

Geometric Sampling

Lecture 1

Probability

Events, probabilities,
conditional probabilities,
union bound

Random variables,
expected value,
linearity of expectation
Markov's inequality

Algorithms

Approx. 3-SAT,

Big idea: we only want the average

$$\begin{array}{l} \text{Linearity of Expectation} \\ E[X+Y] = E[X] + E[Y] \quad (\text{stochastic variables}) \\ \text{"average sum = sum of averages"} \end{array}$$

let $Y_i = \begin{cases} 1 & \text{if } i\text{th clause satisfied} \\ 0 & \text{otherwise} \end{cases}$ ($i=1, \dots, m$)

$$E[Y_i] = 7/8 \quad (1 \text{ clause case})$$

$$E[\text{\# clauses satisfied}] = E\left[\sum_{i=1}^m Y_i\right] = \sum_{i=1}^m E[Y_i] = 7/8 m$$

let $Y_i = \begin{cases} 1 & \text{if } i\text{th clause satisfied} \\ 0 & \text{otherwise} \end{cases}$

$$Z = \sum_{i=1}^m Y_i = \text{\# clauses satisfied}$$

$$E[Z] = \frac{1}{2} E[Z | x_1 = \text{true}] + \frac{1}{2} E[Z | x_1 = \text{false}]$$

$\Rightarrow$ either $E[Z | x_1 = \text{true}]$ or $E[Z | x_1 = \text{false}]$ is $\geq E[Z]$!

QuickSort,

Random divide and conquer

① pick random 'pivot'

② Divide A into two parts



Expected running time of QuickSort

For each $i, j \in [n]$ w/ $i < j$,

$X_{ij} = \begin{cases} 1 & \text{if } i\text{th smallest element is compared to } j\text{th smallest} \\ 0 & \text{otherwise} \end{cases}$

$\sum_{i,j} X_{ij}$ = total # comparisons made by the algorithm
 $\approx$ total running time

$E[\sum_{i,j} X_{ij}]$ = average running time

$$E[\sum_{i,j} X_{ij}] = \sum_{i,j} E[X_{ij}]$$

now analyze $E[X_{ij}]$ in isolation

$$E[X_{ij}] = \Pr[X_{ij} = 1] = \Pr[i\text{th element is compared to } j\text{th}]$$

i th and j th elem selected before any element in between happens w/ probability $\frac{2}{j-i+1}$

$$E[X_{ij}] = \frac{2}{j-i+1}$$

by linearity of expectation

$$E[\sum_{i,j} X_{ij}] = \sum_{i,j} \frac{2}{j-i+1}$$

$$= 2 \sum_{i,j} \frac{1}{j-i+1} = 2 \sum_{k=1}^n \sum_{i=1}^k \frac{1}{k+1}$$

$$= 2 \sum_{k=1}^n \frac{k}{k+1} \leq 2n \left(1 + \frac{1}{k+1}\right) \leq O(n \log n)$$

with harmonic series

Quickselect (HW)

Lecture 2

Probability

Hash Functions,

Universal Hash Functions

$h: [n] \rightarrow k$ s.t. for all pairs $i, j \in [n]$,
 $\Pr[h(i) = h(j)] \leq \frac{1}{k}$ (much weaker than ideal hash)

e.g. $h(x) = (ax + b) \bmod p \bmod k$

where • p prime, $> k$

• a in $\{1, \dots, p-1\}$ unif. random

• b in $\{0, \dots, p-1\}$ unif. random



Markov's Inequality

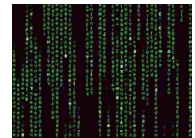
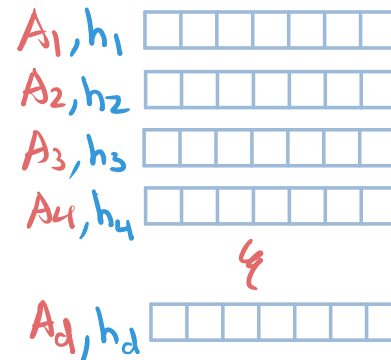
Markov's Inequality (special case)

Let $X \geq 0$ be nonnegative RV

$$\Pr[X \geq 2E[X]] \leq \frac{1}{2}$$

Algorithms

Frequency Estimation, Heavy Hitters (in streams)

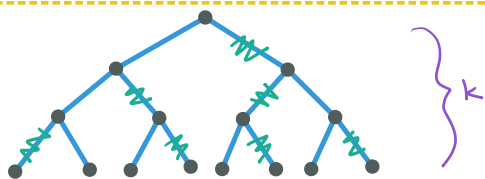


estimate f_i w/
 $\min_{j=1, \dots, d} A_j[h_j(i)]$

Lecture 4

Probability

Galton-Watson processes



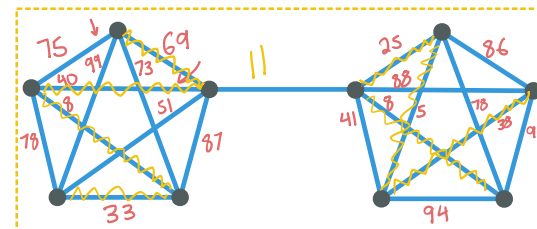
let T = complete binary tree of height k ,
w/ every edge deleted w/ prob $\leq 1/2$

Fact: $P[\text{some leaf connected to root}] \geq \frac{1}{k}$

"Galton-Watson process"

Algorithms

Random Contractions for min cut



Randomized contractions

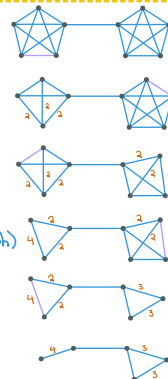
Input: $G=(V,E)$, weights $c \in \mathbb{R}_{\geq 0}^E$
(unweighted)

① while $|E| > 1$

② sample $e \in E$ in proportion to c

③ "contract" e

④ return set of edges (in original graph)
contracted to remaining edge



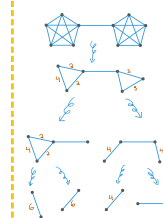
Amplification by branching

Instead of restarting at the beginning, restart partway through

High-level algorithm

① randomly contract until prob.
of ruining the min-cut is $\sim 1/2$

② Branch:
make two copies of current graph
recurse on both copies



Lecture 5

Probability

Multiplicative Chernoff Bound

Multiplicative Chernoff bound

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

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

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

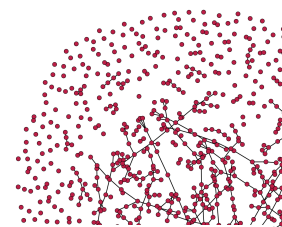
- $P[X_1 + \dots + X_n \geq (1 + \epsilon)\mu] \leq e^{-\epsilon^2 \mu / 3}$
- $P[X_1 + \dots + X_n \leq (1 - \epsilon)\mu] \leq e^{-\epsilon^2 \mu / 2}$

Algorithms (/ results)

Random graphs & the giant component

Erdős-Rényi graphs $G(n, p)$

- n vertices
- each edge sampled independently w/ probability p



Theorem 5.3. Consider a random graph $G \sim \mathcal{G}(n, p)$ for $p = c/n$, where c is a constant.

- If $c > 1$, then with high probability, there is exactly one connected component of G with $\Omega(n)$ vertices, and all other components have size $\leq O(\log n)$.
- For $c < 1$, then with high probability, all connected components of G has size $< O(\log n)$.

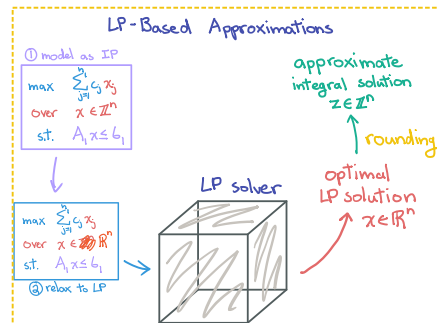
Lecture 6

Probability (& techniques)

Approximations via linear programming & randomized rounding

Linear programs (LPs)
Similar to IPs, except no integrality constraints

min/max $\sum_{j=1}^n c_j x_j$ linear objective
over $x \in \mathbb{R}^n$ linear integral variables
st. $A_1 x \leq b_1$ linear constraints
 $A_2 x = b_2$
 $A_3 x \geq b_3$



Continuous optimization vs Discrete optimization

randomization

Algorithms

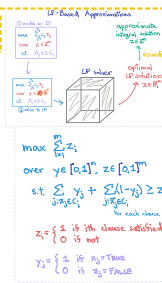
Max Sat Set Cover Covering Integer Programs

For each variable x_j , independently, randomly set $x_j = 1$ with probability p_j , and $x_j = 0$ otherwise.

$$E[\# \text{ sat. clauses}] \geq (1 - 1/e) \text{OPT}$$

k	oblivious	LP	average
1	1/2	$z_i \geq (3/4)z_i$	$\geq (3/4)z_i$
2	3/4	$(3/4)z_i \geq (3/4)z_i$	$\geq (3/4)z_i$
3	7/8	$(1 - 1/e)z_i \geq (3/4)z_i$	$\geq (3/4)z_i$
4	15/16	$(1 - 1/e)z_i \geq (3/4)z_i$	$\geq (3/4)z_i$
5	31/32	$(1 - 1/e)z_i \geq (3/4)z_i$	$\geq (3/4)z_i$
...	...	...	...
k	$1 - 1/2^k$	$(1 - 1/e)z_i \geq (3/4)z_i$	$\geq (3/4)z_i$

Can we get $\frac{3}{4}$ -APX?



Rounding LP solution
 $x \in [0, 1]^n \rightsquigarrow F \subseteq \{S_1, \dots, S_n\}$

For each set S_j , independently, take S_j with probability x_j .

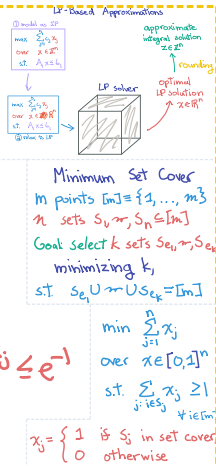
$$E[|F|] = \sum x_j \leq \text{OPT}$$

Fix a point i .

$$E[\# \text{ sets in } F \text{ covering } i] = \sum_{j \in S_i} x_j \geq z_i$$

$$P[F \text{ not cover } i] = \prod_{j \in S_i} (1 - x_j) \leq e^{-\sum_{j \in S_i} x_j} \leq e^{-z_i}$$

$$P[F \text{ cover } i] \geq 1 - e^{-1}$$



Lecture 7

Probability

Continuous Random Variables

Variance, Chebyshev's Variance Reduction

Median trick

Variance

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

Chebyshev's inequality

$$P[|X - E[X]| \geq d] \leq \frac{\text{Var}(X)}{d^2}$$

"median trick", "median of means"

- ① take mean to reduce variance, get constant prob. of error
- ② take median to amplify success prob. exponentially

Algorithms

Distinct Elements



Algorithm w/
limited space

Goal: estimate # of distinct elements in the stream

Suppose you had an ideal hash function

$$h: [m] \rightarrow [0, 1]$$

and the following facts:

let $Y_1, \dots, Y_k \in [0, 1]$ independent and uniformly random

$$X = \min \{Y_1, \dots, Y_k\}$$

$$\text{then } E[X] = \frac{1}{k+1}$$

$$\textcircled{1} X \leftarrow 1$$

② for each element e

$$\textcircled{a} X \leftarrow \min\{X, h(e)\}$$

$$/* E[X] = \frac{1}{k+1}, k = (\text{\# distinct elements}) */$$

$$\textcircled{3} \text{ return } \frac{1}{X} - 1$$

$$\text{recall } E[X] = \frac{1}{k+1}$$

$$\text{goal: show } |X - E[X]| \leq \frac{\epsilon}{k+1}$$

w/ good prob.

$$P[|X - E[X]| \geq \frac{\epsilon}{k+1}]$$

$$\leq \text{Var}(X) \cdot \left(\frac{k+1}{\epsilon}\right)^2 \leftarrow \text{Chebyshev's inequality}$$

where

$$\text{Var}(X) = \frac{k}{(k+1)^2(k+2)} \approx \frac{1}{k^2}$$

Q: how to reduce $\text{Var}(X)$?

we can get $(1 \pm \epsilon)$ -APX

w/ 10% error prob. in $O(\frac{1}{\epsilon^2})$ space

what about high probability?

① make q copies $Y_1, \dots, Y_q$ of

② return median of $Y_1, \dots, Y_q$

Algo correct $\Leftrightarrow \geq \frac{1}{2}$ of Y_i 's correct

let $Z = \# \text{correct } Y_i$'s $E[Z] = \frac{q}{2}$

$$P[Z \leq (1 - \frac{\epsilon}{4})E[Z]] \leq e^{-\left(\frac{\epsilon}{4}\right)^2 E[Z]/2} \leq \frac{1}{n^{O(1)}} \text{ for } q = O(\log n)$$

ideal hash $h: [m] \rightarrow [0, 1]$

let $Y_1, \dots, Y_k \in [0, 1]$

indep. and unifs. rand.

$$X = \min \{Y_1, \dots, Y_k\}$$

$$\text{then } E[X] = \frac{1}{k+1}$$

$$\text{and } E[X^2] = \frac{2}{(k+1)(k+2)}$$

$$\text{Var}(X) = E[(X - E[X])^2]$$

$$X_1, \dots, X_k \text{ indep. copies of } X$$

$$\bar{X} = \frac{1}{k}(X_1 + \dots + X_k)$$

$$\text{Var}(\bar{X}) = \frac{1}{k} \text{Var}(X)$$

$$P[X_1 + \dots + X_k \leq (1 - \epsilon)k] \leq e^{-\epsilon^2 k/4}$$

Lecture 8

Probability

Gaussian random variables

Gaussian random variables

$$x \sim \mathcal{N}(\mu, \sigma^2)$$

mean variance

Def: the nicest, most pleasant, convenient random variable ever

Fact let $x \sim \mathcal{N}(\mu, \sigma^2)$, $d \in \mathbb{R}$.

$$\text{then } x+d \sim \mathcal{N}(\mu+d, \sigma^2)$$

$$dx \sim \mathcal{N}(d\mu, d^2\sigma^2)$$

Fact let $x_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $x_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$

$$\text{then } x_1 + x_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Fact let $x \sim \mathcal{N}(0, 1)$, $t < \frac{1}{2}$

$$\text{then } E[e^{tx^2}] = \frac{1}{\sqrt{1-2t}}$$

Algorithms

Dimensionality reduction

Bag-of-words

literally any sentence is fine

any ↑ is literally sentence English
five encodes text in high dimensions

Lemma 8.8. Let $A \sim \mathcal{N}^{k \times d}$ be a random matrix where each coordinate A_{ij} is an independently drawn sample from $\mathcal{N}$. Let $\epsilon > 0$ be sufficiently small. For any vector x ,

$$\mathbb{P}\left[(1-\epsilon)\|x\|^2 \leq \frac{1}{k}\|Ax\|^2 \leq (1+\epsilon)\|x\|^2\right] \geq 1 - 2e^{-k/8}$$

$$k \left\{ \left[\frac{1}{k} \|Ax\|^2 \right] \leftarrow \alpha_i \right.$$

$$E[\|Ax\|^2] = \sum_{i=1}^k E[(Ax)_i^2] = k\|x\|^2$$

$$E[(\sum_{i=1}^k A_{ij} x_j)^2] = \|x\|^2$$

$$A_{ij} x_j \sim \mathcal{N}(0, x_j^2)$$

$$\sum_{j=1}^n A_{ij} x_j \sim \mathcal{N}(0, \|x\|^2)$$

$$(Ax) \sim \mathcal{N}^k(0, \|x\|^2)$$

$$\Pr[\|Ax\|^2 \geq k\|x\|^2 \cdot (1+\epsilon)] \leq e^{-\epsilon^2 k/8}$$

Dimensionality Reduction - high level

$O(\log n)$ dimensions suffice

where $n = \# \text{ points}$

$$\mathbb{R}^{\text{very large}} \xrightarrow{f(x)} \mathbb{R}^{O(\log n)}$$

One can map n points into $\mathbb{R}^{O(\log n/\epsilon^2)}$ while preserving all distances up to $(1 \pm \epsilon)$ -APX

Lecture 9

Probability

Locality-sensitive hashing

Fix $\theta \in [0, \pi]$. given q , either:

(a) output $p \in P$ w/ $\angle(p, q) \leq 4\theta$

(b) declare no $p \in P$ w/ $\angle(p, q) \leq \theta$

define $h: S^{d-1} \rightarrow \{-1, 1\}^k$ by

$$h_i(x) = \text{sign}(\langle g_i, x \rangle) \text{ w/ } g_i \sim \mathcal{N}^d$$

$$k = \pi(\log(n)/2\theta)$$



d-dimensional sphere

when querying y :

- if there is $x \in P$ w/ $\angle(x, y) \leq \theta$, then it collides w/ probability $\geq 1/\sqrt{n}$

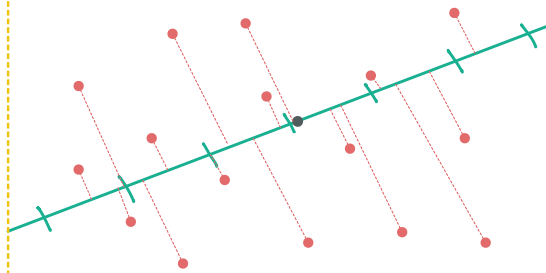
- expect to collide w/ at most one point $x \in P$ w/ $\angle(x, y) > 2\theta$

Lemma let $x, y \in S^{d-1}$
then $g \sim \mathcal{N}^d$

$$P[\text{sign}(\langle g, x \rangle) \neq \text{sign}(\langle g, y \rangle)] = \frac{\angle(x, y)}{\pi}$$

Random line embedding & Buckets

$$h(x) = \lfloor \langle g, x \rangle + d \rfloor \text{ w/ } g \sim \mathcal{N}^d, d \in [0, 1] \text{ unifs. rand.}$$

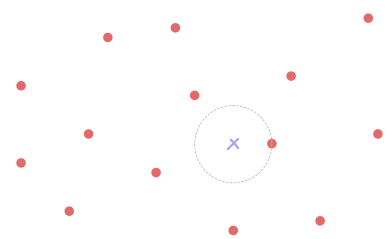


Algorithms

Approximate nearest neighbors

Nearest neighbor search

Input: n points $P \subset \mathbb{R}^d$



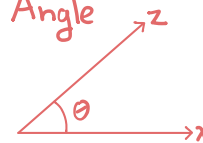
Query: given $q \in \mathbb{R}^d$, find closest point in P to q

Types of distances

- Euclidean distance

$$\|x - z\|_2 = \sqrt{\sum_i (x_i - z_i)^2}$$

- Angle



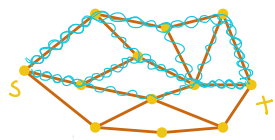
Lecture 10

Probability (and techniques)

LP duality

L_1 -metric embeddings

Max Flow Min Cut



Path Packing LP

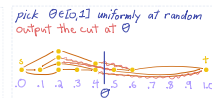
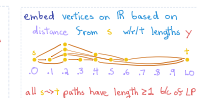
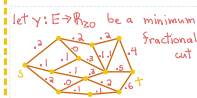
maximize $\sum_{p \in P} x_p$
over $x: P_{st} \rightarrow \mathbb{R}_{\geq 0}$
(all s-t paths)

s.t. $\sum_{p \in P} x_p \leq c_e$ for all edges e
"Ax ≤ c" (capacity of e)

Path Covering LP (i.e. cut)

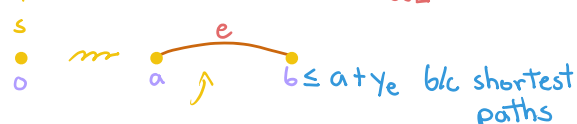
minimize $\sum_{e \in E} c_e y_e$
over $y: E \rightarrow \mathbb{R}_{\geq 0}$
(total capacity of "fractional cut")

s.t. $\sum_{e \in P} y_e \geq 1$ for all $p \in P_{st}$
"A'y ≥ b" (continuous analog of s-t cut)



$E[\text{capacity of cut}]$

$$\stackrel{\text{(linearity of expectation)}}{=} \sum_{e \in E} c_e \Pr[e \text{ cut}] \leq \sum_{e \in E} c_e y_e$$

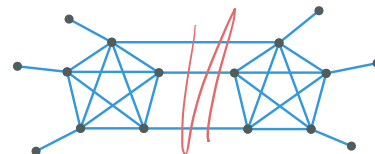


Algorithms

Sparsest Cut

Special case: uniform sparsest cut
w/ uniform demands $b(u,v) = 1 \forall u,v$

$$\text{"sparsity"} = \frac{c(\delta(s))}{|S| |T|}$$



useful primitive for graph partitioning, divide+conquer

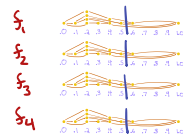
Q: how to round LP to sparse cut?

min $\sum_{e \in E} c_e d(u,v)$ over metrics $d: V \times V \rightarrow \mathbb{R}_{\geq 0}$
s.t. $\sum_{i=1}^k b_i d(s_i, t_i) \geq 1$

Suppose $d(u,v)$ was of the form

$$d(u,v) = \sum_{i=1}^k |f_i(u) - f_i(v)| \text{ for } f_i: V \rightarrow [0,1]^k$$

(assume $\min f_i(a) = 0, \max f_i(a) = 1$)



" L_1 -metric"

Lecture 11

Probability (and techniques)

Algorithms

Other metric-based problems

(sometimes via correspondence between cuts $\leftrightarrow$ metrics)

Buy-at-bulk network design
k-median clustering
metric labeling
group Steiner tree
metrical task systems
vehicle routing
hierarchical placement
distributed k-server
covering Steiner tree



&

these problems are easy for tree metrics

Approximation by tree metrics

Input: undirected $G=(V,E)$ w/ edge weights $w: E \rightarrow \mathbb{R}_{\geq 0}$

(high-level)

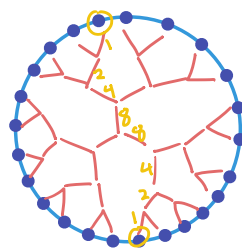
Goal: compute tree T (containing V as nodes)
so that shortest paths on T are close to T

$$d_T(u,v) \geq d_G(u,v)$$

$$\text{For } e=f_u, v \quad \begin{matrix} \text{(distance in } T) \\ \text{(stretch } e) = \frac{d_T(u,v)}{d_G(u,v)} \\ \text{(distance in } G) \end{matrix}$$

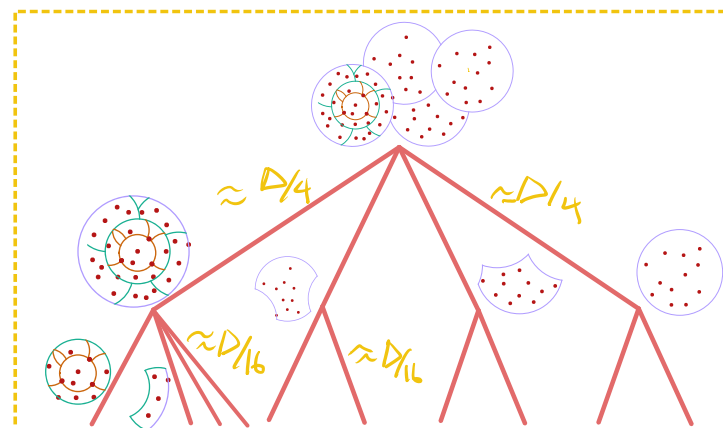
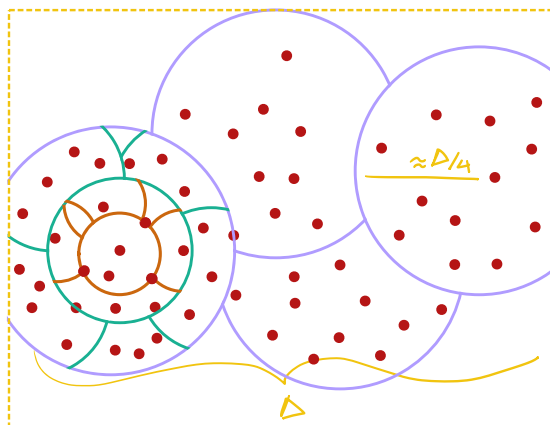
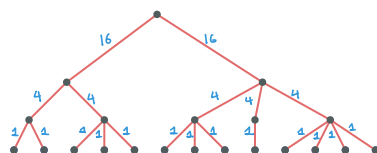
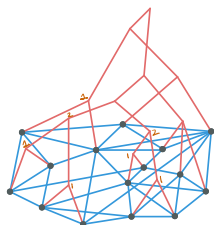
ideally

$$(\text{stretch } e) = \text{polylog}(n) \quad \forall e$$



Randomized Tree metrics

Theorem In randomized polynomial time,
randomized hierarchical dominating tree metric w/
 $E[\text{stretch } e] \leq O(\log n) \quad \forall e \in E$



Lecture 12

Probability (and techniques)

Additive
Chernoff Inequality

Double-sampling

VC dimension

Algorithms

ϵ -samples
and ϵ -nets

Concentration phenomena

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

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

Q: probability that

$X_1 + \dots + X_n$ 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

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

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

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

Additive Chernoff bound

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

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

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

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

(proof at end if time permits)

$X_1, \dots, X_n \in [0,1]$ independent, $\epsilon \in (0,1)$

Multiplicative Chernoff

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

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

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

Additive Chernoff

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

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

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

Lecture 12

Probability (and techniques)

✓ Additive
Chernoff Inequality

Double-sampling

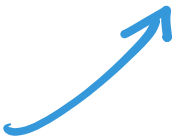
VC dimension

Algorithms

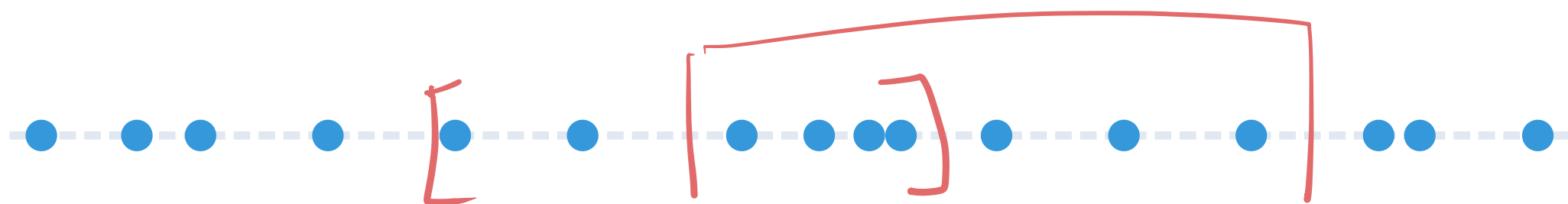
ϵ -samples
and ϵ -nets

Range Space

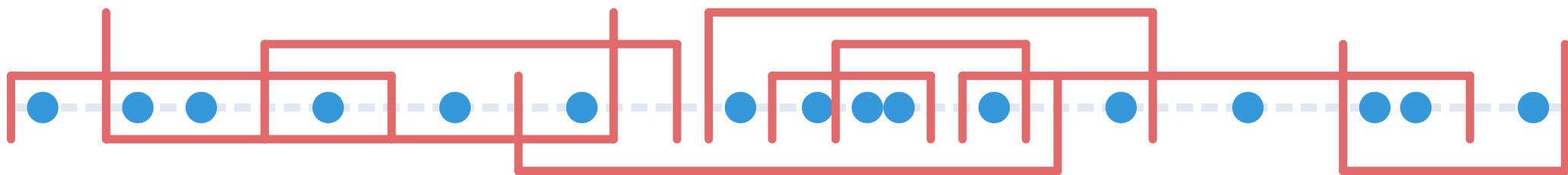
(P, R)

points 

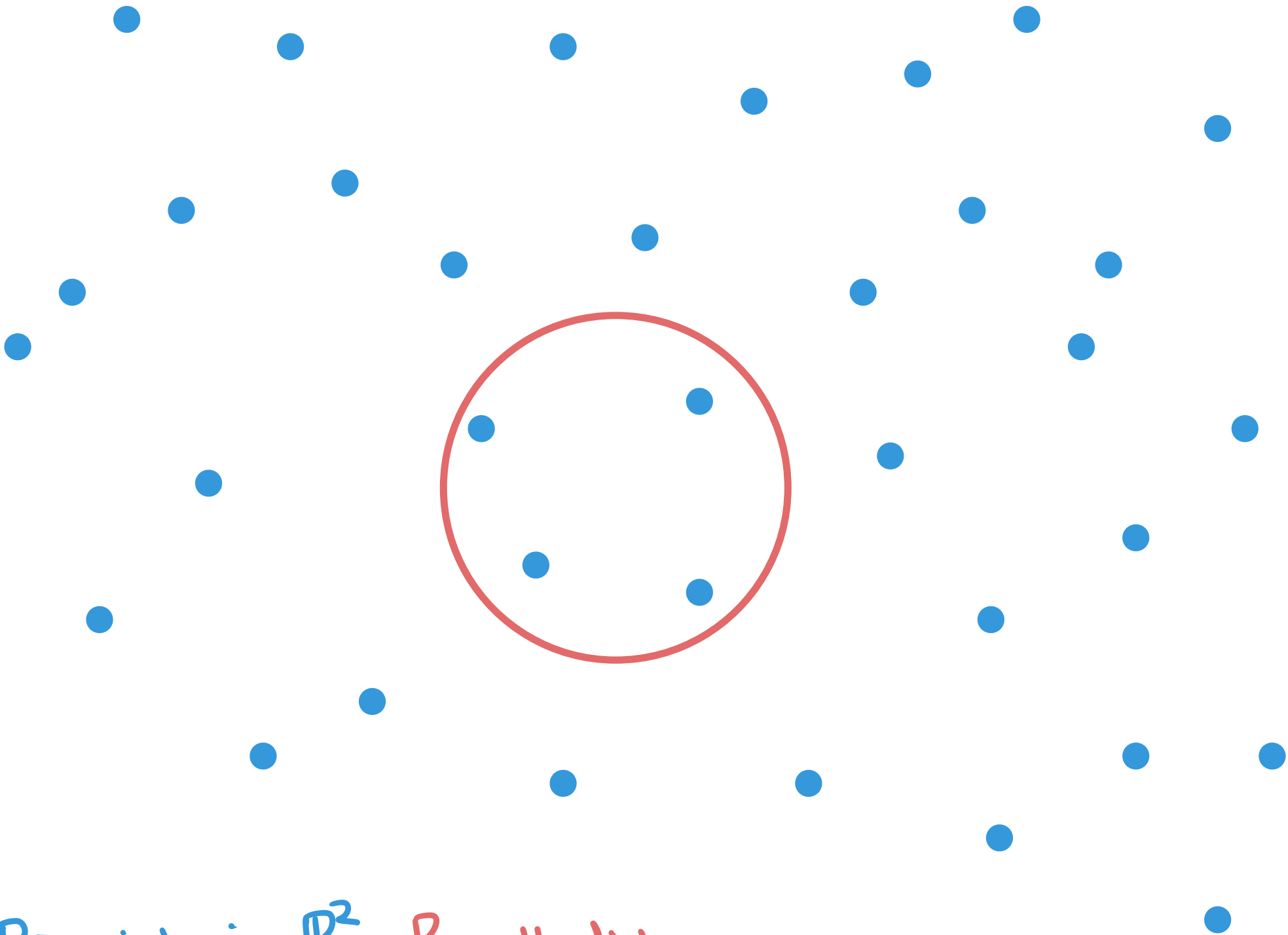
 ranges
(subsets of P)



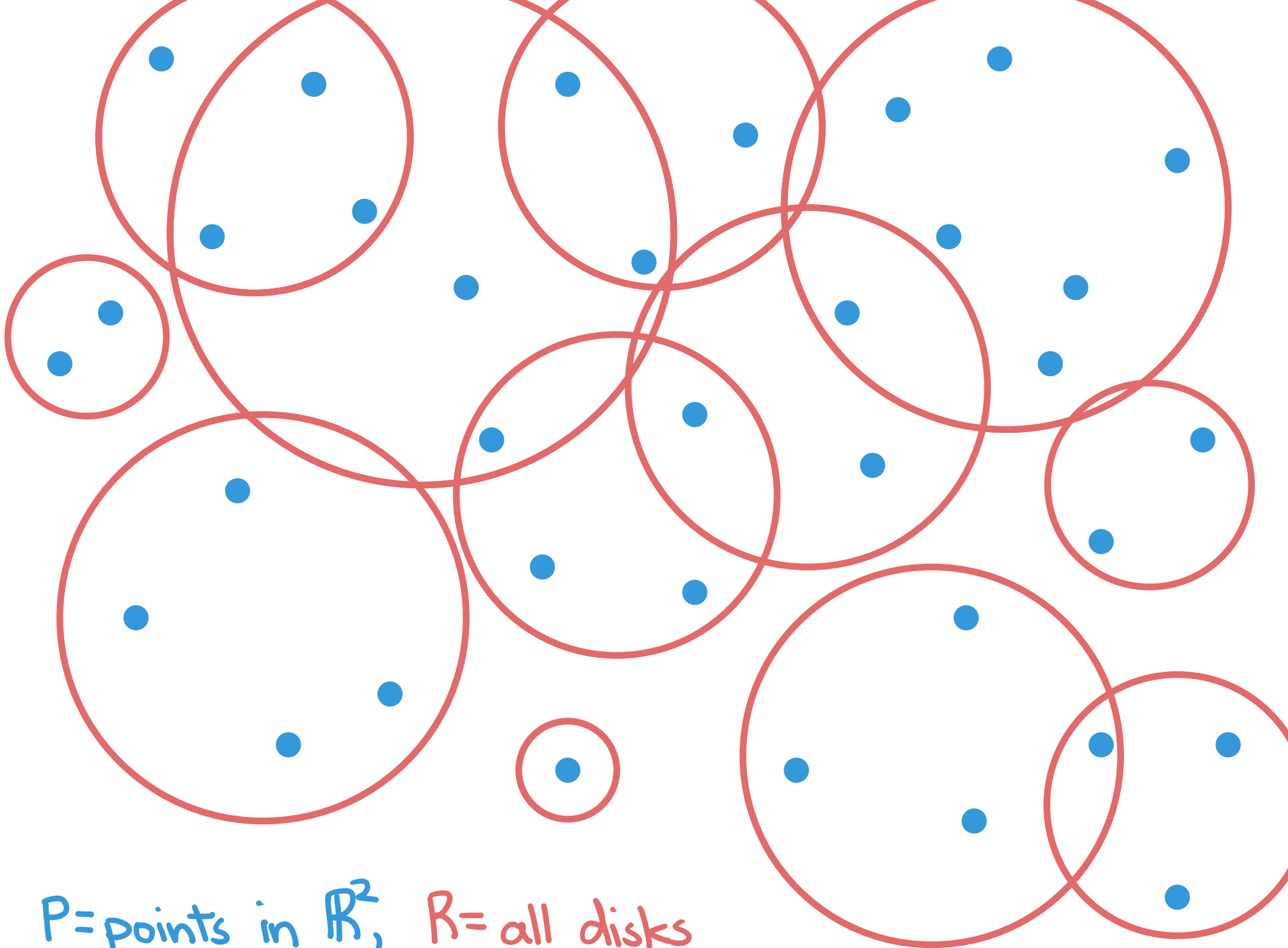
P = points in $\mathbb{R}$, R = intervals



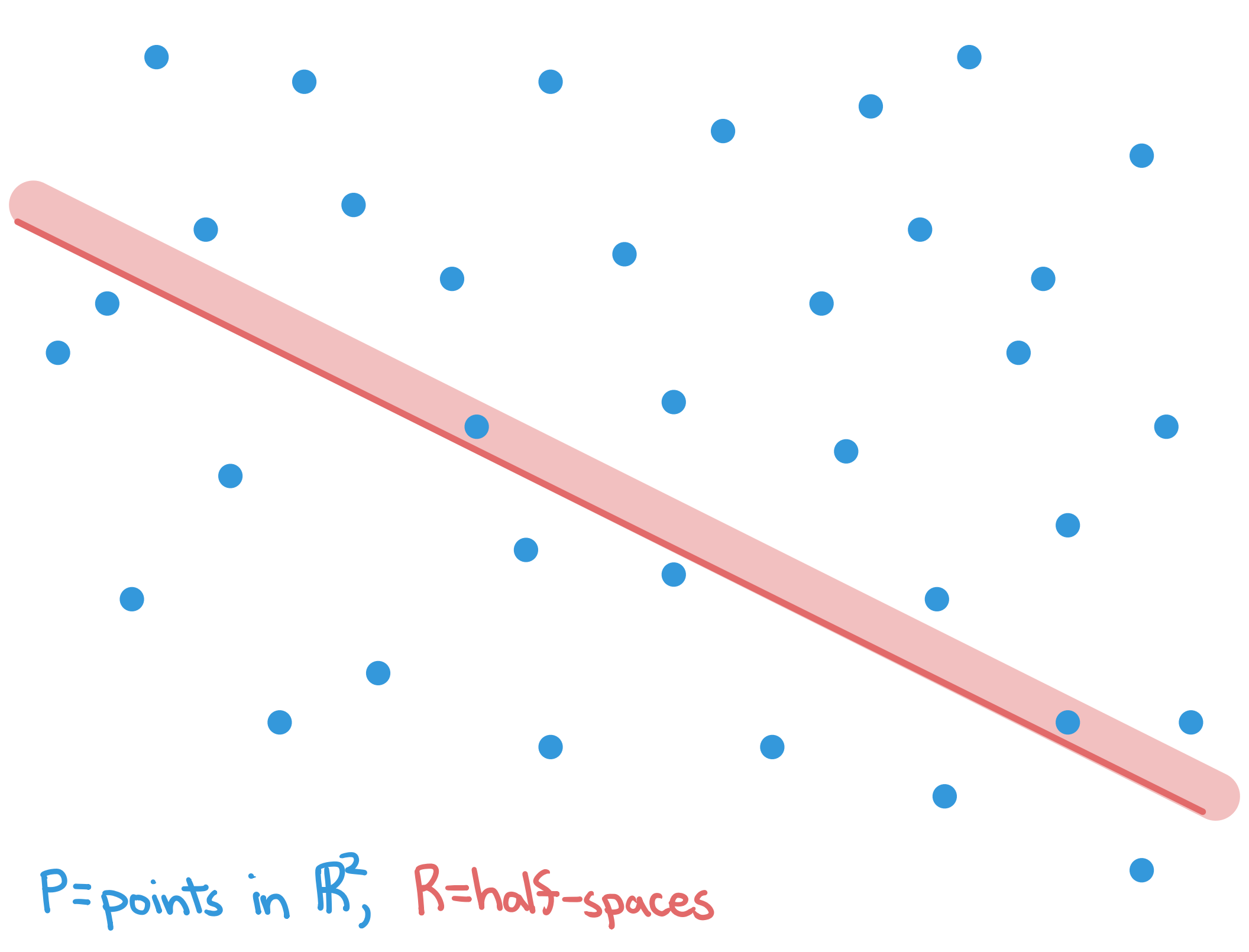
P = points in $\mathbb{R}$, R = intervals

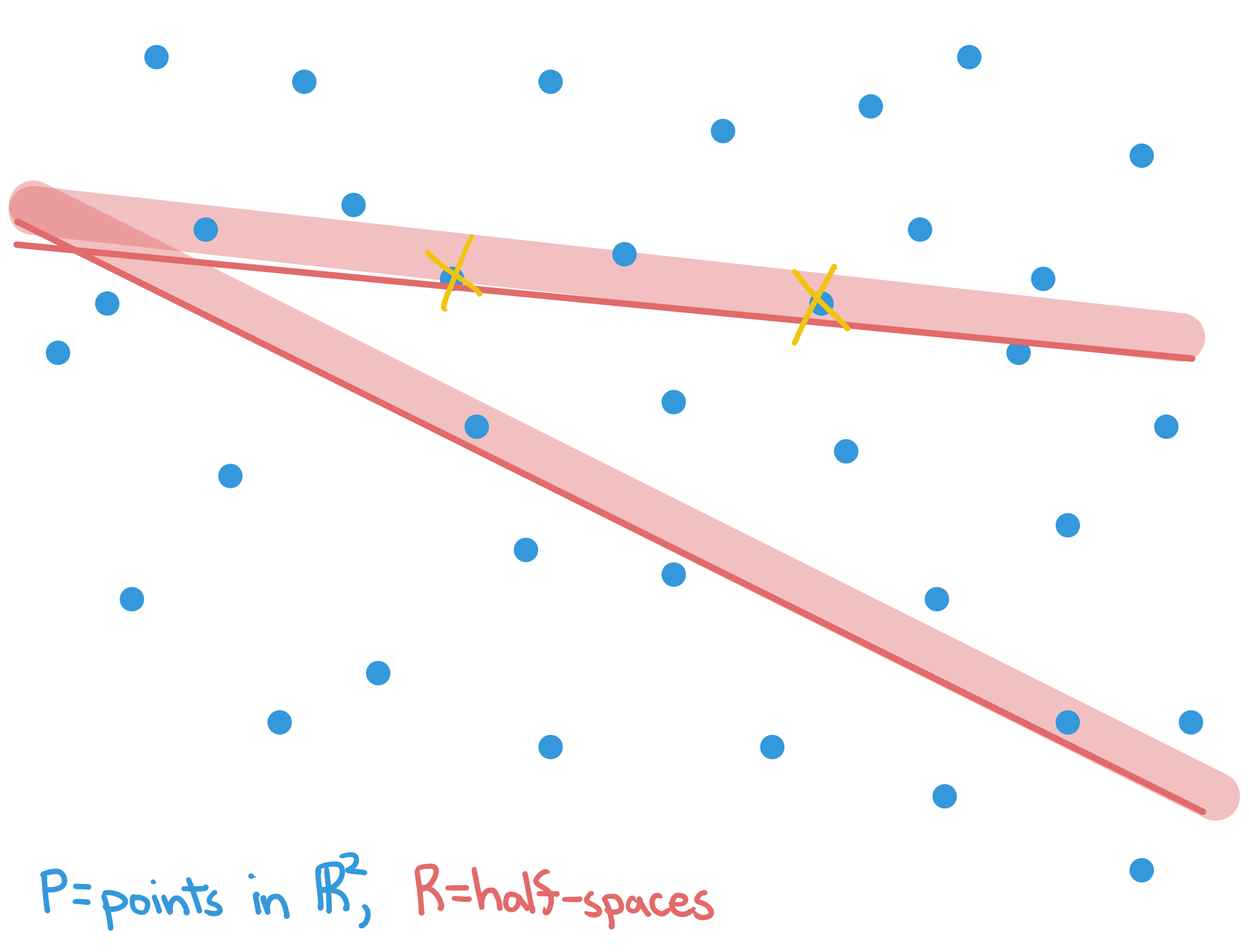


$P = \text{points in } \mathbb{R}^2, R = \text{all disks}$



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





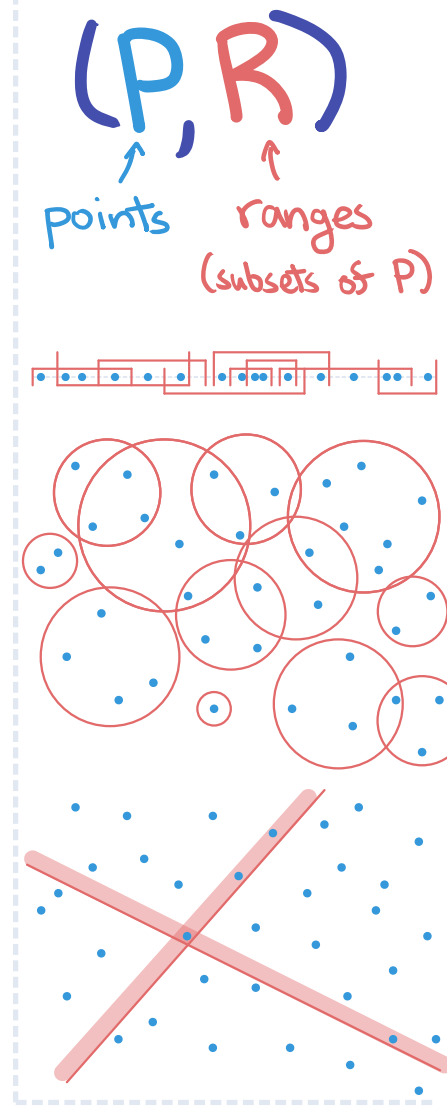
P =points in $\mathbb{R}^2$, R =half-spaces

Typical queries: for $r \in R$:

Q: is r empty?

Q: how many points does r contain?

Q: does r contain $\geq \epsilon$ -fraction of P ?
(for $\epsilon \in (0,1)$)

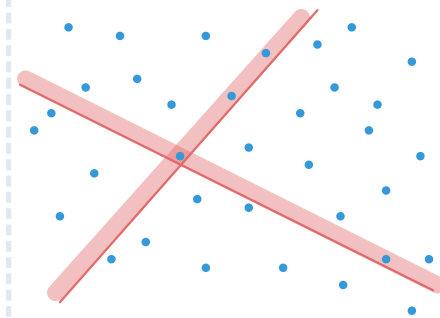
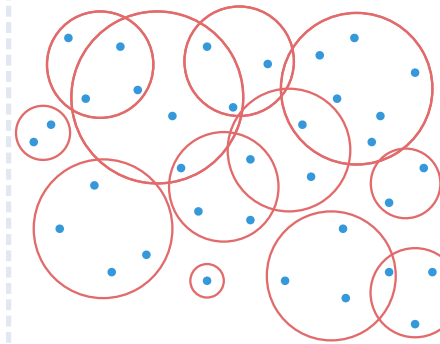


measure $\mu(r) = \frac{|P \cap r|}{|P|}$

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

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

(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 $\geq \epsilon$ -fraction of P ?

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

ϵ -sample

Q is an ϵ -sample if

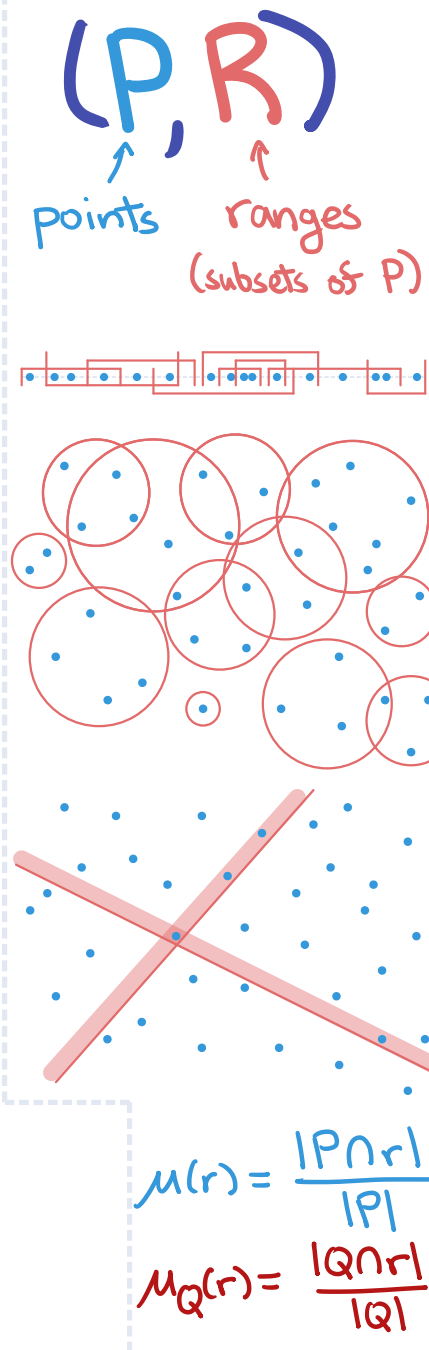
$$|\mu_Q(r) - \mu(r)| \leq \epsilon \quad \forall r \in R$$

ϵ -net

Q is an ϵ -net if

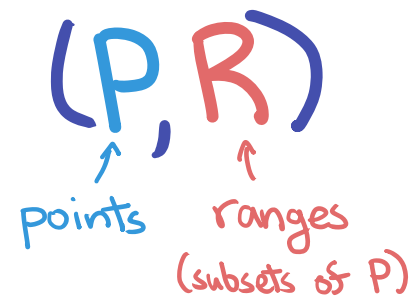
$$\mu_Q(r) > 0 \text{ whenever } \mu(r) \geq \epsilon$$

(ϵ -sample $\Rightarrow$ ϵ -net)



ϵ -sample

$$|\mu_Q(r) - \mu(r)| \leq \epsilon \quad \forall r \in R$$

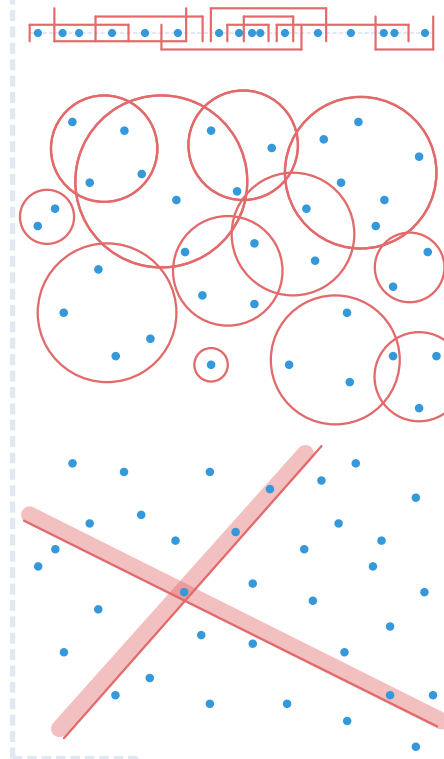


(!) $|Q| < (1-\epsilon)|P|$ is impossible
w/out any assumptions on R

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

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

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



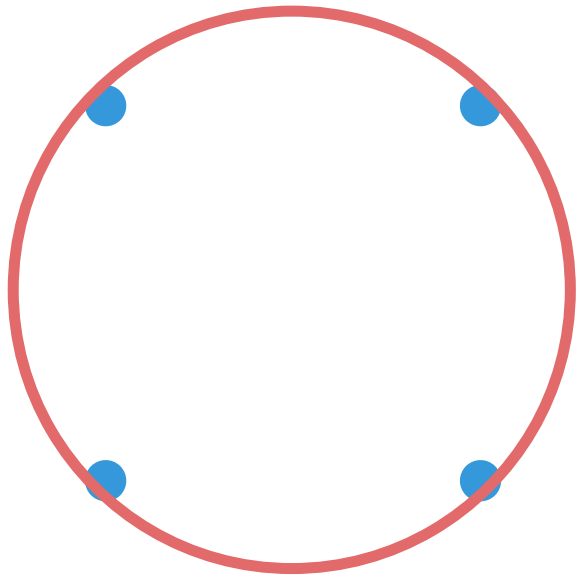
$$\mu(r) = \frac{|P \cap r|}{|P|}$$
$$\mu_Q(r) = \frac{|Q \cap r|}{|Q|}$$

(same for ϵ -nets)

(!) $|Q| < (1-\epsilon)|P|$ 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

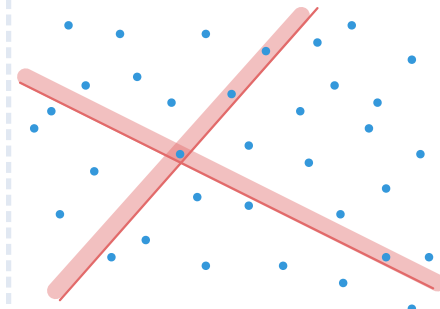
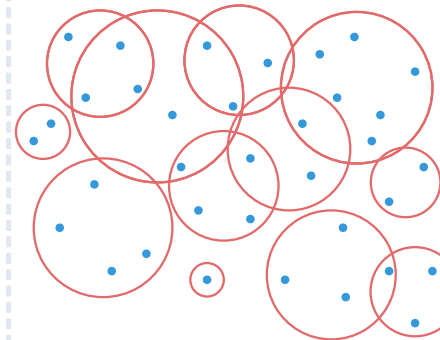


e.g. cannot cover
(only) opposite
corners of a square
w/ a disk

(P, R)

points

ranges
(subsets of P)



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

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

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

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

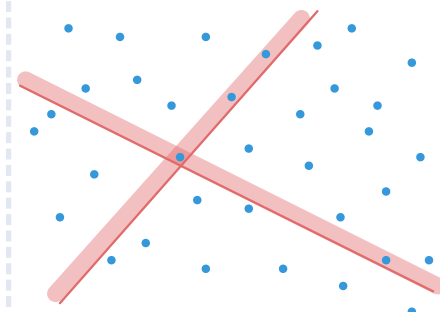
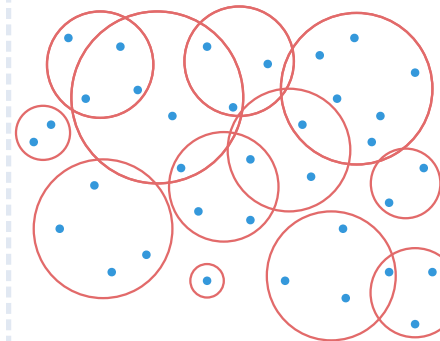
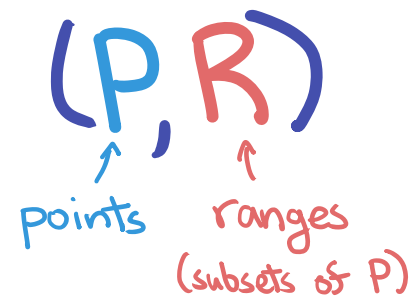
many natural range spaces cannot induce (nearly) all subsets of P

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

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

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

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



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

ϵ -sample $|\mu_Q(r) - \mu(r)| \leq \epsilon \quad \forall r \in R$

let Q sample k points w/ repetition

for range $r \in R$:

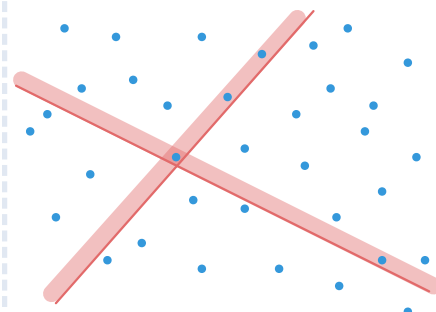
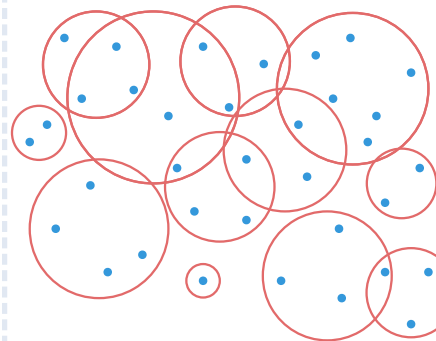
$$\mu_Q(r) = \frac{1}{k}(X_1 + \dots + X_k) \text{ w/}$$

$$X_i = \begin{cases} 1 & \text{if } i\text{th point in } r \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} &P[|\mu_Q(r) - \mu(r)| \geq \epsilon] \\ &= P\left[\left|\frac{1}{k}(X_1 + \dots + X_k) - \mu(r)\right| \geq \epsilon\right] \\ &\leq 2e^{-2\epsilon^2 k} \end{aligned}$$

(P, R)

points $\uparrow$ ranges
(subsets of P)



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

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

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

$$P\left[\frac{X_1 + \dots + X_n}{n} \geq \mu + \epsilon\right], P\left[\frac{X_1 + \dots + X_n}{n} \leq \mu - \epsilon\right] \leq e^{-2\epsilon^2 n} \quad g(n) = \max \{|R \wedge Q| : Q \subseteq P, |Q| \leq n\}$$

ϵ -sample $|\mu_Q(r) - \mu(r)| \leq \epsilon \quad \forall r \in R$

let Q sample k points w/ repetition

✓ $P[|\mu_Q(r) - \mu(r)| \geq \epsilon] \leq 2e^{-2\epsilon^2 k}$ for $r \in R$

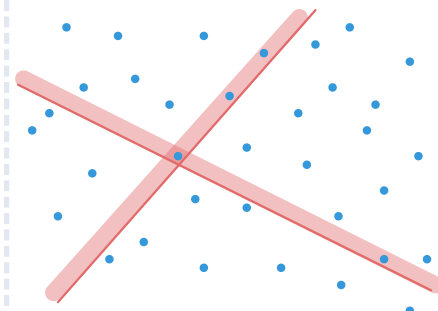
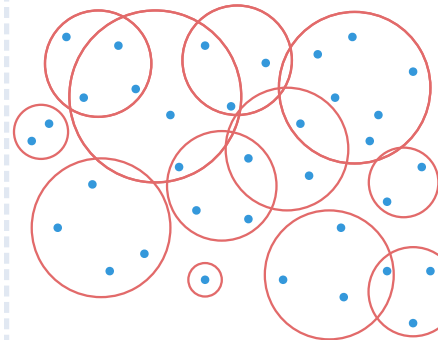
how many r 's?

$$|R \cap P| \leq g(|P|)$$

(P, R)

points

ranges
(subsets of P)



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

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

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

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

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

ϵ -sample $|\mu_Q(r) - \mu(r)| \leq \epsilon \quad \forall r \in R$

let Q sample k points w/ repetition

✓ $P[|\mu_Q(r) - \mu(r)| \geq \epsilon] \leq 2e^{-2\epsilon^2 k}$ for $r \in R$

✓ $|R \cap P| \leq g(|P|)$

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

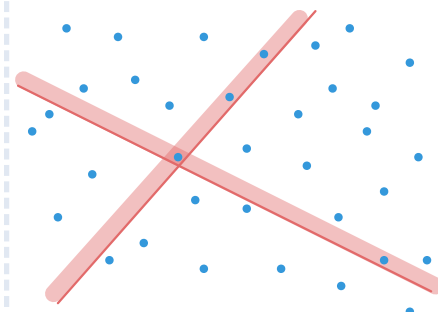
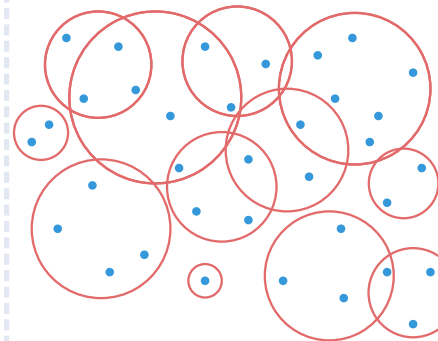
$P_r[Q \text{ not } \epsilon\text{-sample}]$

$\leq \sum_{r \in R \cap P} P[|\mu_Q(r) - \mu(r)| \geq \epsilon]$
(union bound)

$\leq g(|P|) e^{-\epsilon^2 k} \leq \frac{1}{2}$

(P, R)

points $\uparrow$ ranges
(subsets of P)



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

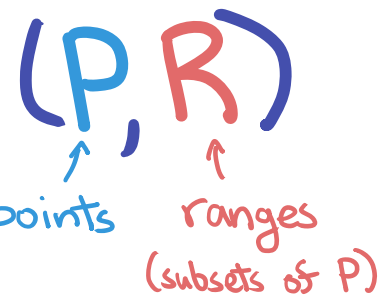
$\mu_Q(r) = \frac{|Q \cap r|}{|Q|}$

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

$P[\frac{X_1 + \dots + X_n}{n} \geq \mu + \epsilon], P[\frac{X_1 + \dots + X_n}{n} \leq \mu - \epsilon] \leq e^{-2\epsilon^2 n}$

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

ϵ -sample $|\mu_Q(r) - \mu(r)| \leq \epsilon \quad \forall r \in R$



let Q sample k points w/ repetition

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

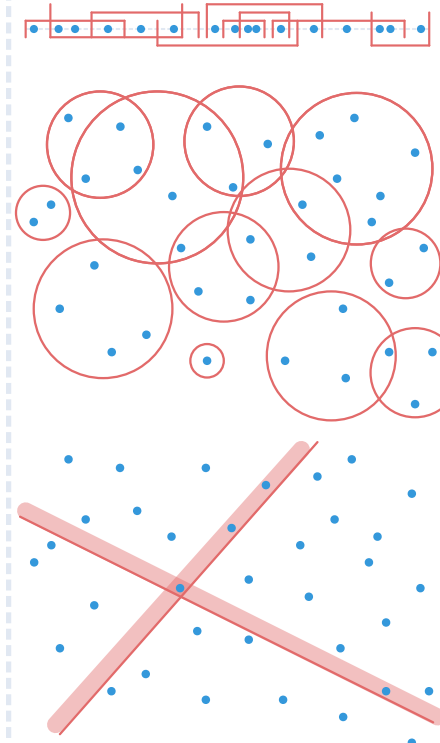
($= O\left(\frac{\ln(|P|)}{\epsilon^2}\right)$ for $g(n) = n^{O(1)}$)

(w/ constant probability)

Also (exercise)

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

(w/ constant probability)



Better?

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

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

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

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

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

ϵ -sample $|\mu_Q(r) - \mu(r)| \leq \epsilon \quad \forall r \in R$

let Q sample k points w/ repetition

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

ϵ -sample theorem

[VC71]

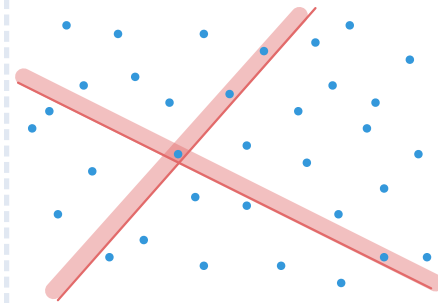
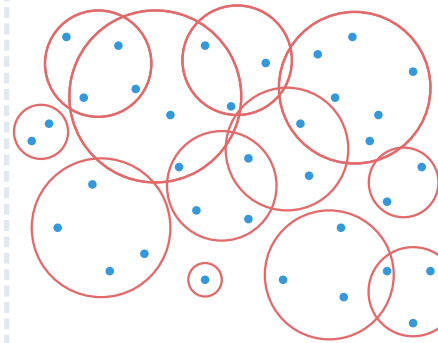
$$k \geq O\left(\frac{\ln(g(2k))}{\epsilon^2}\right) \Rightarrow \epsilon\text{-sample} \quad (\text{w/ constant probability})$$

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

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

↑ independent of $|P|$ (!)

(P, R)
 ↑ points ↑ ranges
 (subsets of P)



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

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

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

$$P\left[\frac{X_1 + \dots + X_n}{n} \geq \mu + \epsilon\right], P\left[\frac{X_1 + \dots + X_n}{n} \leq \mu - \epsilon\right] \leq e^{-2\epsilon^2 n} \quad g(n) = \max \{|R \wedge Q| : Q \subseteq P, |Q| \leq n\}$$

ϵ -sample $|\mu_Q(r) - \mu(r)| \leq \epsilon \quad \forall r \in R$

let Q sample k points w/ repetition

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

(P, R)

points $\uparrow$ ranges
(subsets of P)



ϵ -net theorem

[HW87]

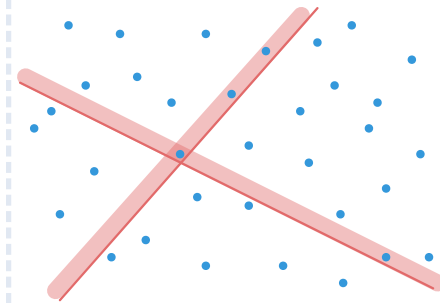
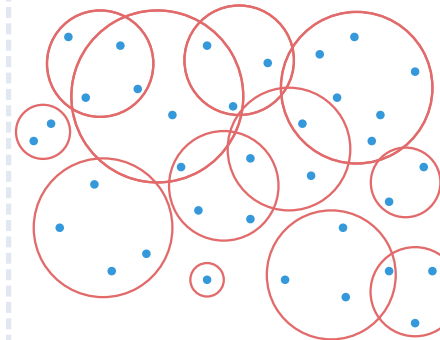
$$k \geq O\left(\frac{\ln(g(2k))}{\epsilon}\right) \Rightarrow \epsilon\text{-net}$$

(w/ constant probability)

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

$$k \geq O\left(\frac{d \ln(d/\epsilon)}{\epsilon}\right) \Rightarrow \epsilon\text{-net}$$

$\uparrow$ independent of $|P|$ (!)



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

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

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

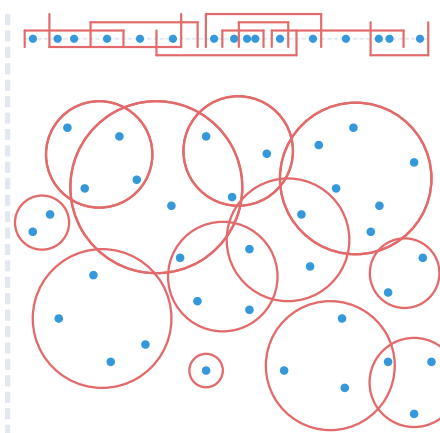
$$P\left[\frac{X_1 + \dots + X_n}{n} \geq \mu + \epsilon\right], P\left[\frac{X_1 + \dots + X_n}{n} \leq \mu - \epsilon\right] \leq e^{-2\epsilon^2 n} \quad g(n) = \max \{|R \wedge Q| : Q \subseteq P, |Q| \leq n\}$$

ϵ -sample theorem

[VC71]

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

(w/ constant probability)



Proof

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

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

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

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

$$\epsilon\text{-sample } |\mu_Q(r) - \mu(r)| \leq \epsilon \quad \forall r$$

$$\epsilon\text{-net } \mu_Q(r) > 0 \quad \forall \mu(r) \geq \epsilon$$

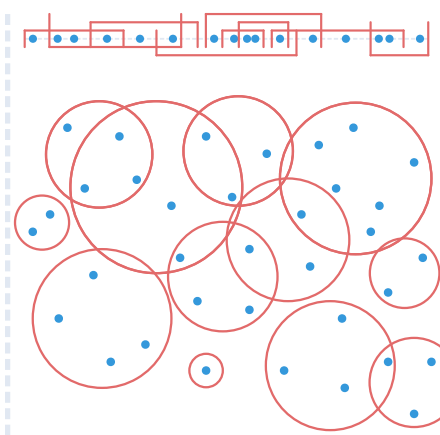
$$\left\{ \begin{array}{l} P\left[\frac{X_1 + \dots + X_n}{n} \geq \mu + \epsilon\right] \\ P\left[\frac{X_1 + \dots + X_n}{n} \leq \mu - \epsilon\right] \end{array} \right\} \leq e^{-2\epsilon^2 n}$$

ϵ -sample theorem

[VC71]

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

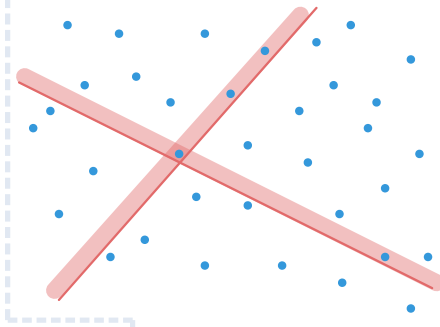
(w/ constant probability)



Proof let Q sample k points (w/ repetition)

define event

$$A = \text{(event that)} |\mu_Q(r) - \mu(r)| > \epsilon \text{ for some } r$$



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

goal: prove $P[A] \leq \frac{1}{2}$

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

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

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

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

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

let Q_1 sample k points
(w/ repetition)

$$\epsilon\text{-sample theorem}$$

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

define event

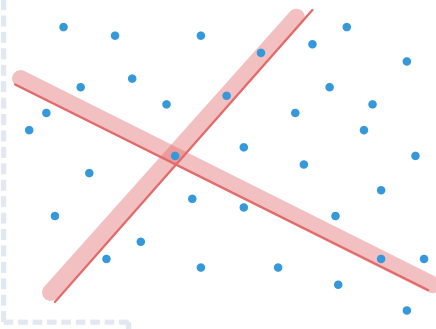
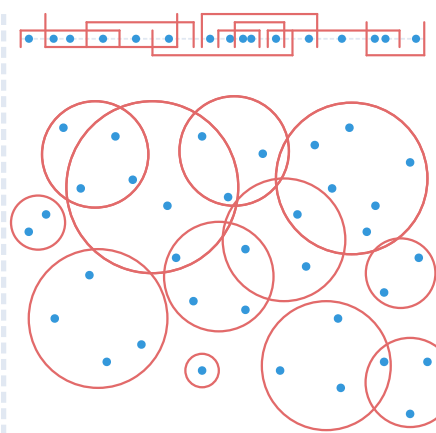
$$A = \text{(event that)} |\mu_1(r) - \mu(r)| > \epsilon \text{ for some } r$$

(!) let Q_2 sample k points

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

$$C = \text{(event that)} |\mu_1(r) - \mu(r)| > \epsilon \text{ and}$$

$$|\mu_2(r) - \mu_1(r)| > \frac{\epsilon}{2} \text{ for some } r$$



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

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

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

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

$$\epsilon\text{-sample } |\mu_Q(r) - \mu(r)| \leq \epsilon \quad \forall r$$

$$\epsilon\text{-net } \mu_Q(r) > 0 \quad \forall \mu(r) \geq \epsilon$$

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

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

$$\epsilon\text{-sample theorem}$$

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

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

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

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

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

$$P[B] \geq P[C] = P[A, C] = \underbrace{P[C|A]}_{\geq 1/2} \cdot P[A]$$

$A \Rightarrow r$ s.t. $|\mu_1(r) - \mu(r)| > \epsilon$

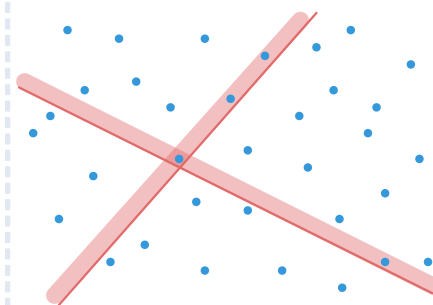
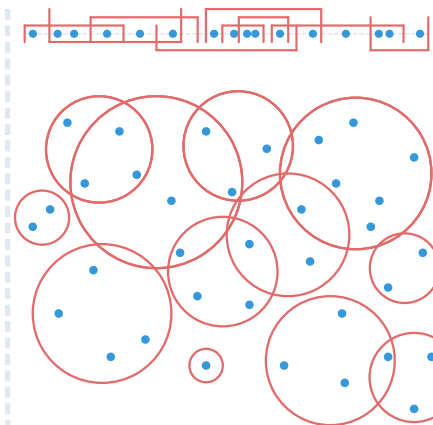
$A \subseteq C$

Chernoff

$$P[|\mu_2(r) - \mu(r)| > \epsilon/2] \leq \frac{1}{2}$$

$$|\mu_1(r) - \mu_2(r)| \geq |\mu_1(r) - \mu(r)| - |\mu_2(r) - \mu(r)|$$

$$> \epsilon - \epsilon/2 = \epsilon/2$$



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

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

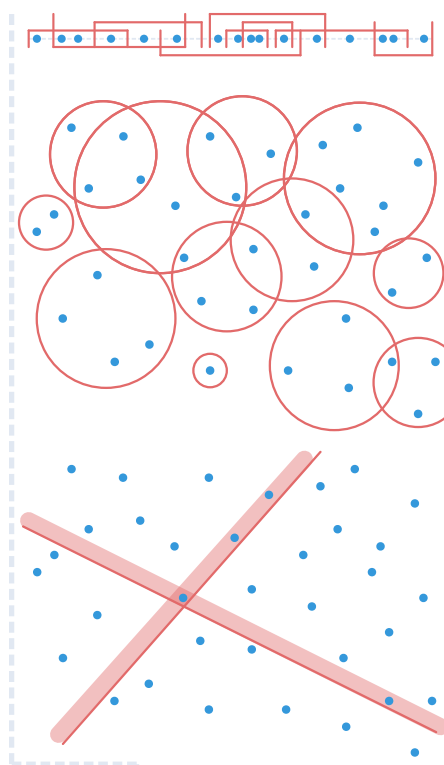
$$R \cap Q = \{r \cap Q : r \in R\}$$

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

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

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

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



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

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

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

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

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

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

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

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

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

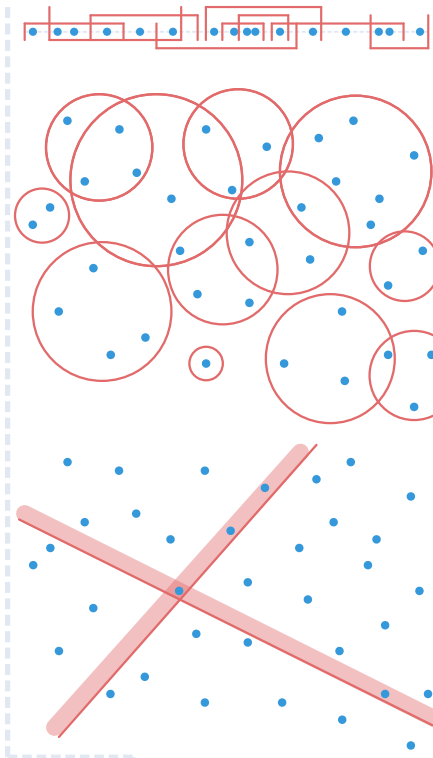
$B = (\text{event that}) |\mu_2(r) - \mu(r)| > \frac{\epsilon}{2}$ for some r

$$\checkmark P[A] \leq 2P[B]$$

goal: prove $P[B] \leq \frac{1}{4}$

ϵ -sample theorem

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



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

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

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

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

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

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

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

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

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

$B = (\text{event that}) |\mu_2(r) - \mu(r)| > \frac{\epsilon}{2}$ for some r

$$\checkmark P[A] \leq 2P[B]$$

Suppose we generate Q_1, Q_2 as follows:

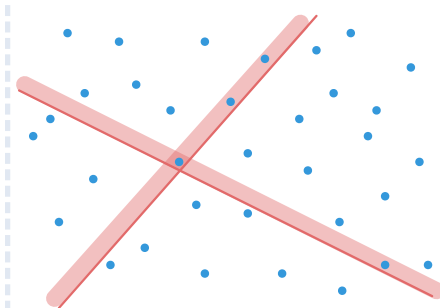
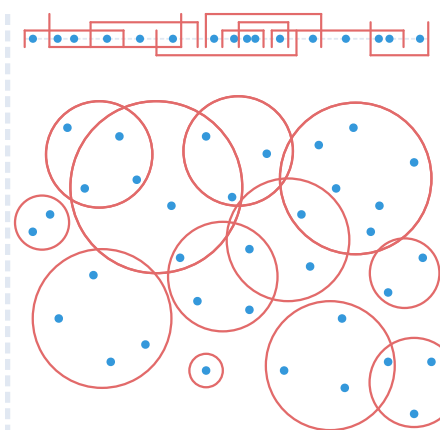
① sample $2k$ points Q_0

② randomly partition Q_0 into halves Q_1, Q_2

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

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

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



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

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

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

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

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

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

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

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

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

✓ $P[A] \leq 2P[B]$

① sample $2k$ points Q_0

② randomly partition Q_0 into halves Q_1, Q_2

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

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

Additive Chernoff:

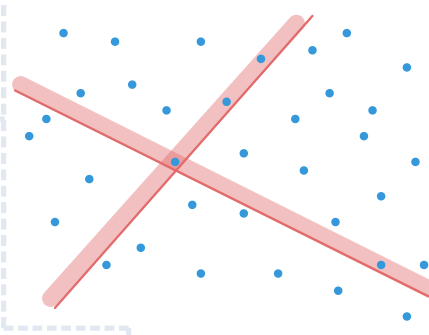
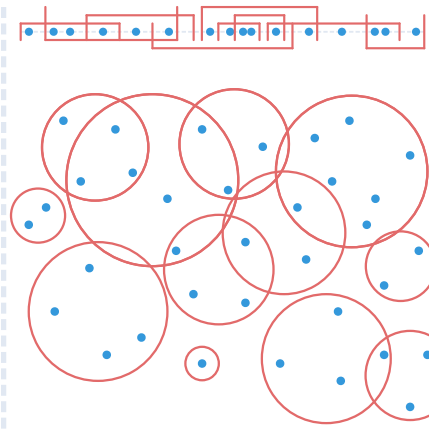
$$\left. \begin{aligned} P[|\mu_1(r) - \mu_0(r)| \geq \frac{\varepsilon}{4}] \\ P[|\mu_2(r) - \mu_0(r)| \geq \frac{\varepsilon}{4}] \end{aligned} \right\} \leq \frac{1}{8g(2k)}$$

$$\forall r \in R \cap Q_0$$

ε -sample theorem

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

↑



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

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

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

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

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

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

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

Union bound $\Rightarrow$

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

$$\forall r \in R \cap Q_0 \text{ w/ prob } \geq 3/4$$

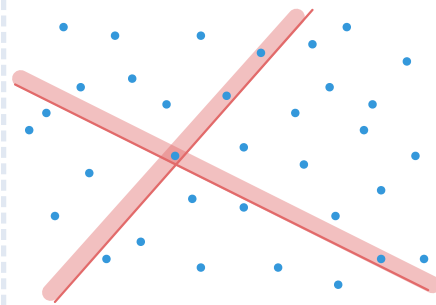
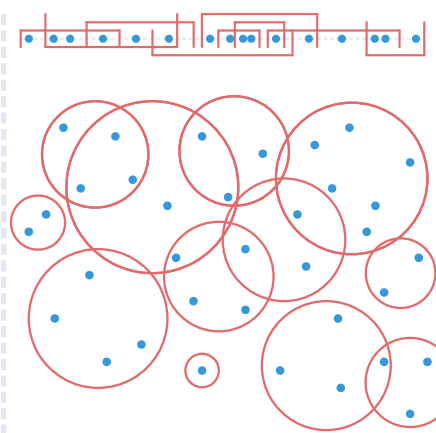
$$\forall r \in R \cap Q_0,$$

$$|\mu_1(r) - \mu_2(r)| \leq |\mu_1(r) - \mu_0(r)| + |\mu_2(r) - \mu_0(r)| \leq \varepsilon/2$$

$$r' \in R \rightarrow r = r' \cap Q_0 \rightarrow$$

same $\forall r \in R$

$$\mu_1(r') = \mu_1(r)$$
$$\mu_2(r') = \mu_2(r)$$



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

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

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

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

$$\varepsilon\text{-sample } |\mu_Q(r) - \mu(r)| \leq \varepsilon \quad \forall r$$

$$\varepsilon\text{-net } \mu_Q(r) > 0 \quad \forall \mu(r) \geq \varepsilon$$

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



ϵ -sample theorem

[VC71]

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

(w/ constant probability)

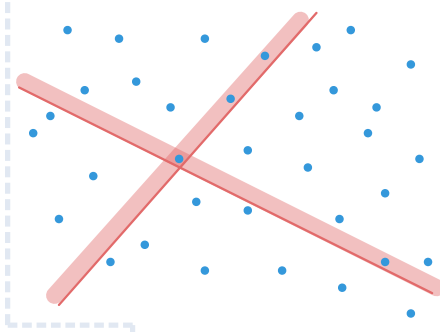
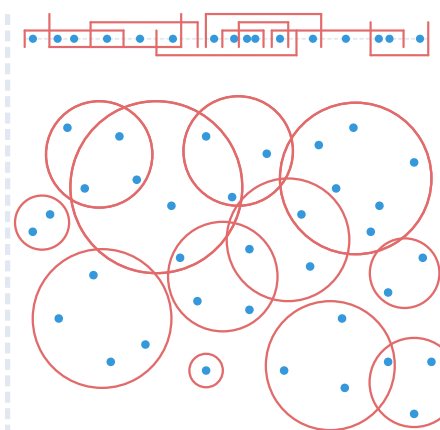
similar,
exercise

ϵ -net theorem

[HW87]

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

(w/ constant probability)



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

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

next question:

how to bound $g(n)$?

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

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

$$\epsilon\text{-sample } |\mu_Q(r) - \mu(r)| \leq \epsilon \quad \forall r$$

$$\epsilon\text{-net } \mu_Q(r) > 0 \quad \forall \mu(r) \geq \epsilon$$

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

Lecture 12

Probability (and techniques)

✓ Additive
Chernoff Inequality

✓ Double-sampling

VC dimension

Algorithms

✓ ϵ -samples
and ϵ -nets

VC dimension

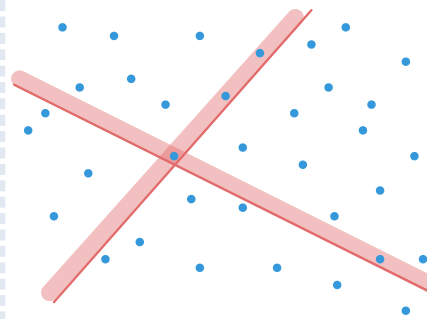
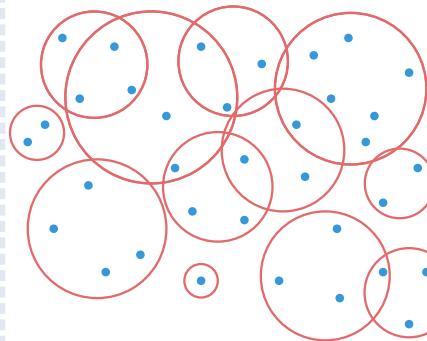
for QCP, R shatters Q if

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

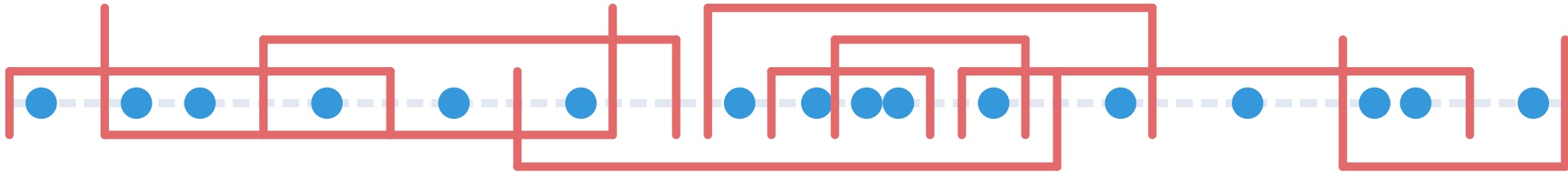
VC dimension of $(P, R) =$

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

(P, R)
points ranges
(subsets of P)



for QCP, R shatters Q if $R \cap Q = \text{all } 2^{|Q|} \text{ subsets of } Q$
VC dimension of $(P, R) = \max |Q| \text{ s.t. } R \text{ shatters } Q$

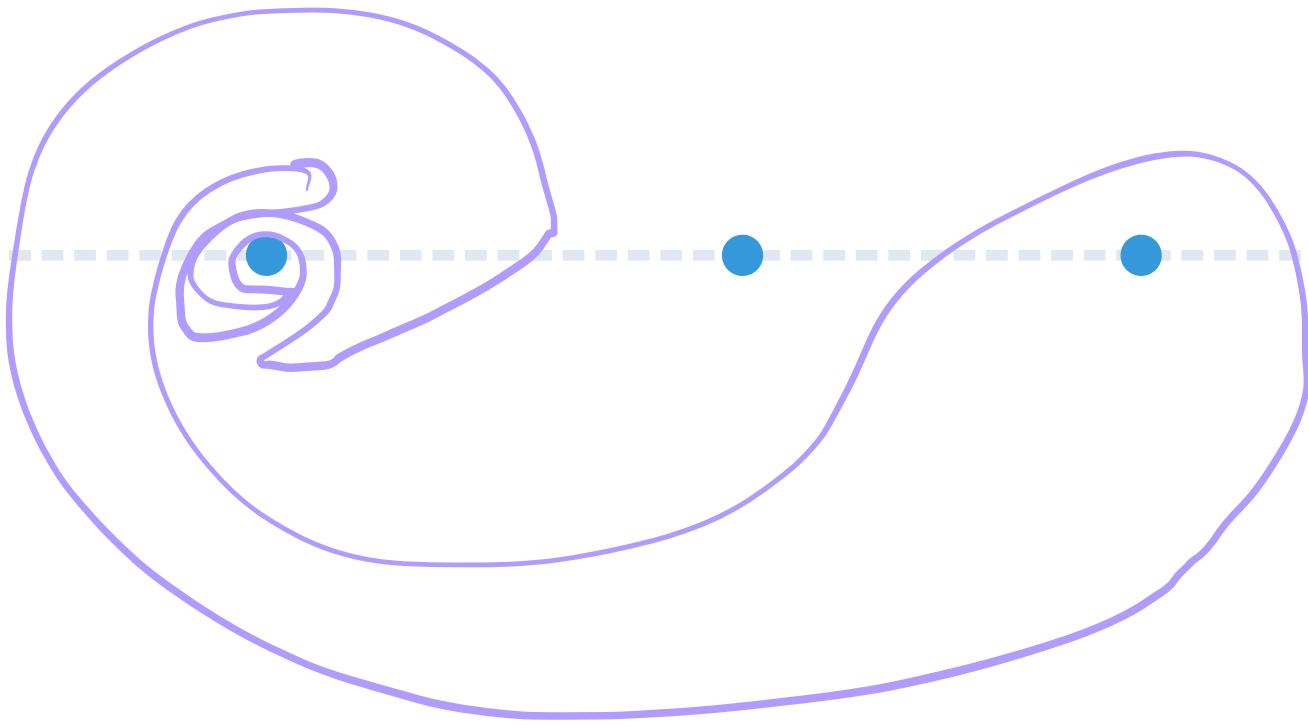


$P = \text{points in } \mathbb{R}, \quad R = \text{intervals}$

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

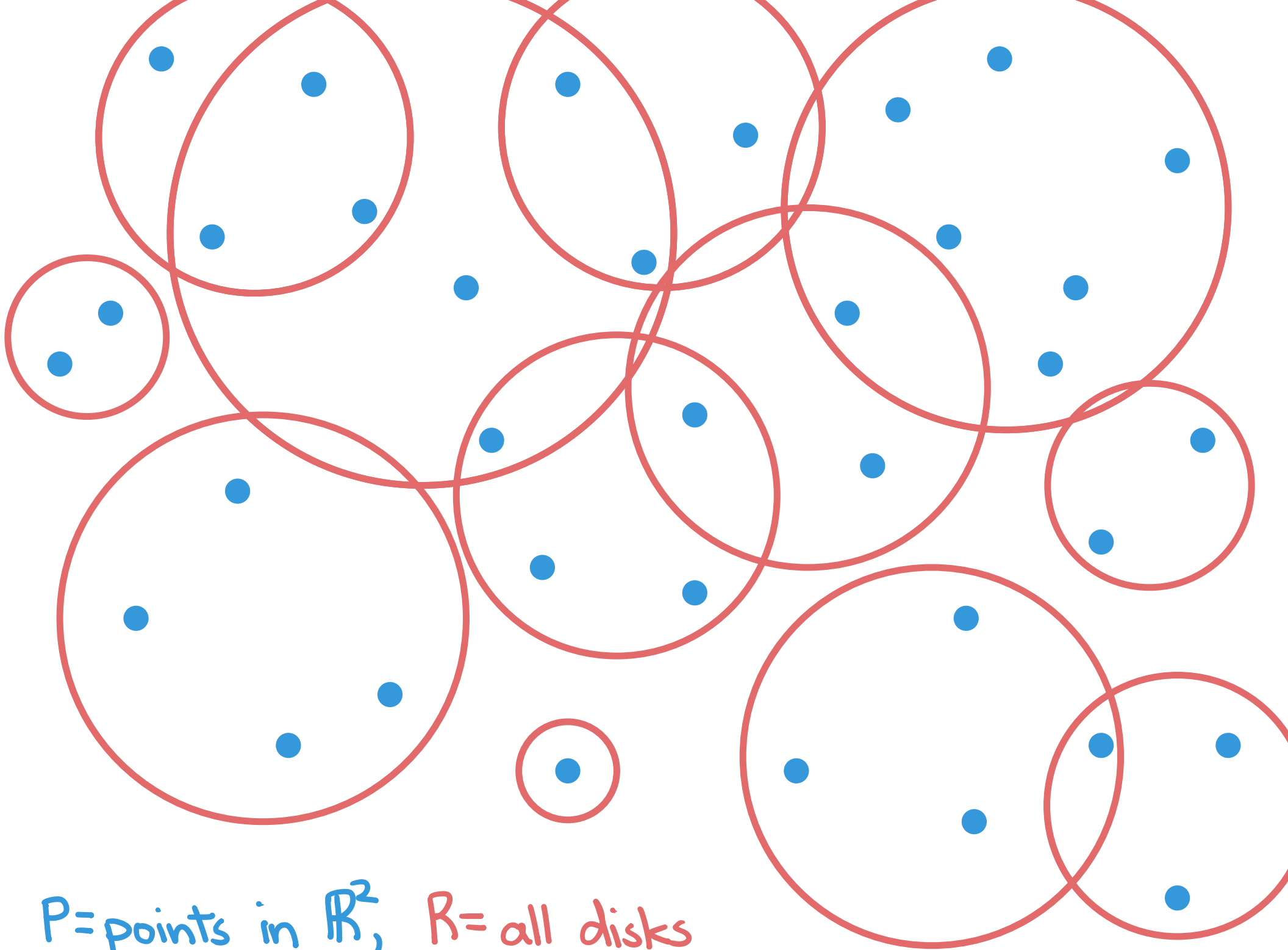
VC dimension of $(P, R) = \max |Q| \text{ s.t. } R \text{ shatters } Q$

VC-dim=3



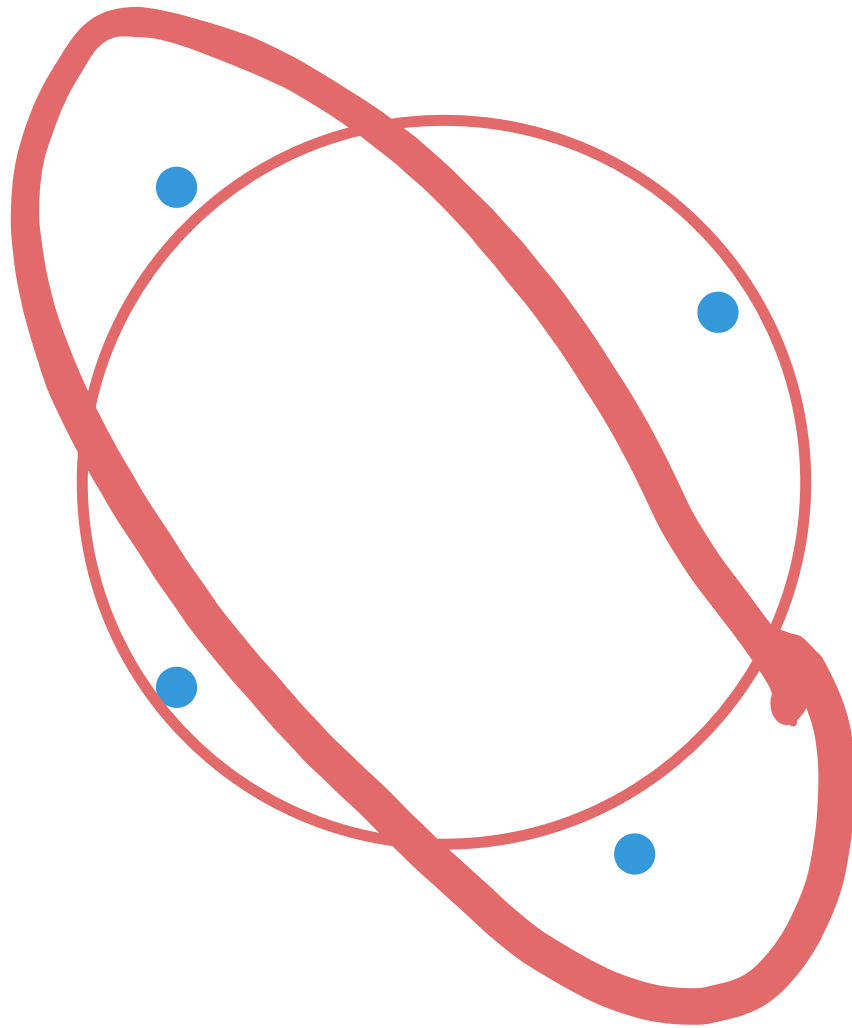
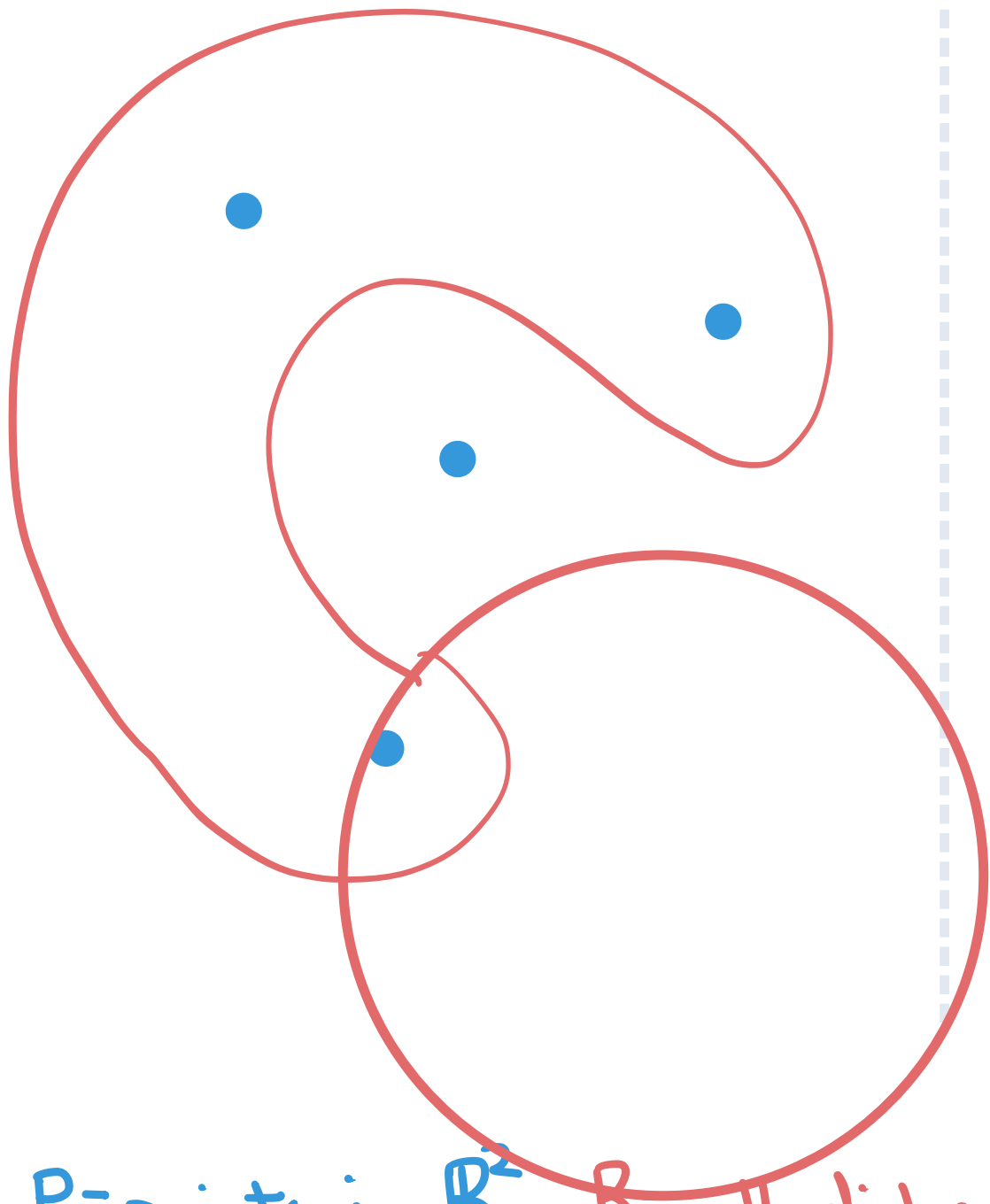
$P = \text{points in } \mathbb{R}$, $R = \text{intervals}$



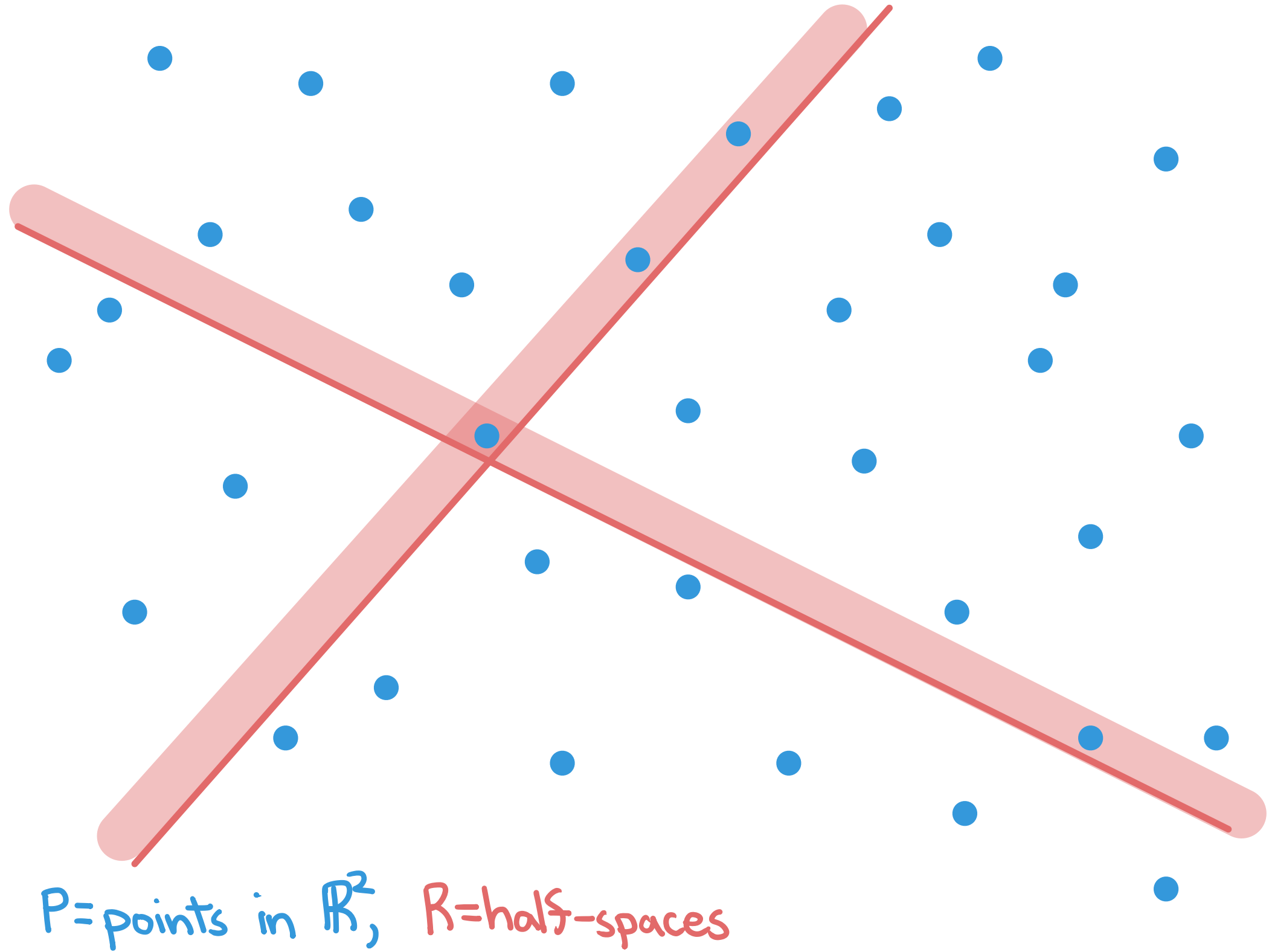


P = points in $\mathbb{R}^2$, R = all disks

VC-dim = 4



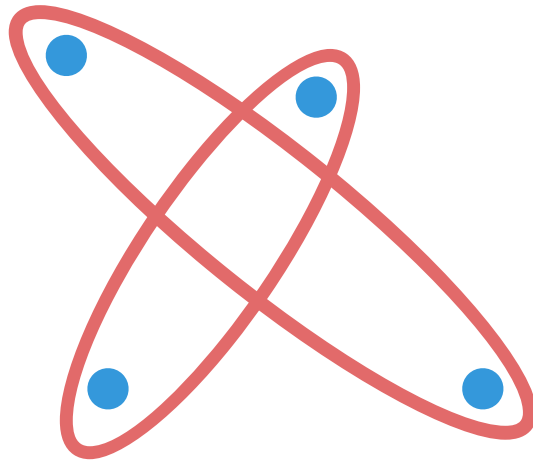
P = points in $\mathbb{R}^2$, R = all disks



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

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

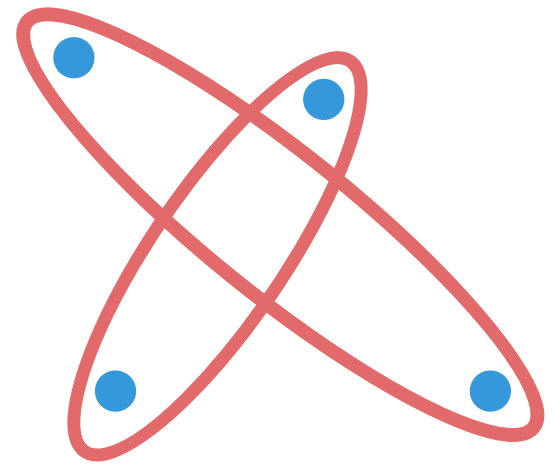
$$\text{convex-hull}(S_1) \cap \text{convex-hull}(S_2) \neq \emptyset$$



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

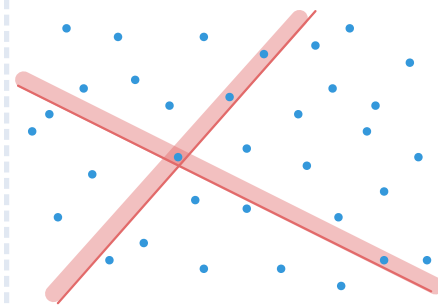
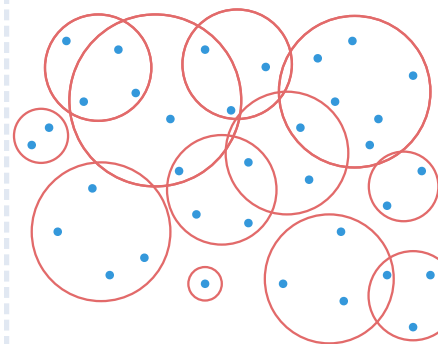
$$g(n) \leq \sum_{i=0}^d \binom{n}{i} \leq n^d$$

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

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

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

(P, R)
points ranges
(subsets of P)



For QCP, R shatters Q if
 $R \cap Q = \text{all subsets of } Q$

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

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

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

Proof let $f(d, n) =$



base cases:

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

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

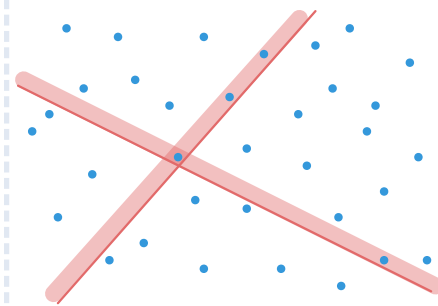
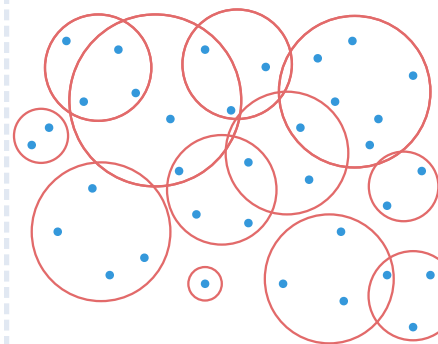
case $n, d > 0$: fix Q w/ $n=|Q|$, $p \in Q$.

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

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

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

(P, R)
 points ranges
 (subsets of P)



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

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

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

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

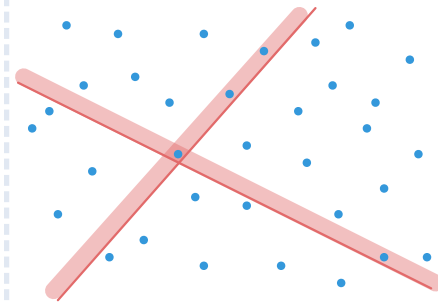
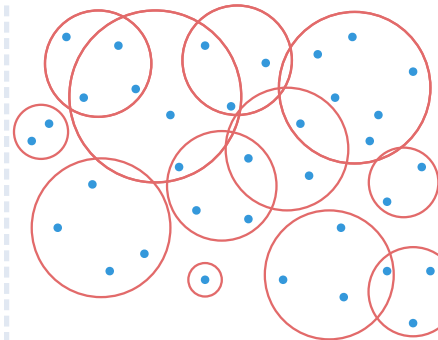
$R_p = \{r - p : r \in R, r \cup \{p\} \in R, r \setminus \{p\} \in R\}$ $R^p = \{r - p : r \in R\}$

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



(P, R)

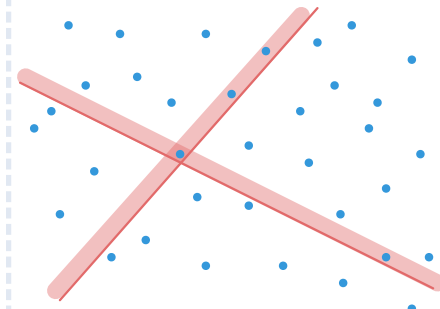
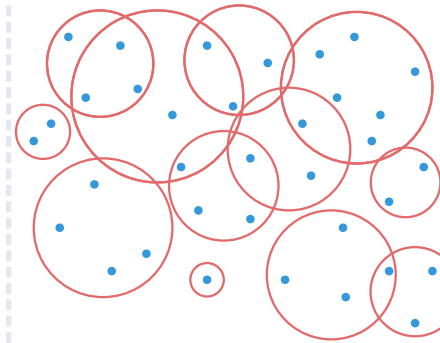
points ranges
(subsets of P)



For QCP, R shatters Q if
 $R \cap Q = \text{all subsets of } Q$

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

(P, R)
 ↑ ↑
 points ranges
 (subsets of P)



For QCP, R shatters Q if
 $R \cap Q = \text{all subsets of } Q$

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

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

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

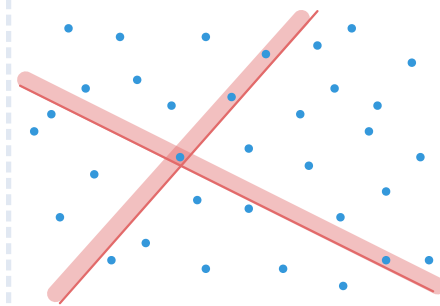
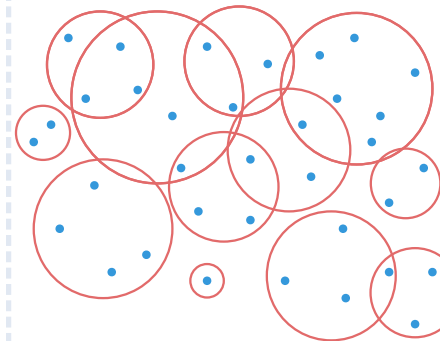
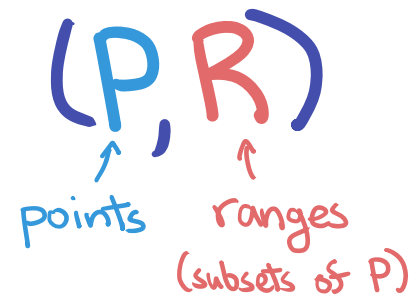
$R_p = \{r - p : r \in R, r \cup \{p\} \in R, r \setminus \{p\} \in R\}$ $R^p = \{r - p : r \in R\}$

✓ $|R_p| + |R^p| = |R|$

$(Q - p, R_p)$ has VC-dim $d-1$

$(Q - p, R^p)$ has VC-dim d

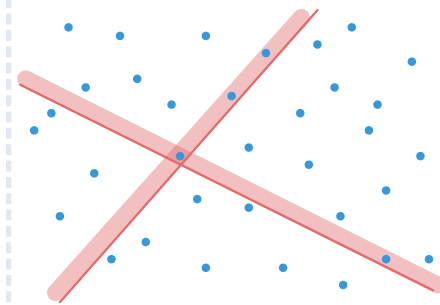
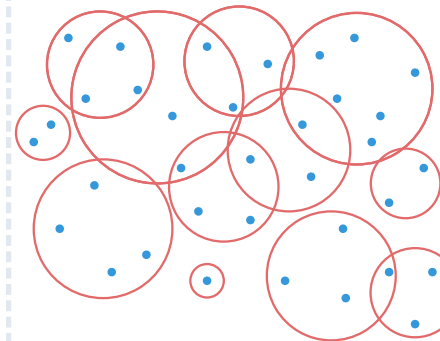
$|R| =$



For QCP, R shatters Q if
 $R \cap Q = \text{all subsets of } Q$

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

(P, R)
 ↑ ↑
 points ranges
 (subsets of P)



For QCP, R shatters Q if
 $R \cap Q = \text{all subsets of } Q$

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

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

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

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

✓ $|R_p| + |R^P| = |R|$

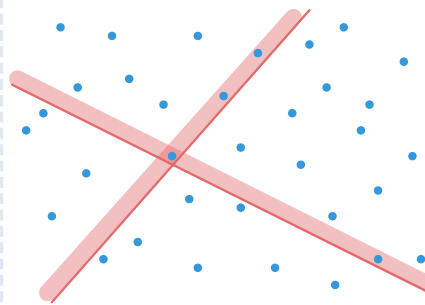
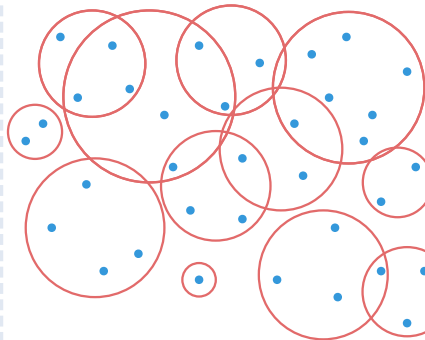
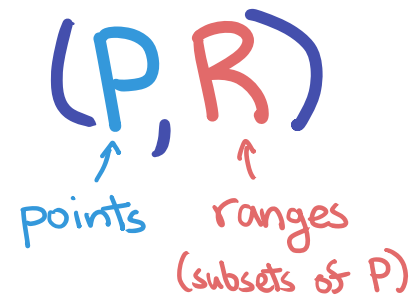
$(Q - p, R_p)$ has VC-dim $d-1$

$(Q - p, R^P)$ has VC-dim d

$|R| = |R_p| + |R^P| \leq f(d-1, n-1) + f(d, n-1)$

$= \sum_{i=0}^{d-1} \binom{n-1}{i} + \sum_{i=0}^d \binom{n-1}{i}$

$\leq \sum_{i=0}^d \binom{n-1}{i-1} + \sum_{i=0}^d \binom{n-1}{i}$



For QCP, R shatters Q if $R \cap Q = \text{all subsets of } Q$

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

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

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

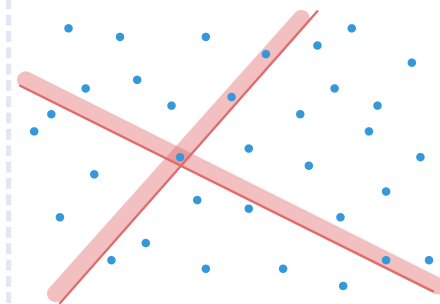
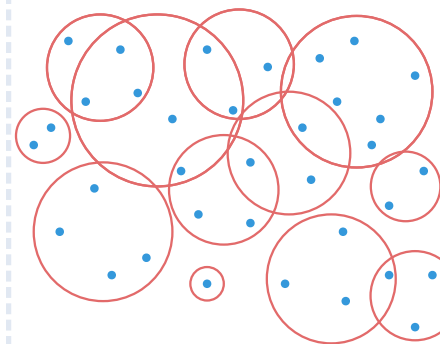
$$g(n) \leq \sum_{i=0}^d \binom{n}{i} \leq n^d$$

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

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

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

(P, R)
 ↑ points ↑ ranges
 (subsets of P)



For QCP, R shatters Q if
 $R \cap Q = \text{all subsets of } Q$

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

Lecture 12

Probability (and techniques)

✓ Additive
Chernoff Inequality

✓ Double-sampling

✓ VC dimension

Algorithms

✓ ϵ -samples
and ϵ -nets