \mu + \varepsilon\right],\ P\left[\frac{X_1 + \cdots + X_n}{n} \le \mu - \varepsilon\right] \le e^{-2\varepsilon^2 n}$$

$$g(n) = \max\{ |R \wedge Q| : Q \subseteq P,\ |Q| \le n \}$$

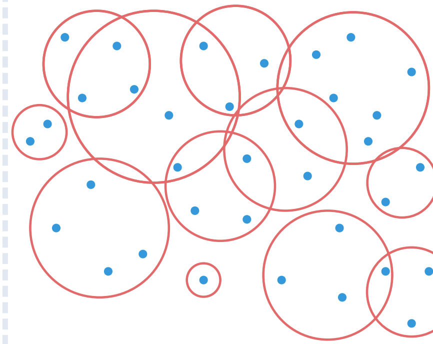
ε-sample $\quad |\mu_Q(r)-\mu(r)| \le \varepsilon \quad \forall r \in R$

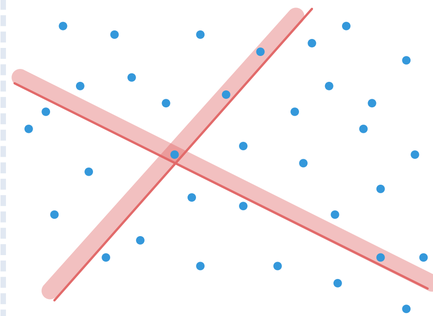
(P, R)

points $\quad$ ranges (subsets of P)

let Q sample k points w/ repetition

$k=O\left(\frac{\ln(g(|P|))}{\varepsilon^2}\right) \Rightarrow$ ε-sample $\quad k=O\left(\frac{\ln(g(|P|))}{\varepsilon}\right) \Rightarrow$ ε-net (w/ constant probability)





ε-net theorem [HW87]

$$k \ge O\left(\frac{\ln(g(2k))}{\varepsilon}\right) \Rightarrow \text{ε-net}$$

(w/ constant probability)

e.g. $g(n)=O(n^d)$

$$k \ge O\left(\frac{d\ln(1/\varepsilon)}{\varepsilon}\right) \Rightarrow \text{ε-net}$$

independent of $|P|$ (!)

$\mu(r)=\frac{|P\cap r|}{|P|}$

$\mu_Q(r)=\frac{|Q\cap r|}{|Q|}$

$R \wedge Q = \{r\cap Q : r\in R\}$

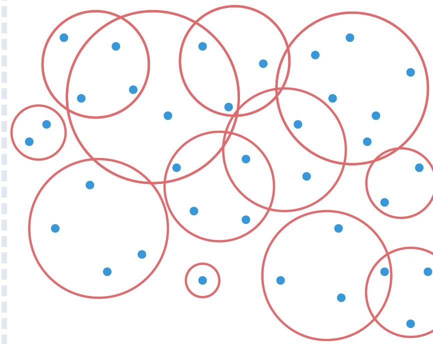
$P\left[\frac{X_1+\cdots+X_n}{n}\ge \mu+\varepsilon\right], P\left[\frac{X_1+\cdots+X_n}{n}\le \mu-\varepsilon\right] \le e^{-2\varepsilon^2 n}$

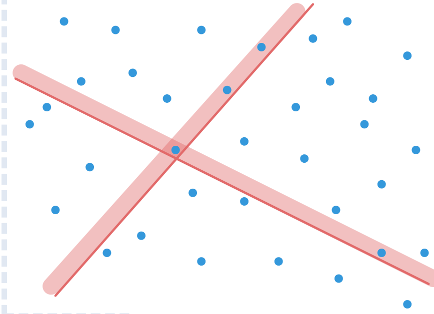
$g(n)=\max\{|R\wedge Q| : Q\subseteq P, |Q|\le n\}$

ε-sample theorem [VC71]

$$k \geq O\left(\frac{\ln(g(2k))}{\varepsilon^2}\right) \Rightarrow \varepsilon\text{-sample}$$

(w/ constant probability)





Proof let Q_1 sample k points (w/ repetition)

define event

$A = \binom{\text{event}}{\text{that}}\ |\mu_1(r) - \mu(r)| > \varepsilon$ for some r

goal: prove $P[A] \leq 1/2$

$\mu(r) = \frac{|P \cap r|}{|P|}$

$\mu_Q(r) = \frac{|Q \cap r|}{|Q|}$

$R \wedge Q = \{ r \cap Q : r \in R \}$

$g(n) = \max \{ |R \wedge Q| : |Q| \leq n \}$

ε-sample $|\mu_Q(r) - \mu(r)| \leq \varepsilon \ \forall r$

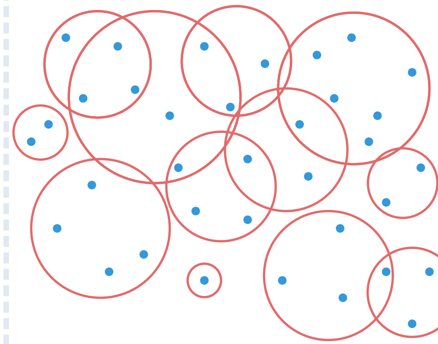
ε-net $\mu_Q(r) > 0 \ \forall \mu(r) \geq \varepsilon$

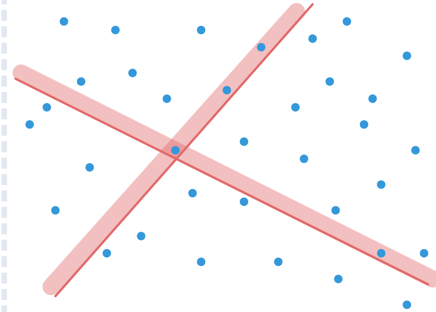
$$\left.\begin{matrix} P\left[\frac{X_1 + \cdots + X_n}{n} \geq \mu + \varepsilon\right] \\ P\left[\frac{X_1 + \cdots + X_n}{n} \leq \mu - \varepsilon\right] \end{matrix}\right\} \leq e^{-2\varepsilon^2 n}$$

let Q_1 sample k points (w/ repetition)

ε-sample theorem

$k \geq O\left(\frac{\ln(g(2k))}{\varepsilon^2}\right)$





define event

$A =$ (event that) $|\mu_1(r) - \mu(r)| > \varepsilon$ for some r

(!) let Q_2 sample k points

$B =$ (event that) $|\mu_2(r) - \mu_1(r)| > \frac{\varepsilon}{2}$ for some r

$\mu(r) = \frac{|P \cap r|}{|P|}$

$\mu_Q(r) = \frac{|Q \cap r|}{|Q|}$

$C = A \wedge B$

$=$ (event that) $|\mu_1(r) - \mu(r)| > \varepsilon$ and $|\mu_2(r) - \mu_1(r)| > \frac{\varepsilon}{2}$ for some r

$R \wedge Q = \{r \cap Q : r \in R\}$

$g(n) = \max\{|R \wedge Q| : |Q| \leq n\}$

ε-sample $|\mu_Q(r) - \mu(r)| \leq \varepsilon \;\; \forall r$

ε-net $\mu_Q(r) > 0 \;\; \forall \mu(r) \geq \varepsilon$

$$\left.\begin{aligned} P\left[\frac{X_1 + \cdots + X_n}{n} \geq \mu + \varepsilon\right] \\ P\left[\frac{X_1 + \cdots + X_n}{n} \leq \mu - \varepsilon\right] \end{aligned}\right\} \leq e^{-2\varepsilon^2 n}$$

let Q_1, Q_2 sample k points (w/ repetition)

$A = \left(\begin{smallmatrix}\text{event}\\ \text{that}\end{smallmatrix}\right)$ $|\mu_1(r) - \mu(r)| > \varepsilon$ for some r

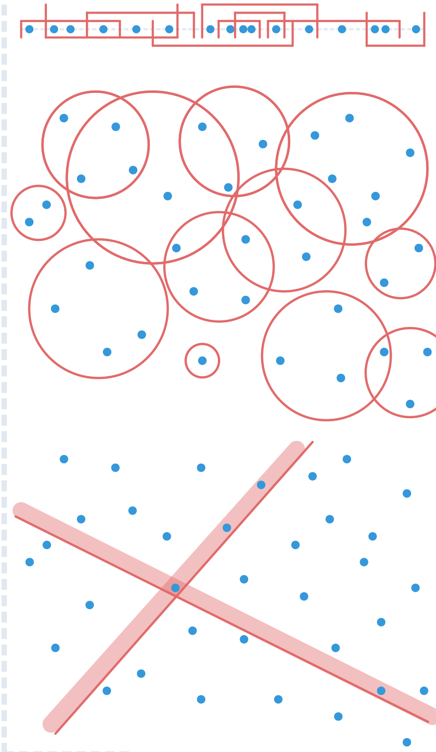
$B = \left(\begin{smallmatrix}\text{event}\\ \text{that}\end{smallmatrix}\right)$ $|\mu_2(r) - \mu_1(r)| > \frac{\varepsilon}{2}$ for some r

$C = \left(\begin{smallmatrix}\text{event}\\ \text{that}\end{smallmatrix}\right)$ $|\mu_1(r) - \mu(r)| > \varepsilon$ and $|\mu_2(r) - \mu_1(r)| > \frac{\varepsilon}{2}$ for some r

claim: $P[A] \le 2P[B]$

ε-sample theorem

$k \ge O\left(\frac{\ln(g(2k))}{\varepsilon^2}\right)$



$\mu(r) = \frac{|P \cap r|}{|P|}$

$\mu_Q(r) = \frac{|Q \cap r|}{|Q|}$

$R \wedge Q = \{r \cap Q : r \in R\}$

$g(n) = \max\{|R \wedge Q| : |Q| \le n\}$

ε-sample $|\mu_Q(r) - \mu(r)| \le \varepsilon \ \forall r$

ε-net $\mu_Q(r) > 0 \ \forall \mu(r) \ge \varepsilon$

$$\left.\begin{array}{l} P\left[\frac{X_1 + \cdots + X_n}{n} \ge \mu + \varepsilon\right] \\ P\left[\frac{X_1 + \cdots + X_n}{n} \le \mu - \varepsilon\right] \end{array}\right\} \le e^{-2\varepsilon^2 n}$$

let Q_1, Q_2 sample k points (w/ repetition)

$A =$ (event that) $|\mu_1(r) - \mu(r)| > \varepsilon$ for some r

$B =$ (event that) $|\mu_2(r) - \mu_1(r)| > \frac{\varepsilon}{2}$ for some r

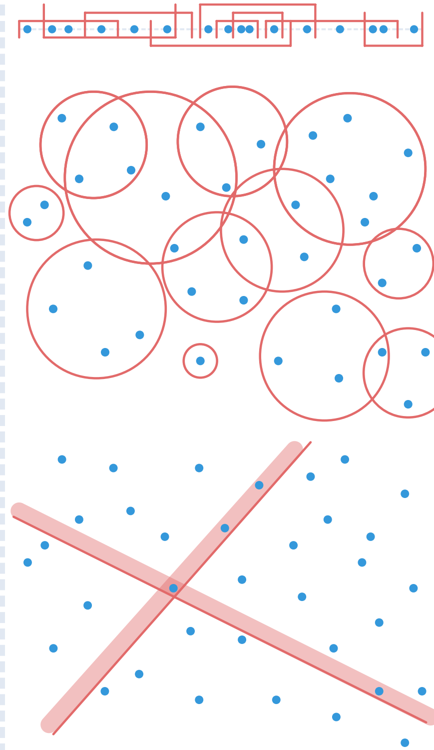
✓ $P[A] \leq 2P[B]$

goal: prove $P[B] \leq \delta$

ideas?

ε-sample theorem

$k \geq O\left(\frac{\ln(g(2k))}{\varepsilon^2}\right)$



$\mu(r) = \frac{|P \cap r|}{|P|}$

$\mu_Q(r) = \frac{|Q \cap r|}{|Q|}$

$R \wedge Q = \{r \cap Q : r \in R\}$

$g(n) = \max\{|R \wedge Q| : |Q| \leq n\}$

ε-sample $|\mu_Q(r) - \mu(r)| \leq \varepsilon \ \forall r$

ε-net $\mu_Q(r) > 0 \ \forall \mu(r) \geq \varepsilon$

$$\left.\begin{array}{l} P\left[\frac{X_1 + \cdots + X_n}{n} \geq \mu + \varepsilon\right] \\ P\left[\frac{X_1 + \cdots + X_n}{n} \leq \mu - \varepsilon\right] \end{array}\right\} \leq e^{-2\varepsilon^2 n}$$

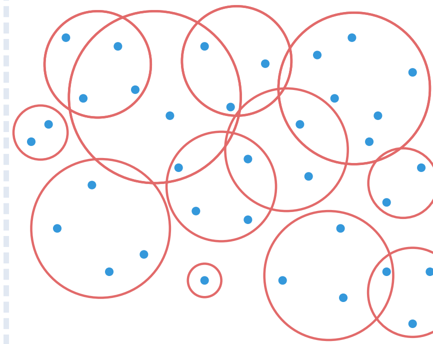
let Q_1, Q_2 sample k points (w/ repetition)

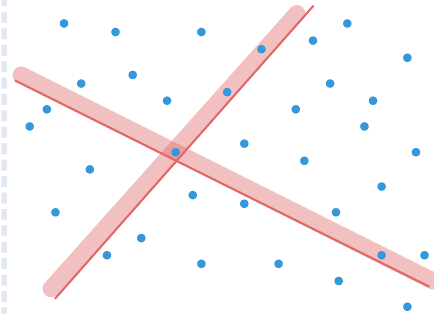
$A = \left(\begin{smallmatrix}\text{event}\\ \text{that}\end{smallmatrix}\right) |\mu_1(r) - \mu(r)| > \varepsilon$ for some r

$B = \left(\begin{smallmatrix}\text{event}\\ \text{that}\end{smallmatrix}\right) |\mu_2(r) - \mu_1(r)| > \frac{\varepsilon}{2}$ for some r

✓ $P[A] \le 2P[B]$

$\boxed{\begin{array}{l}\varepsilon\text{-sample theorem}\\ k \ge O\left(\frac{\ln(g(2k))}{\varepsilon^2}\right)\end{array}}$





Suppose we generate Q_1, Q_2 as follows:

① sample $2k$ points Q_0

② randomly partition Q_0 into halves Q_1, Q_2

$$\mu(r) = \frac{|P \cap r|}{|P|}$$

$$\mu_Q(r) = \frac{|Q \cap r|}{|Q|}$$

$$P[B] = \sum_{Q_0} P[B|Q_0] \cdot P[Q_0]$$

$$\le \max_{Q_0} P[B|Q_0]$$

$$R \wedge Q = \{r \cap Q : r \in R\}$$

$$g(n) = \max\{|R \wedge Q| : |Q| \le n\}$$

ε-sample $|\mu_Q(r) - \mu(r)| \le \varepsilon \;\; \forall r$

ε-net $\mu_Q(r) > 0 \;\; \forall \mu(r) \ge \varepsilon$

$$\left.\begin{array}{l} P\left[\frac{X_1 + \cdots + X_n}{n} \ge \mu + \varepsilon\right] \\ P\left[\frac{X_1 + \cdots + X_n}{n} \le \mu - \varepsilon\right] \end{array}\right\} \le e^{-2\varepsilon^2 n}$$

goal: prove $P[B|Q_0] \le \frac{1}{4} \;\; \forall Q_0$

$A =$ (event that) $|\mu_1(r) - \mu(r)| > \varepsilon$ for some r

$B =$ (event that) $|\mu_2(r) - \mu_1(r)| > \frac{\varepsilon}{2}$ for some r

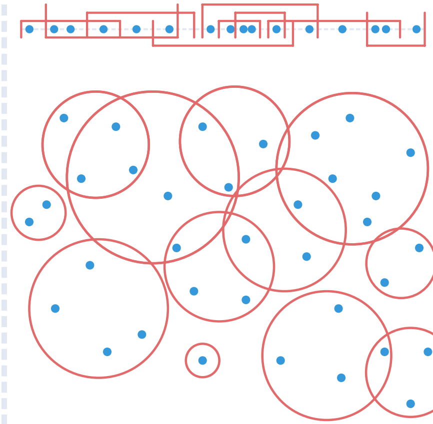
✓ $P[A] \leq 2P[B]$

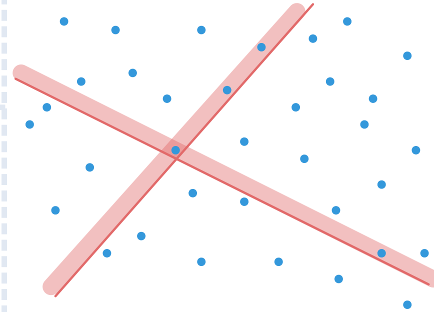
① sample $2k$ points Q_0

② randomly partition Q_0 into halves Q_1, Q_2

ε-sample theorem

$$k \geq O\left(\frac{\ln(g(2k))}{\varepsilon^2}\right)$$





Fix Q_0. let $\mu_0 =$ measure w/r/t Q_0

$$|R \wedge Q_0| \leq g(2k) \quad (!)$$

Additive Chernoff:

$$\left.\begin{array}{l} P\left[|\mu_1(r) - \mu_0(r)| \geq \frac{\varepsilon}{4}\right] \\ P\left[|\mu_2(r) - \mu_0(r)| \geq \frac{\varepsilon}{4}\right] \end{array}\right\} \leq \frac{1}{8g(2k)}$$

$$\forall r \in R \wedge Q_0$$

$$\mu(r) = \frac{|P \cap r|}{|P|}$$

$$\mu_Q(r) = \frac{|Q \cap r|}{|Q|}$$

$$R \wedge Q = \{r \cap Q : r \in R\}$$

$$g(n) = \max\{|R \wedge Q| : |Q| \leq n\}$$

ε-sample $\quad |\mu_Q(r) - \mu(r)| \leq \varepsilon \ \forall r$

ε-net $\quad \mu_Q(r) > 0 \ \forall \mu(r) \geq \varepsilon$

$$\left.\begin{array}{l} P\left[\frac{X_1 + \cdots + X_n}{n} \geq \mu + \varepsilon\right] \\ P\left[\frac{X_1 + \cdots + X_n}{n} \leq \mu - \varepsilon\right] \end{array}\right\} \leq e^{-2\varepsilon^2 n}$$

Union bound $\Rightarrow$

$$|\mu_1(r)-\mu_0(r)| \le \varepsilon/4, \quad |\mu_2(r)-\mu_0(r)| \le \varepsilon/4$$

$\forall r \in R \wedge Q_0$ w/ prob $\ge 3/4$

$\forall r \in R \wedge Q_0$,

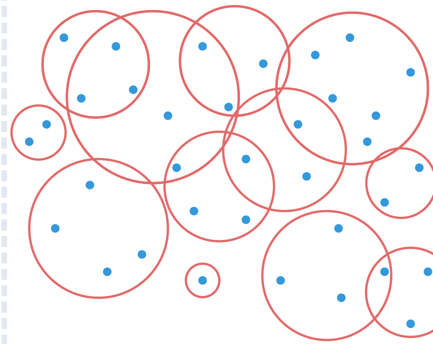
$$|\mu_1(r)-\mu_2(r)| \le |\mu_1(r)-\mu_0(r)| + |\mu_2(r)-\mu_0(r)|$$

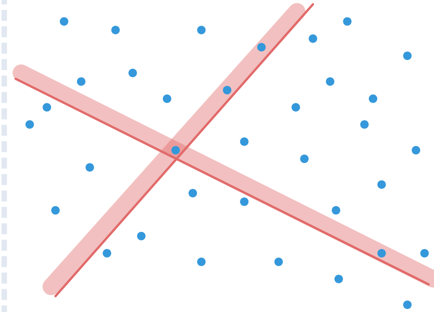
$$\le \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \frac{\varepsilon}{2}$$

$\Rightarrow$ same $\forall r \in R$

so

$$P\left[B = \left(\substack{\text{event} \\ \text{that}}\right) |\mu_2(r)-\mu_1(r)| > \frac{\varepsilon}{2} \text{ for some } r\right] \le \frac{1}{4}$$





$\mu(r) = \frac{|P \cap r|}{|P|}$

$\mu_Q(r) = \frac{|Q \cap r|}{|Q|}$

$R \wedge Q = \{ r \cap Q : r \in R \}$

$g(n) = \max \{ |R \wedge Q| : |Q| \le n \}$

ε-sample $|\mu_Q(r) - \mu(r)| \le \varepsilon \ \forall r$

ε-net $\mu_Q(r) > 0 \ \forall \mu(r) \ge \varepsilon$

$$\left.\begin{array}{l} P\left[\frac{X_1 + \dots + X_n}{n} \ge \mu + \varepsilon\right] \\ P\left[\frac{X_1 + \dots + X_n}{n} \le \mu - \varepsilon\right] \end{array}\right\} \le e^{-2\varepsilon^2 n}$$

ε-sample theorem [VC71]

$$k \ge O\left(\frac{\ln(g(2k))}{\varepsilon^2}\right) \Rightarrow \varepsilon\text{-sample}$$

(w/ constant probability)

similar, exercise

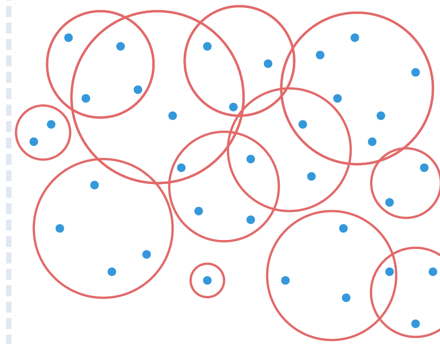
ε-net theorem [HW87]

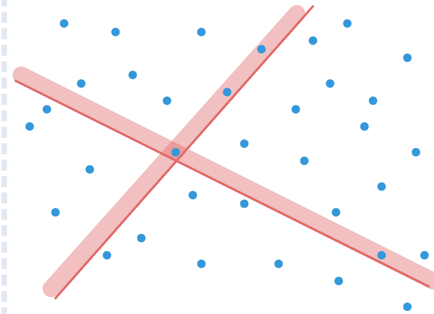
$$k \ge O\left(\frac{\ln(g(2k))}{\varepsilon}\right) \Rightarrow \varepsilon\text{-net}$$

(w/ constant probability)

next question:

how to bound $g(n)$?





$$\mu(r) = \frac{|P \cap r|}{|P|}$$

$$\mu_Q(r) = \frac{|Q \cap r|}{|Q|}$$

$$R \wedge Q = \{ r \cap Q : r \in R \}$$

$$g(n) = \max \{ |R \wedge Q| : |Q| \le n \}$$

ε-sample $\quad |\mu_Q(r) - \mu(r)| \le \varepsilon \;\; \forall r$

ε-net $\quad \mu_Q(r) > 0 \;\; \forall \mu(r) \ge \varepsilon$

$$\left. \begin{array}{l} P\left[\frac{X_1 + \cdots + X_n}{n} \ge \mu + \varepsilon\right] \\ P\left[\frac{X_1 + \cdots + X_n}{n} \le \mu - \varepsilon\right] \end{array} \right\} \le e^{-2\varepsilon^2 n}$$

Lecture 12

Probability (and techniques)	Algorithms
✓ Additive Chernoff Inequality	✓ ε-samples and ε-nets
✓ Double-sampling	
VC dimension	

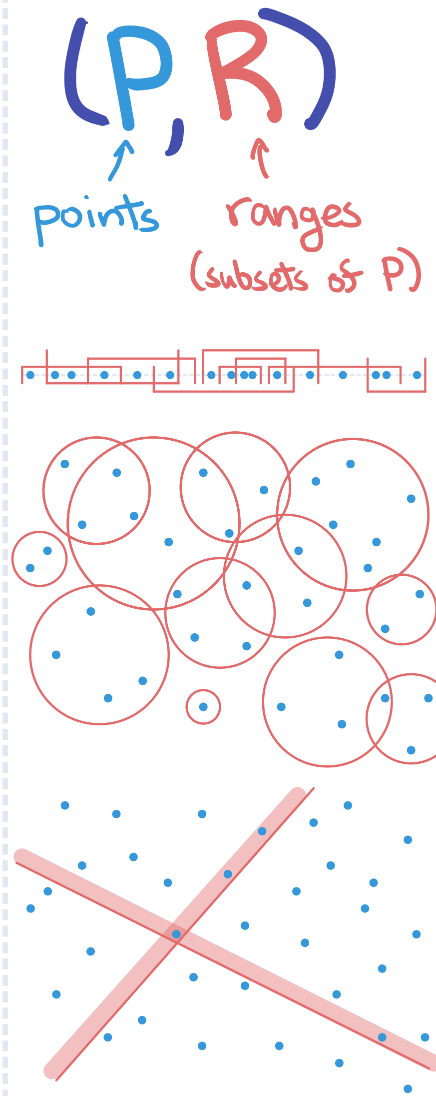
VC dimension

for $Q \subset P$, R shatters Q if

$$R \cap Q = \text{all } 2^{|Q|} \text{ subsets of } Q$$

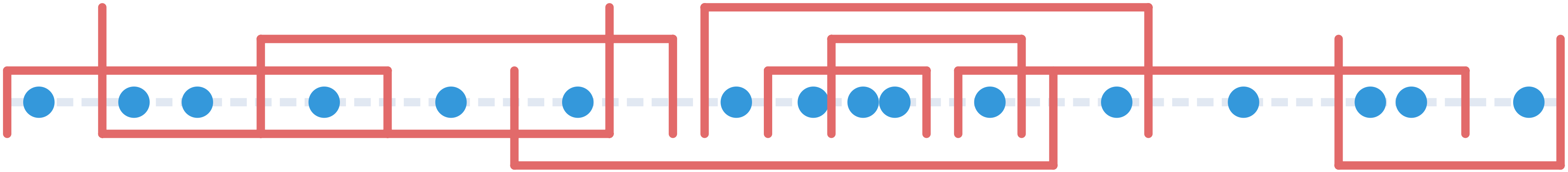
VC dimension of (P, R) =

$$\max |Q| \text{ s.t. } R \text{ shatters } Q$$



for $Q \subset P$, R shatters Q if $R \cap Q$ = all $2^{|Q|}$ subsets of Q

VC dimension of (P, R) = max $|Q|$ s.t. R shatters Q

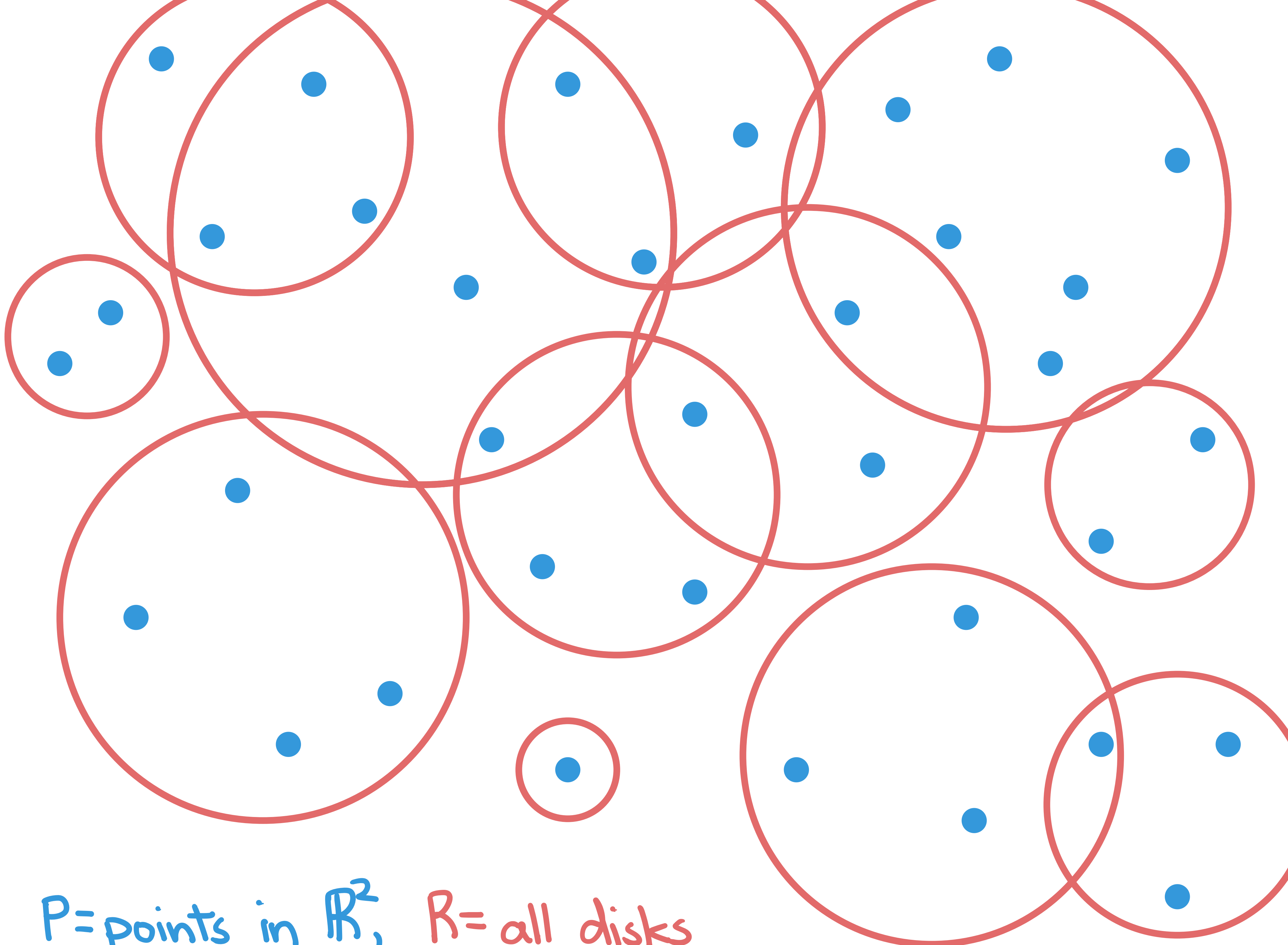


P = points in $\mathbb{R}$, R = intervals for $Q \subset P$, R shatters Q if $R \cap Q$ = all $2^{|Q|}$ subsets of Q

VC dimension of (P, R) = max $|Q|$ s.t. R shatters Q

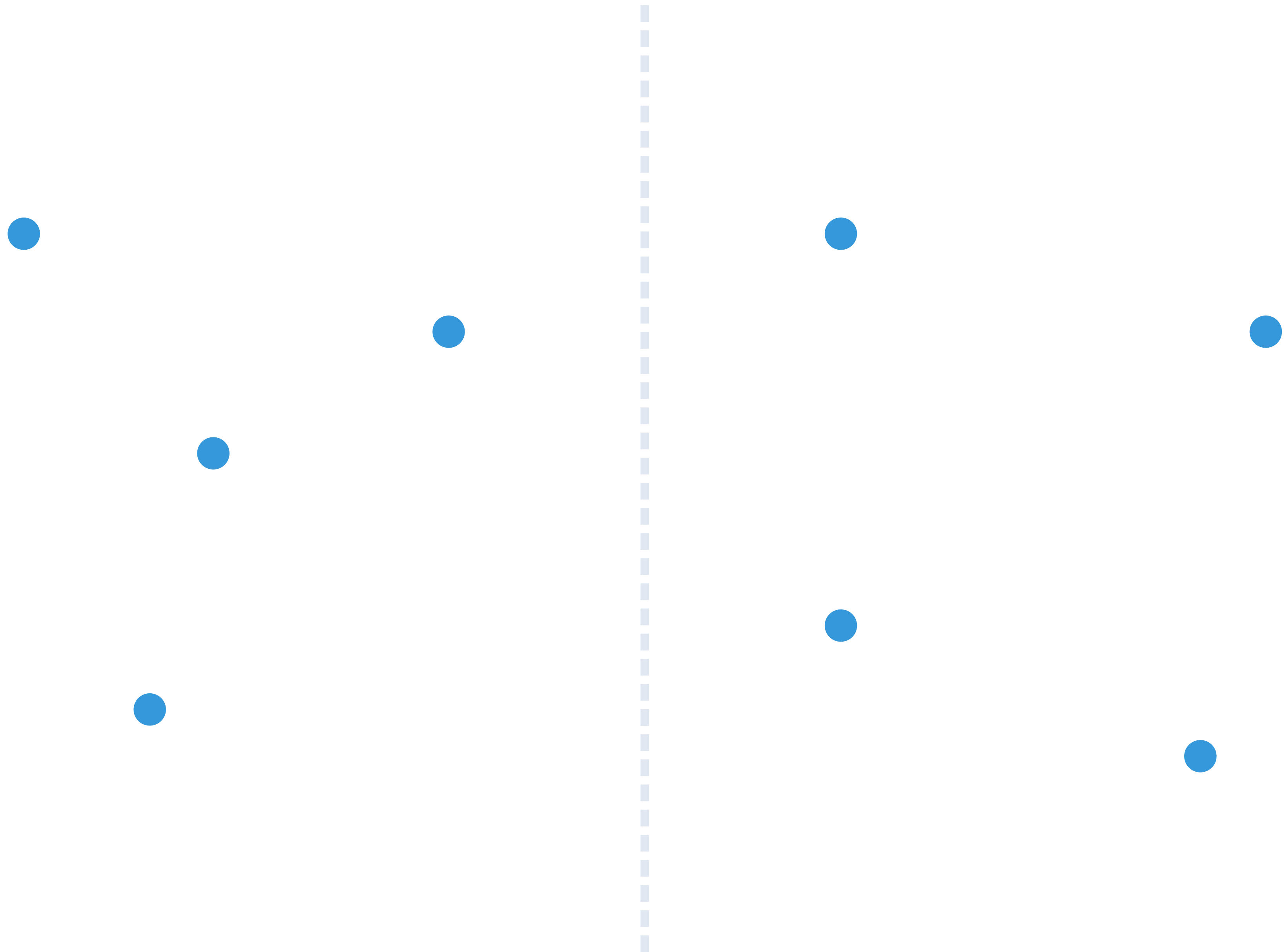
VC-dim = 3

P = points in $\mathbb{R}$, R = intervals

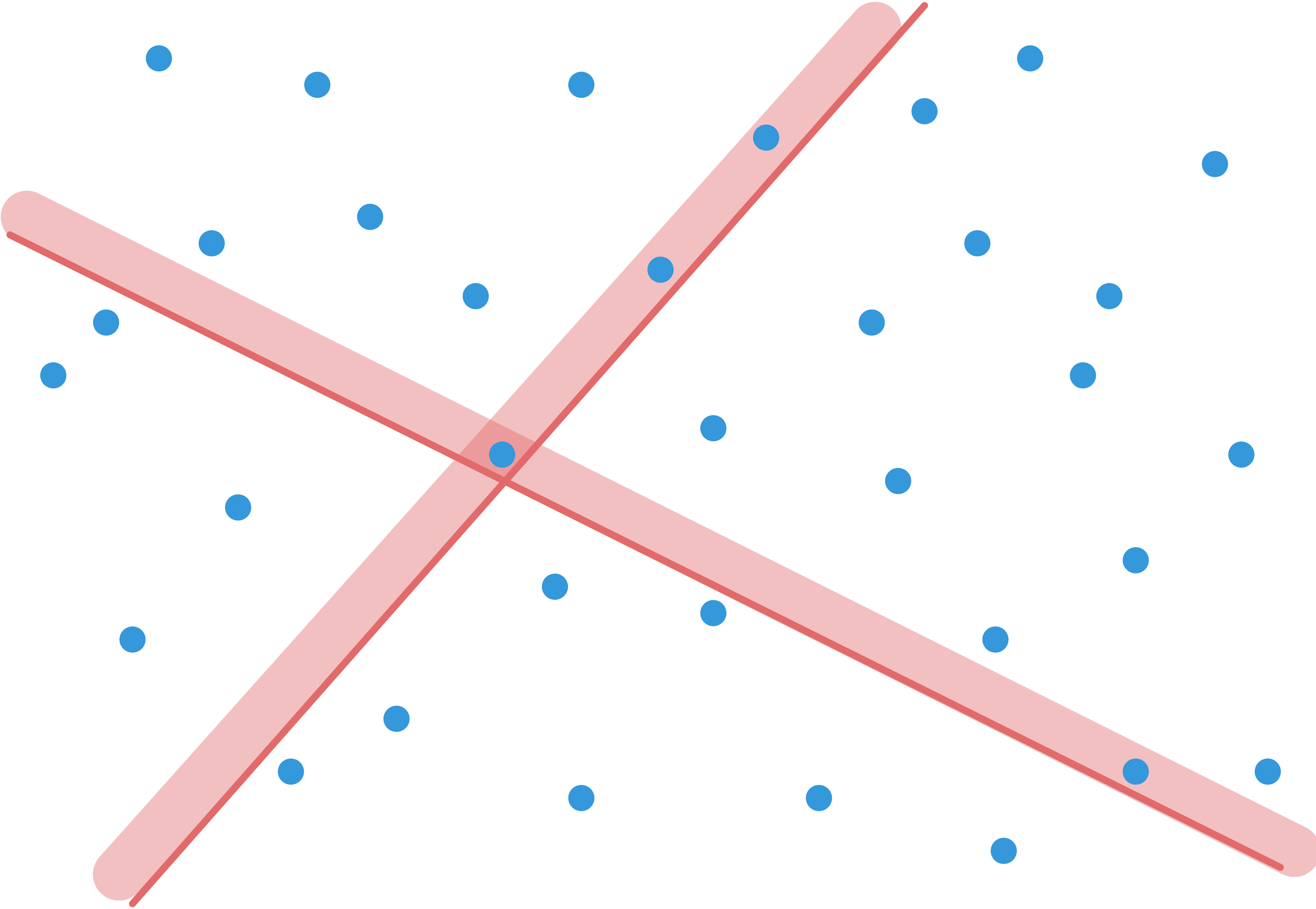


P = points in $\mathbb{R}^2$, R = all disks

VC-dim = 4



P = points in $\mathbb{R}^2$, R = all disks

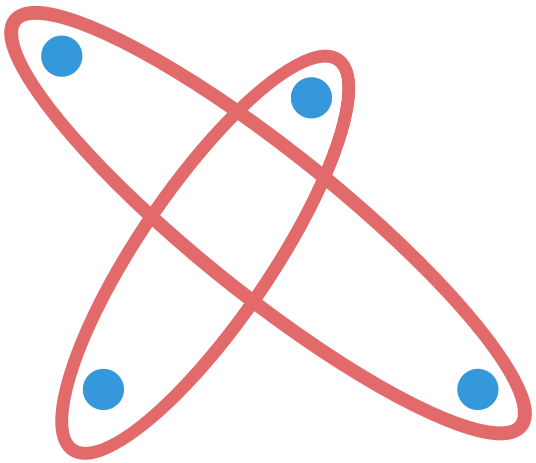


P = points in $\mathbb{R}^2$, R = half-spaces

Radon's theorem $Q \subset \mathbb{R}^d$, $|Q| = d+2$.

$\exists$ partition $Q = S_1 \cup S_2$ s.t.

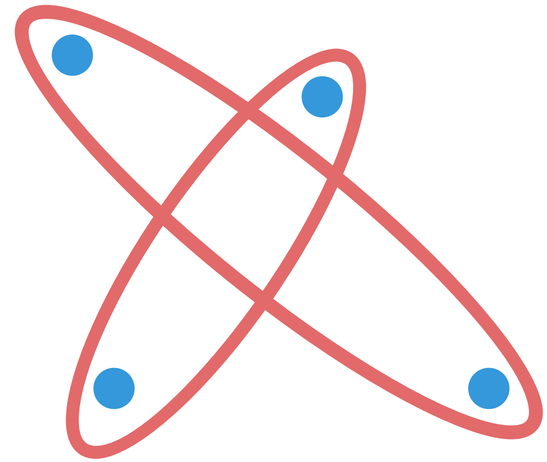
$$\text{convex-hull}(S_1) \cap \text{convex-hull}(S_2) \neq \emptyset$$



Radon's theorem $Q \subset \mathbb{R}^d$, $|Q| = d+2$.

$\exists$ partition $Q = S_1 \cup S_2$ s.t.

$\text{convex-hull}(S_1) \cap \text{convex-hull}(S_2) \neq \emptyset$



$\Rightarrow$ VC-dimension $= d+1$

Sauer's lemma let (P, R) have VC-dim d.

then

$$g(n) \le \sum_{i=0}^{d} \binom{n}{i} \le n^d$$

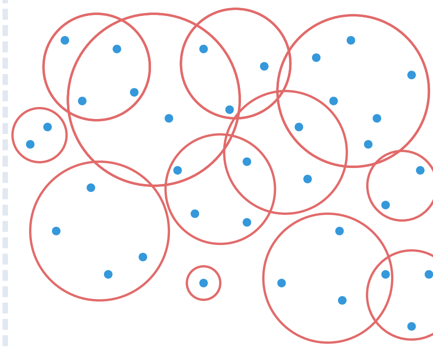
e.g. intervals: $g(n) \le n^2$

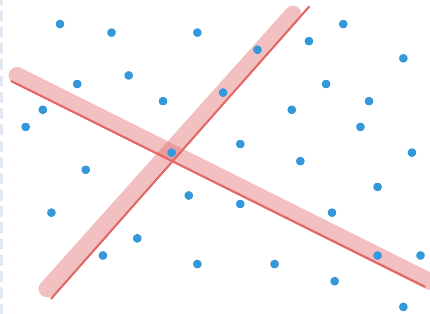
disks in $\mathbb{R}^2$: $g(n) \le n^3$

halfspaces in $\mathbb{R}^d$: $g(n) \le n^{d+1}$

(P, R)

points ranges (subsets of P)





for $Q \subset P$, R shatters Q if $R \wedge Q$ = all subsets of Q

VC dimension of (P, R) = max $|Q|$ s.t. R shatters Q

Sauer's lemma let (P,R) have VC-dim d.

then
$$g(n) \le \sum_{i=0}^{d} \binom{n}{i} \le n^d$$

Proof let $f(d,n) = \sum_{i=0}^{d} \binom{n}{i}$.

base cases:

$n=0 \Rightarrow g(n) = 0 \le f(d,0)$

$VC=0 \Rightarrow g(n) \le 1 \le f(0,n)$

case $n,d>0$: fix Q w/ $n=|Q|$, $p \in Q$.

let $R^p = \{r-p : r \in R\}$

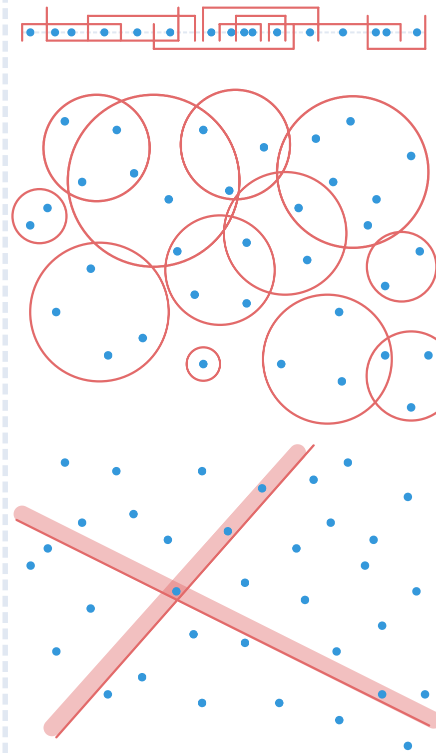
$R_p = \{r-p : r \in R,\ r \cup \{p\} \in R,\ r \setminus \{p\} \in R\}$

("$r+p$" $= r \cup \{p\}$, "$r-p$" $= r \setminus \{p\}$)

(P, R)

points

ranges (subsets of P)



for $Q \subset P$, R shatters Q if $R \wedge Q$ = all subsets of Q

VC dimension of (P,R) = max $|Q|$ s.t. R shatters Q

Sauer's lemma let (P,R) have VC-dim d.

then

$$g(n) \le \sum_{i=0}^{d} \binom{n}{i} \le n^d$$

<u>**Proof**</u> let $f(d,n) = \sum_{i=0}^{d} \binom{n}{i}$.

Base cases: $n=0$, $VC=0$ ✓

case $n,d>0$: fix point p.

Remove p: $R^p = \{r-p : r \in R\}$

R^p has VC-dimension $\le d$

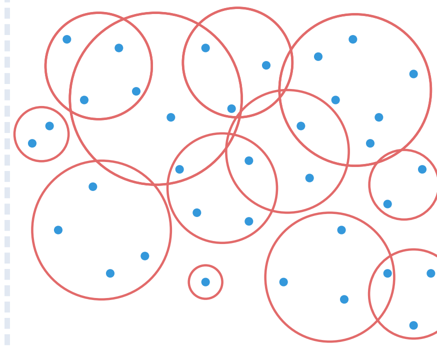
Induction: $|R^p| \le \binom{n-1}{d}$

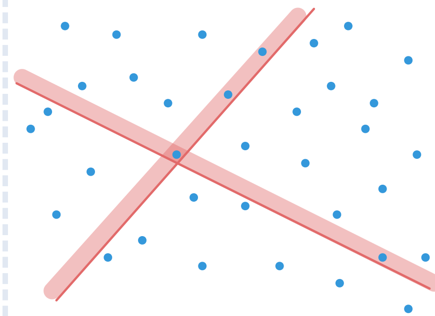
$|R| \ge |R^p|$: how to interpret $|R|-|R^p|$?

("$r+p$" $= r \cup \{p\}$, "$r-p$" $= r \setminus \{p\}$)

(P, R)

points ranges (subsets of P)





for $Q \subset P$, R shatters Q if $R \wedge Q$ = all subsets of Q

VC dimension of (P,R) = max $|Q|$ s.t. R shatters Q

Remove p: $R^p = \{r - p : r \in R\}$ VC-dim $\leq d$

Induction: $|R^p| \leq \binom{n-1}{d}$

(P, R)

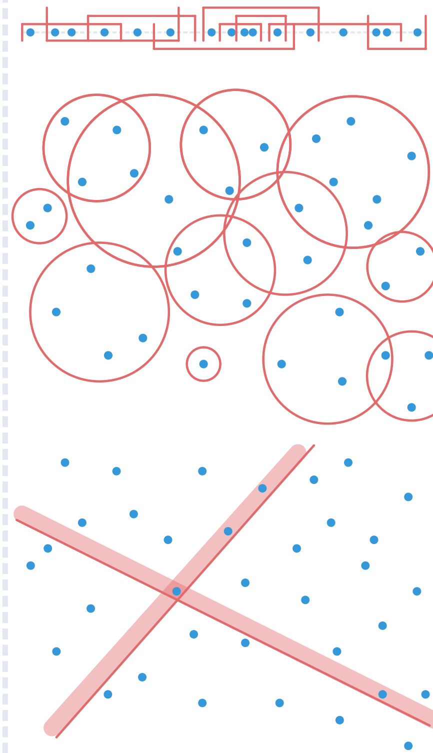
points ranges (subsets of P)

$|R| \geq |R^p|$: how to interpret $|R| - |R^p|$?

"double-counted ranges"

$R_p = \{r - p : r \in R,\ r \cup \{p\} \in R,\ r \setminus \{p\} \in R\}$

claim: $|R_p| + |R^p| = |R|$



for $Q \subset P$, R shatters Q if $R \wedge Q$ = all subsets of Q

VC dimension of (P,R) = max $|Q|$ s.t. R shatters Q

("$r+p$" $= r \cup \{p\}$, "$r-p$" $= r \setminus \{p\}$)

Remove p: $R^p = \{r - p : r \in R\}$ VC-dim $\le d$

Induction: $|R^p| \le \binom{n-1}{d}$

(P, R)

points

ranges (subsets of P)

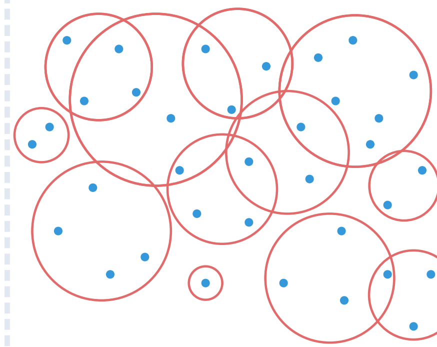
$|R| \ge |R^p|$: how to interpret $|R| - |R^p|$?

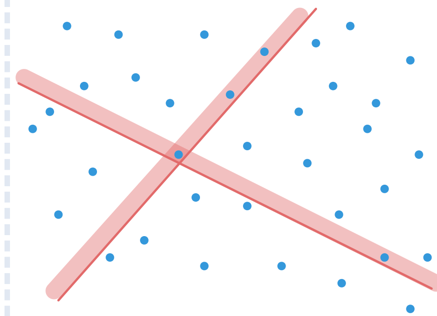
"double-counted ranges"

$$R_p = \{r - p : r \in R, \ r \cup \{p\} \in R, \ r \setminus \{p\} \in R\}$$

✓ $|R_p| + |R^p| = |R|$

R_p has VC-dim $d-1$ (!)





for $Q \subset P$, R shatters Q if $R \cap Q$ = all subsets of Q

VC dimension of (P,R) = max $|Q|$ s.t. R shatters Q

$$f(d,n) = \sum_{i=0}^{d} \binom{n}{i}.$$

Sauer's lemma let (P,R) have VC-dim d.

then

$$g(n) \le \sum_{i=0}^{d} \binom{n}{i} \le n^d$$

let $f(d,n) = \sum_{i=0}^{d} \binom{n}{i}$.

R^P has VC dim d

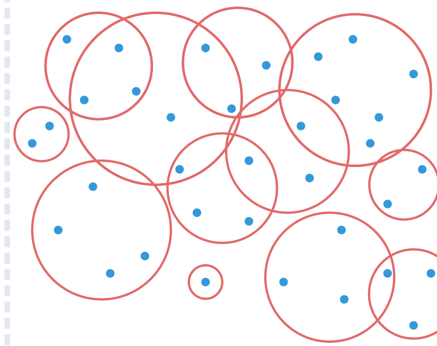
R_p has VC dim $d-1$

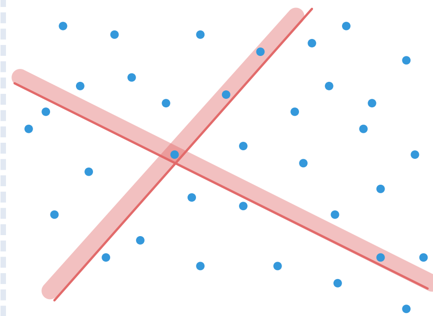
✓ $|R_p| + |R^P| = |R|$

$$\begin{aligned}
|R| &= |R_p| + |R^P| \\
&\le f(d-1,n-1) + f(d,n-1) \\
&= \sum_{i=0}^{d-1} \binom{n-1}{i} + \sum_{i=0}^{d} \binom{n-1}{i} \\
&\le \sum_{i=0}^{d} \binom{n-1}{i-1} + \sum_{i=0}^{d} \binom{n-1}{i}
\end{aligned}$$

(P, R)

points ranges (subsets of P)





for $Q \subset P$, R shatters Q if $R \wedge Q$ = all subsets of Q

VC dimension of (P,R) = max $|Q|$ s.t. R shatters Q

$$\leq \sum_{i=0}^{d} \binom{n}{i} = \Phi(d, n)$$