ε-sampling

range space

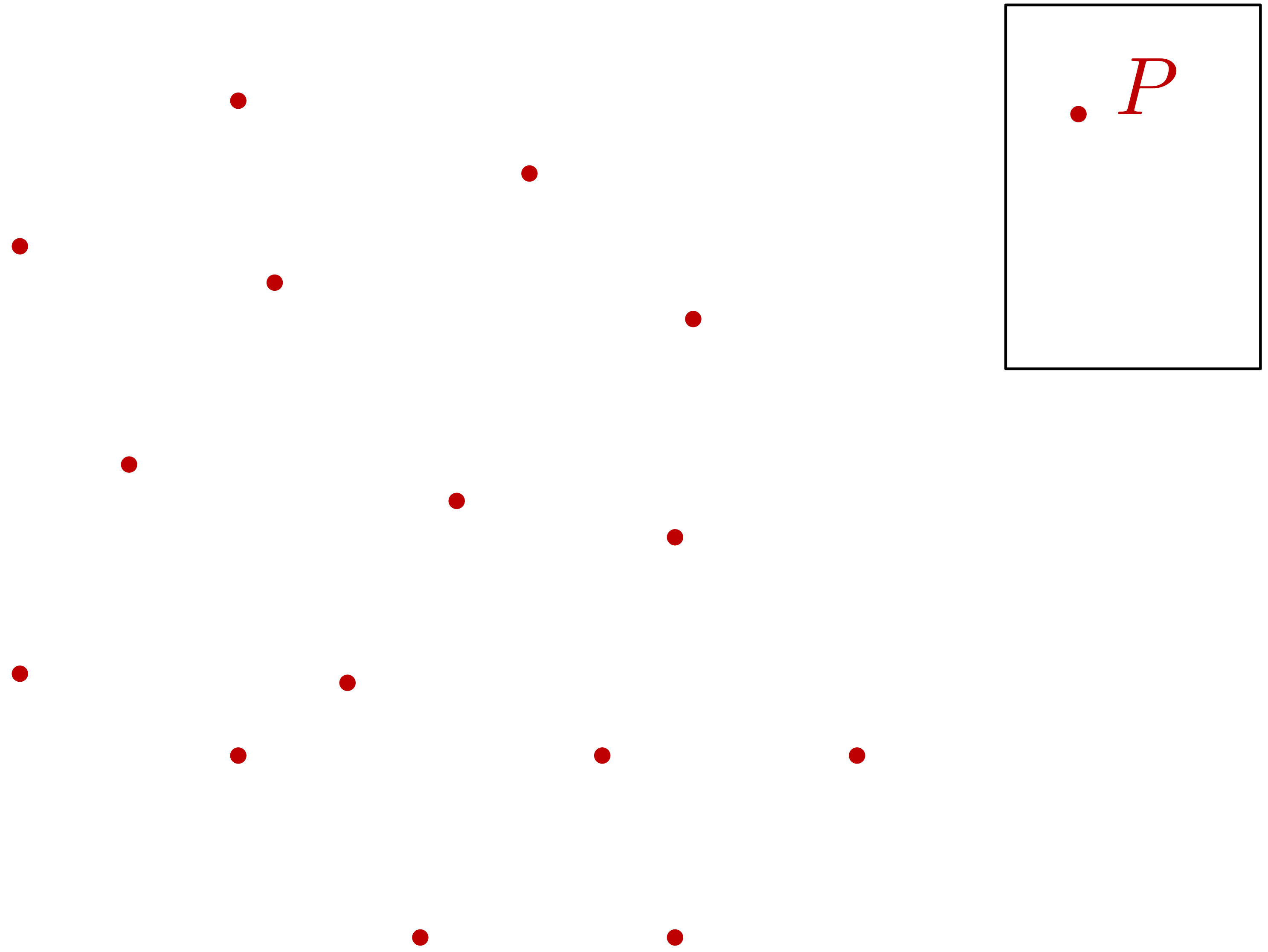
VC-dimension

ε-nets

ε-samples

Motivation: sampling for approximation

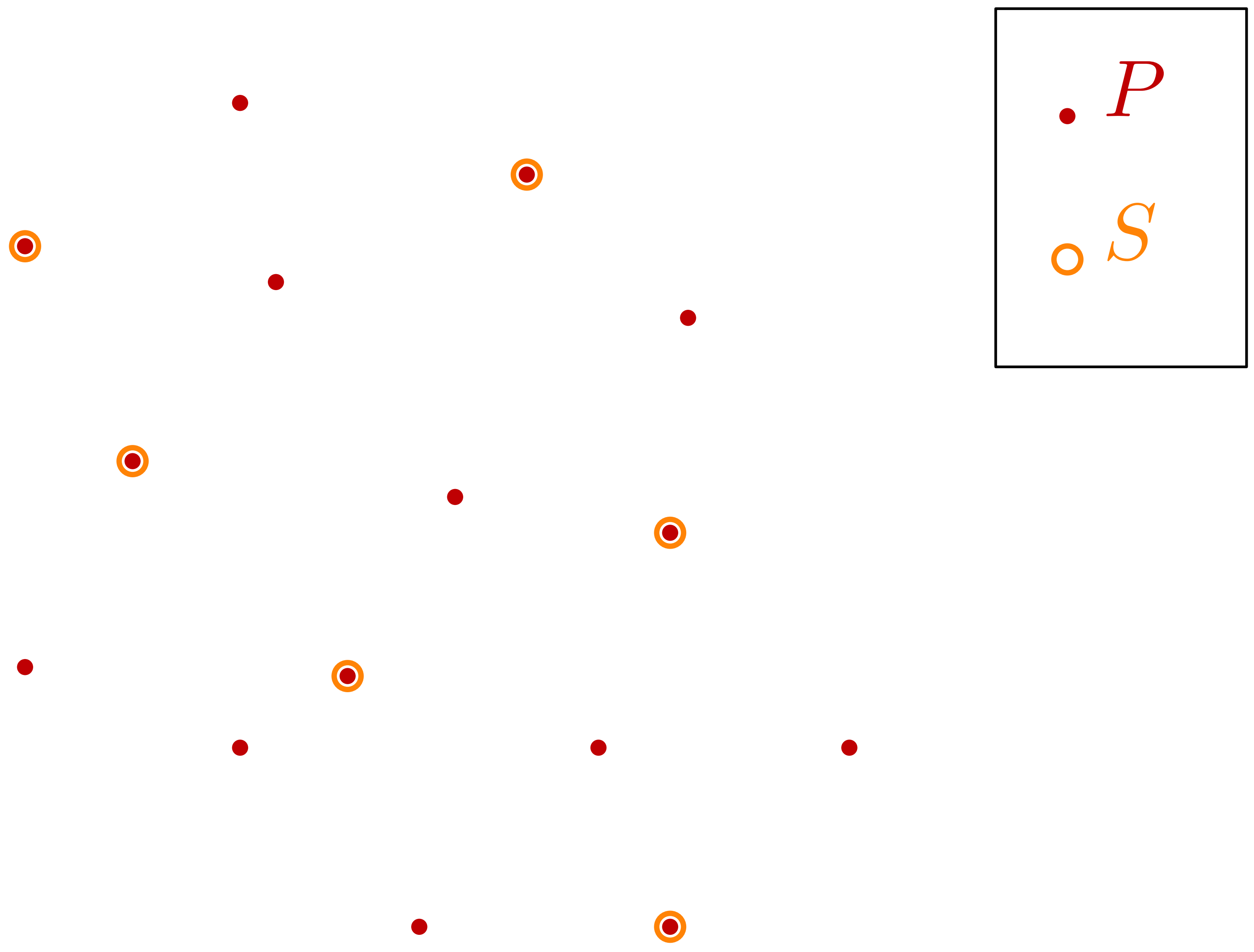
Given P,



Motivation: sampling for approximation

Given P,

how many points do we need to sample ($S \subset P$), such that

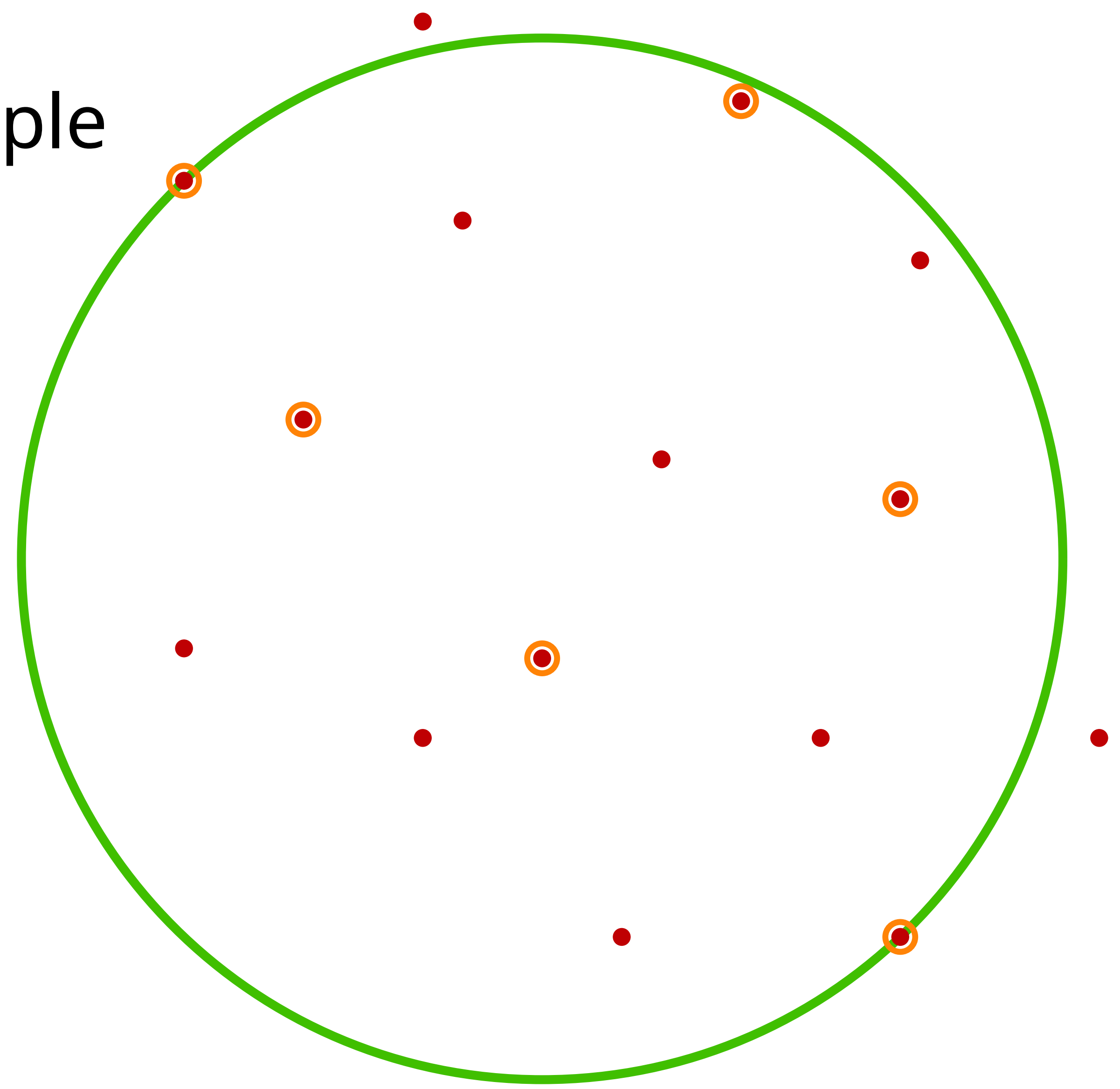


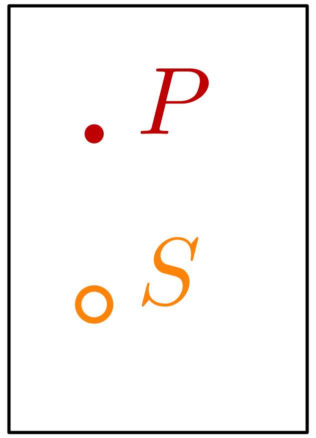
Motivation: sampling for approximation

Given P,

how many points do we need to sample ($S \subset P$), such that

1. the smallest enclosing disk contains 90% of the points in P?





Motivation: sampling for approximation

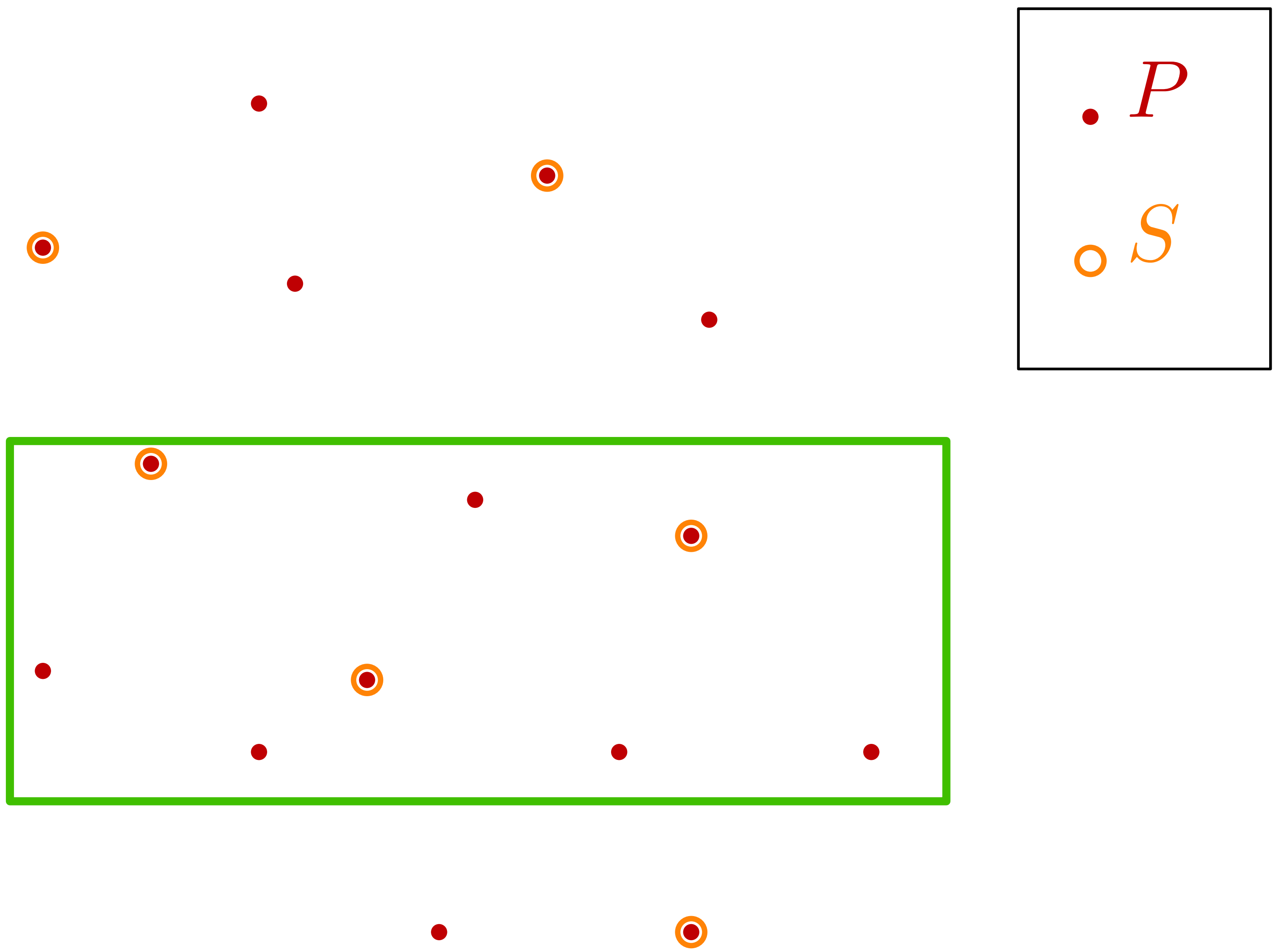
Given P,

how many points do we need to sample ($S \subset P$), such that

1. the smallest enclosing disk contains 90% of the points in P?

2. for any query rectangle r

we can estimate the number of points of P in r?



Motivation: sampling for approximation

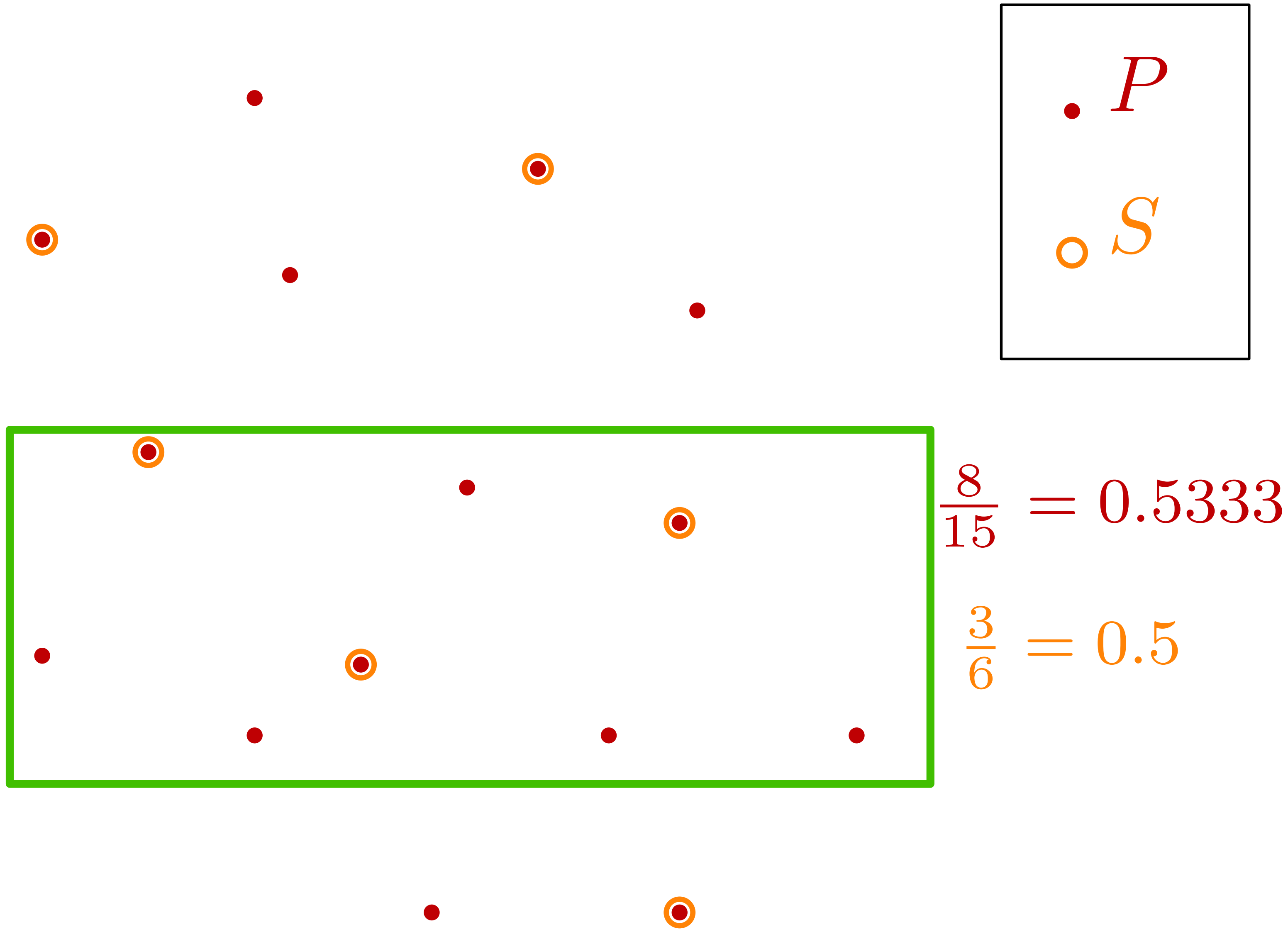
Given P,

how many points do we need to sample ($S \subset P$), such that

1. the smallest enclosing disk contains 90% of the points in P?

2. for any query rectangle r

$$\left| \frac{|r \cap P|}{|P|} - \frac{|r \cap S|}{|S|} \right| \leq 0.25\,?$$



Motivation: sampling for approximation

Given P,

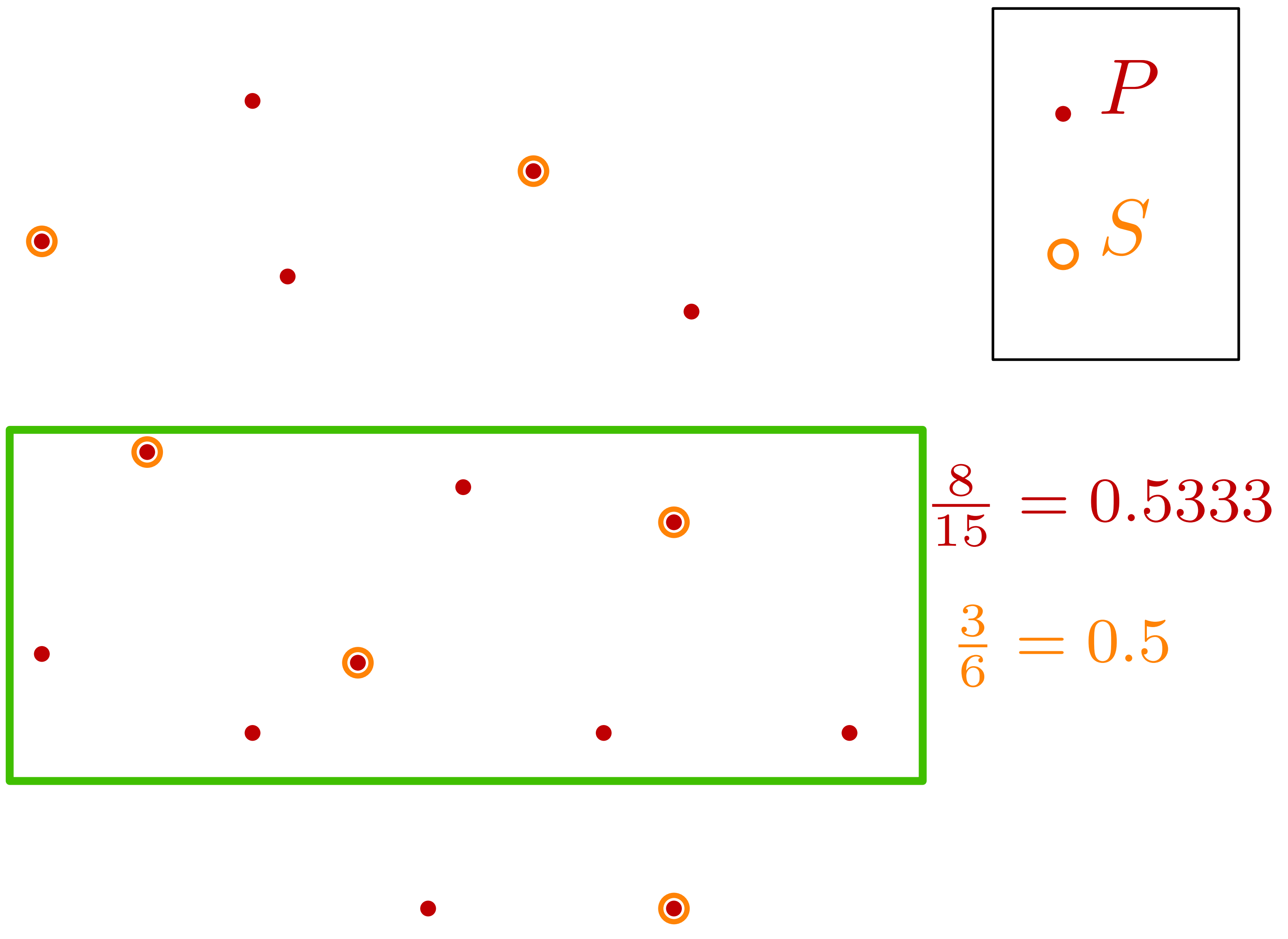
how many points do we need to sample ($S \subset P$), such that

1. the smallest enclosing disk contains 90% of the points in P?

2. for any query rectangle r

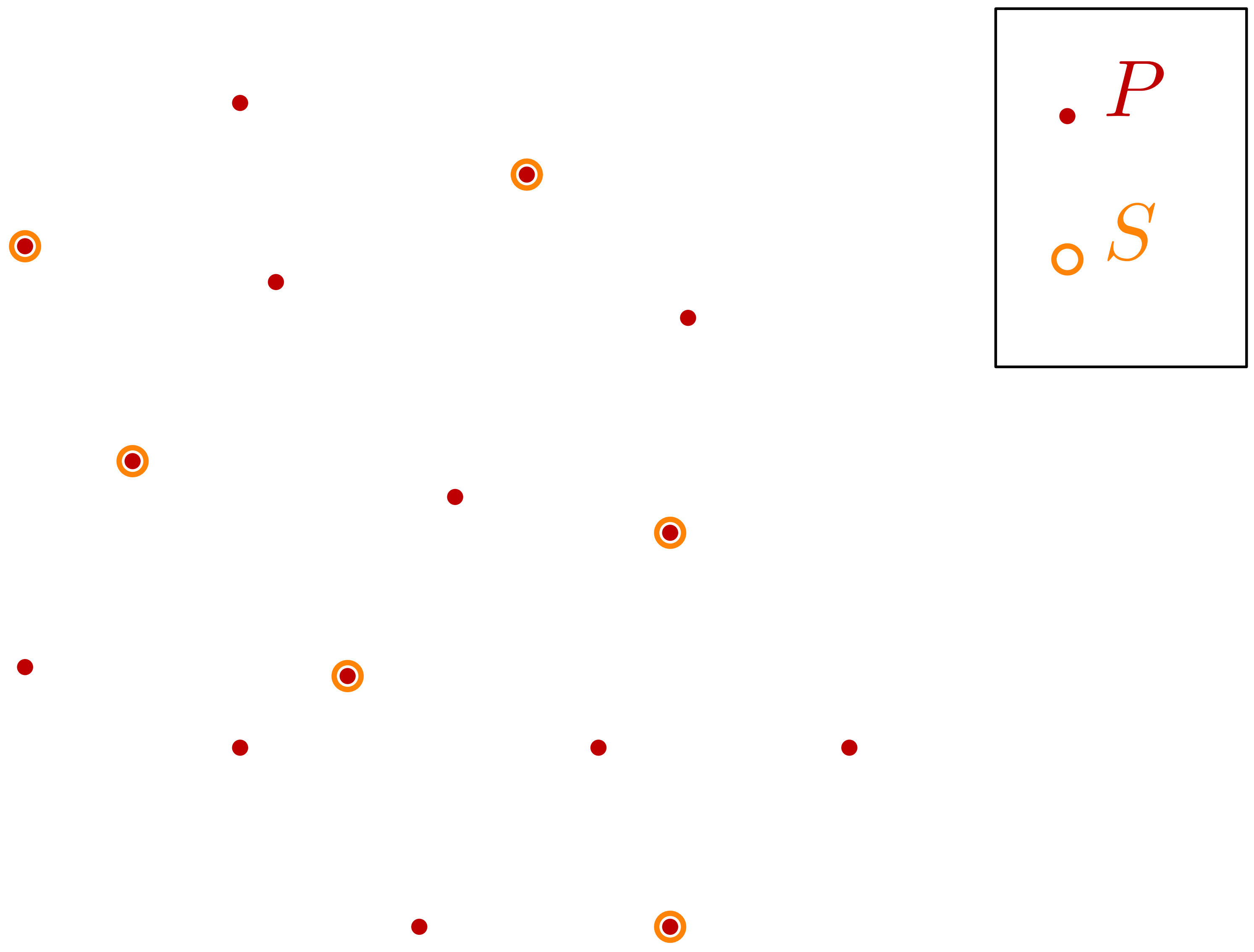
$$\left| \frac{|r \cap P|}{|P|} - \frac{|r \cap S|}{|S|} \right| \leq 0.25 \,?$$

with probability 0.999



Ranges matter

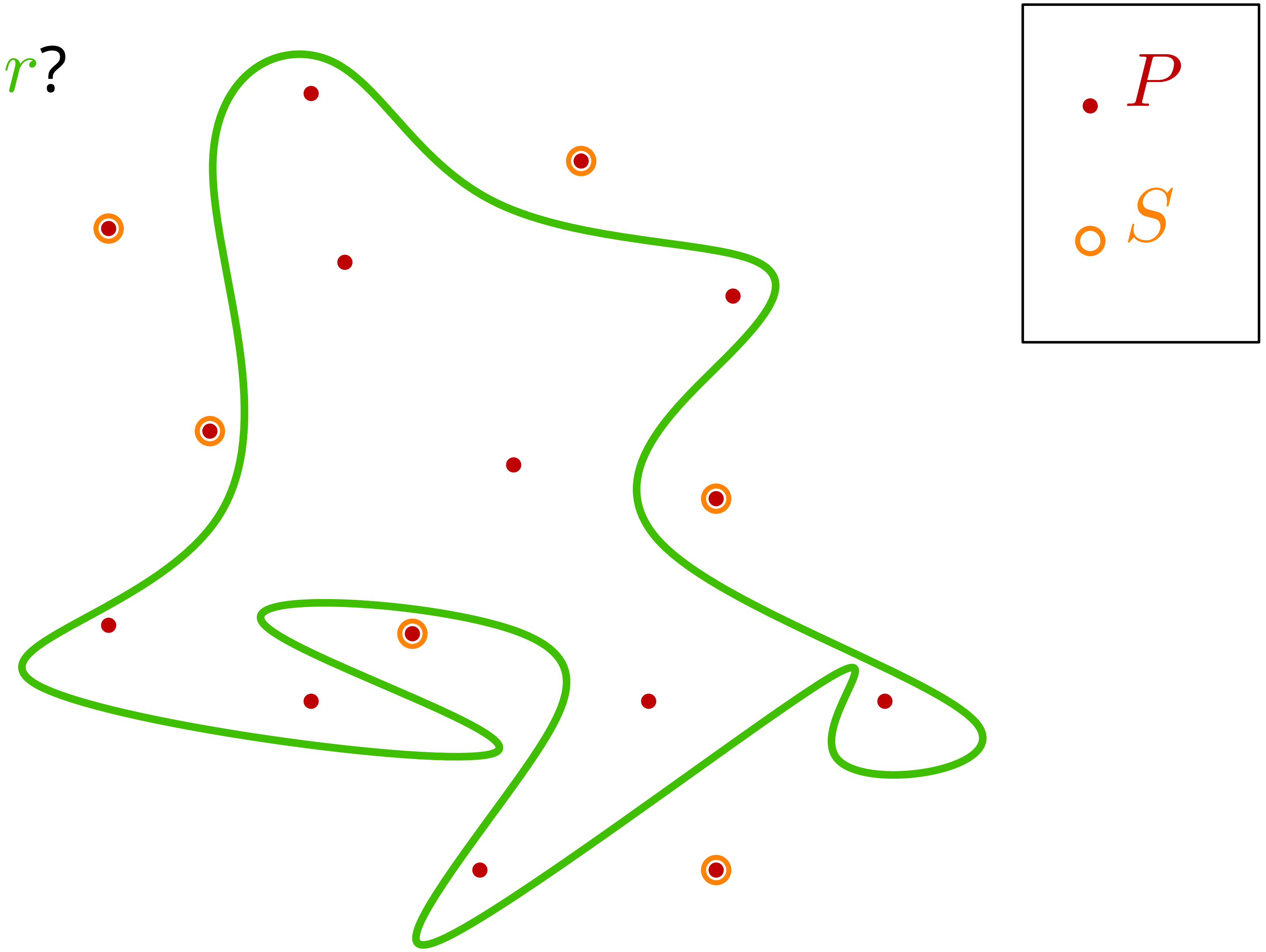
$$\left| \frac{|r \cap P|}{|P|} - \frac{|r \cap S|}{|S|} \right| \leq 0.25$$ for all ranges r?



Ranges matter

$$\left| \frac{|r \cap P|}{|P|} - \frac{|r \cap S|}{|S|} \right| \leq 0.25$$ for all ranges r?

Can't work for general ranges (unless $S \approx P$)

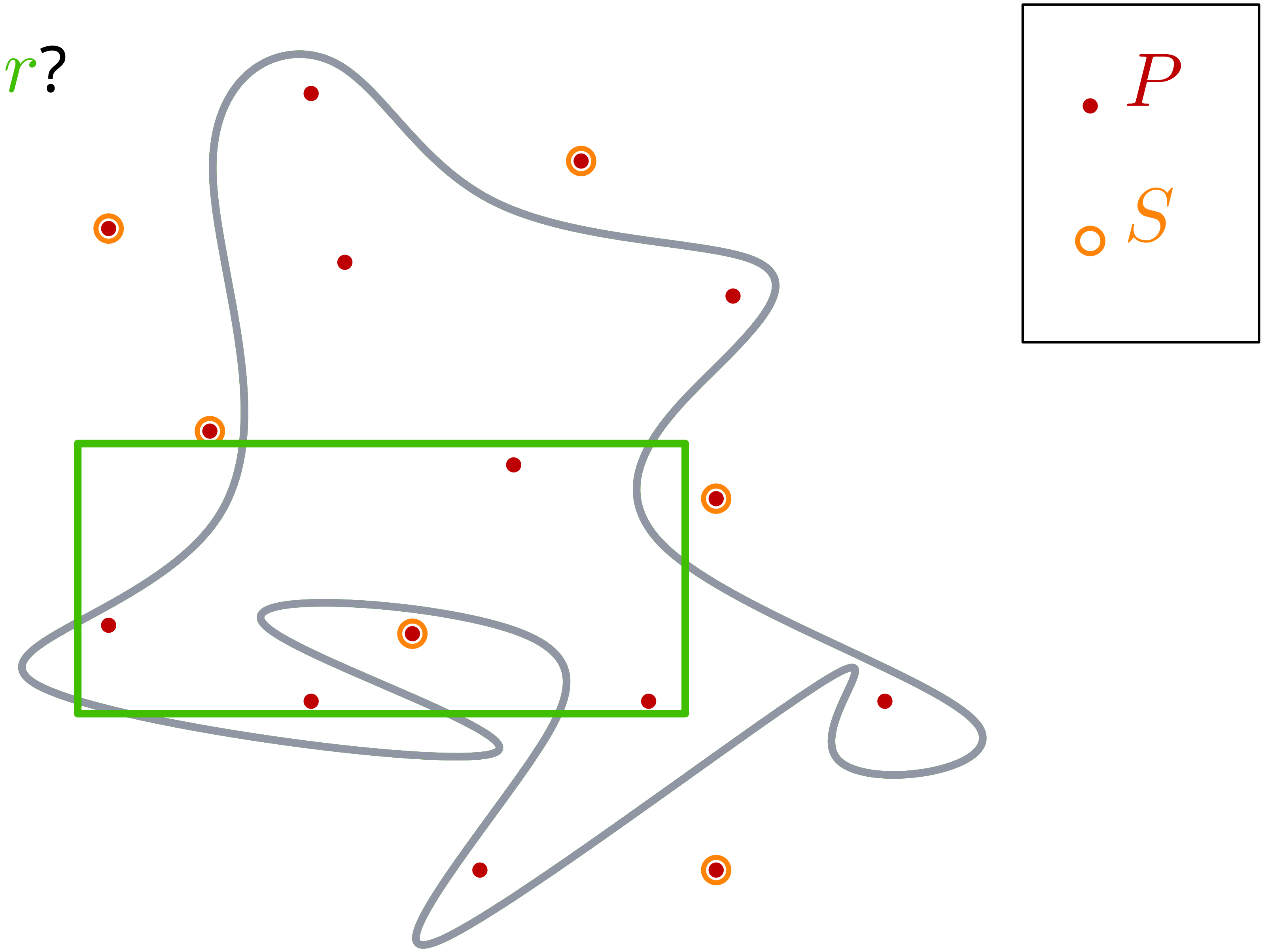


Ranges matter

$$\left| \frac{|r \cap P|}{|P|} - \frac{|r \cap S|}{|S|} \right| \leq 0.25$$ for all ranges r?

Can't work for general ranges (unless $S \approx P$)

Question: Why could this work for (axis-aligned) rectangles?



Ranges matter

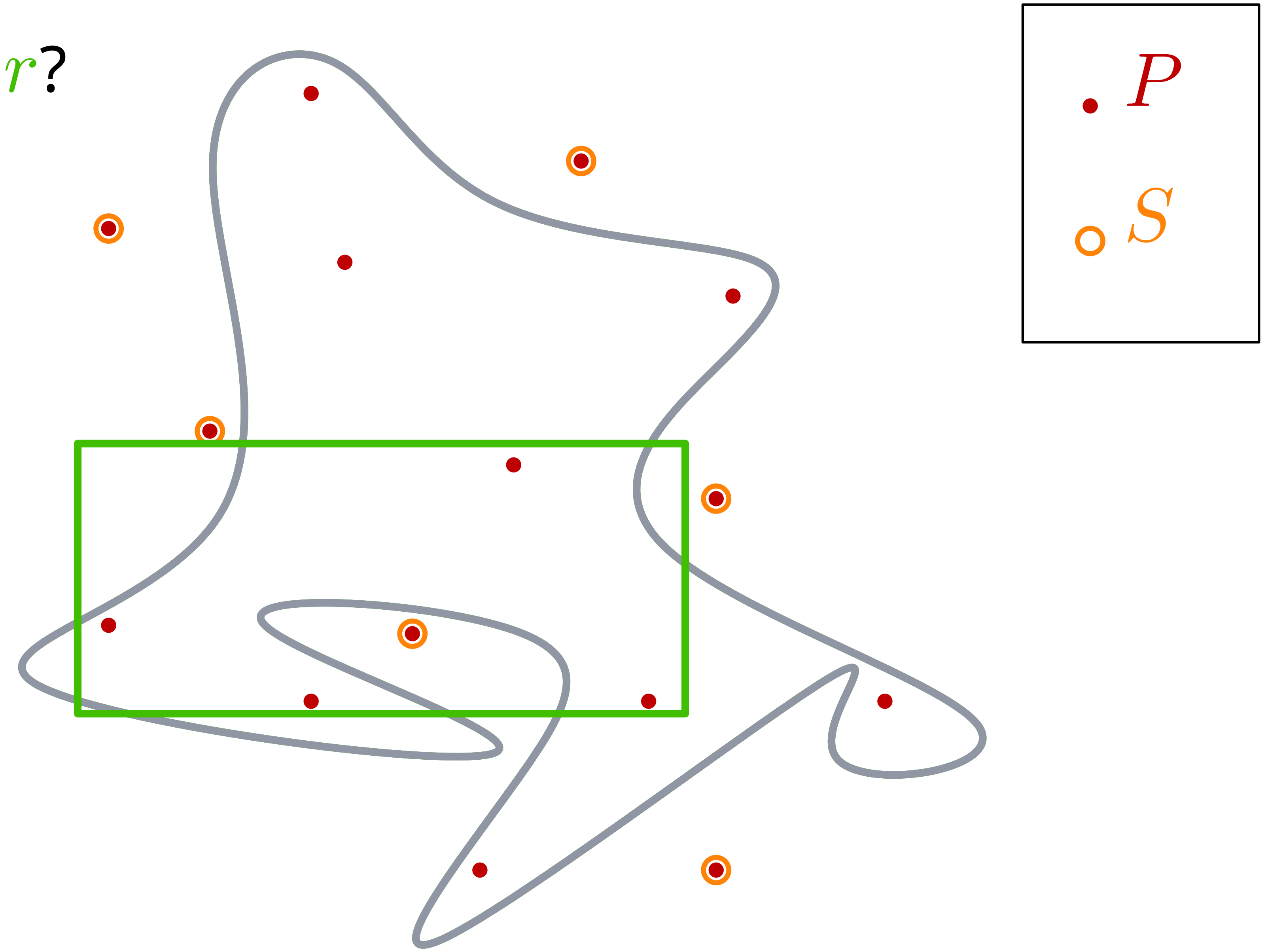
$$\left| \frac{|r \cap P|}{|P|} - \frac{|r \cap S|}{|S|} \right| \leq 0.25$$ for all ranges r?

Can't work for general ranges (unless $S \approx P$)

Question: Why could this work for (axis-aligned) rectangles?

Ideas:

- for 5 points: range with 4 points will contain inner point

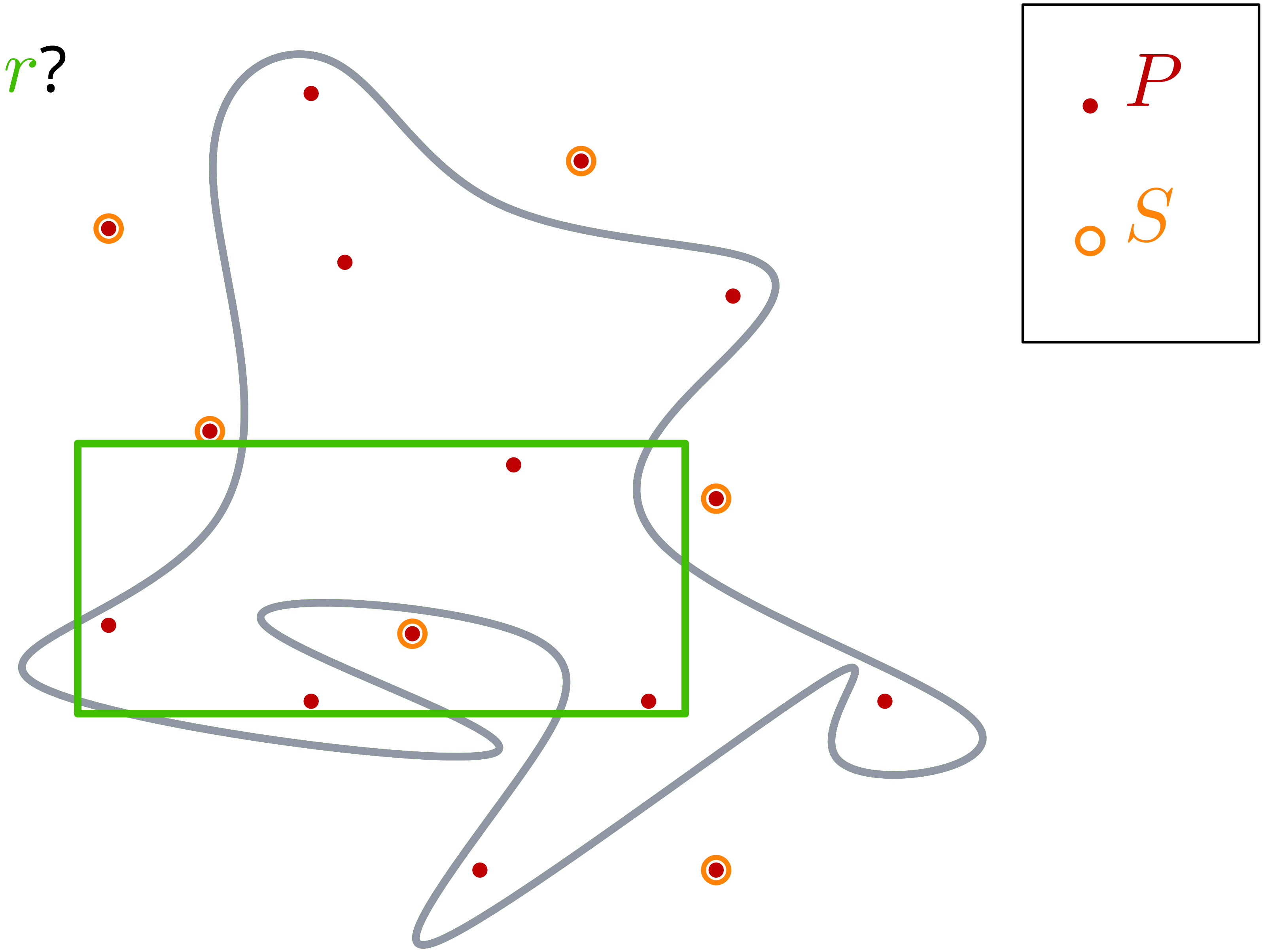


Ranges matter

$$\left| \frac{|r \cap P|}{|P|} - \frac{|r \cap S|}{|S|} \right| \leq 0.25$$ for all ranges r?

Can't work for general ranges (unless $S \approx P$)

Question: Why could this work for (axis-aligned) rectangles?

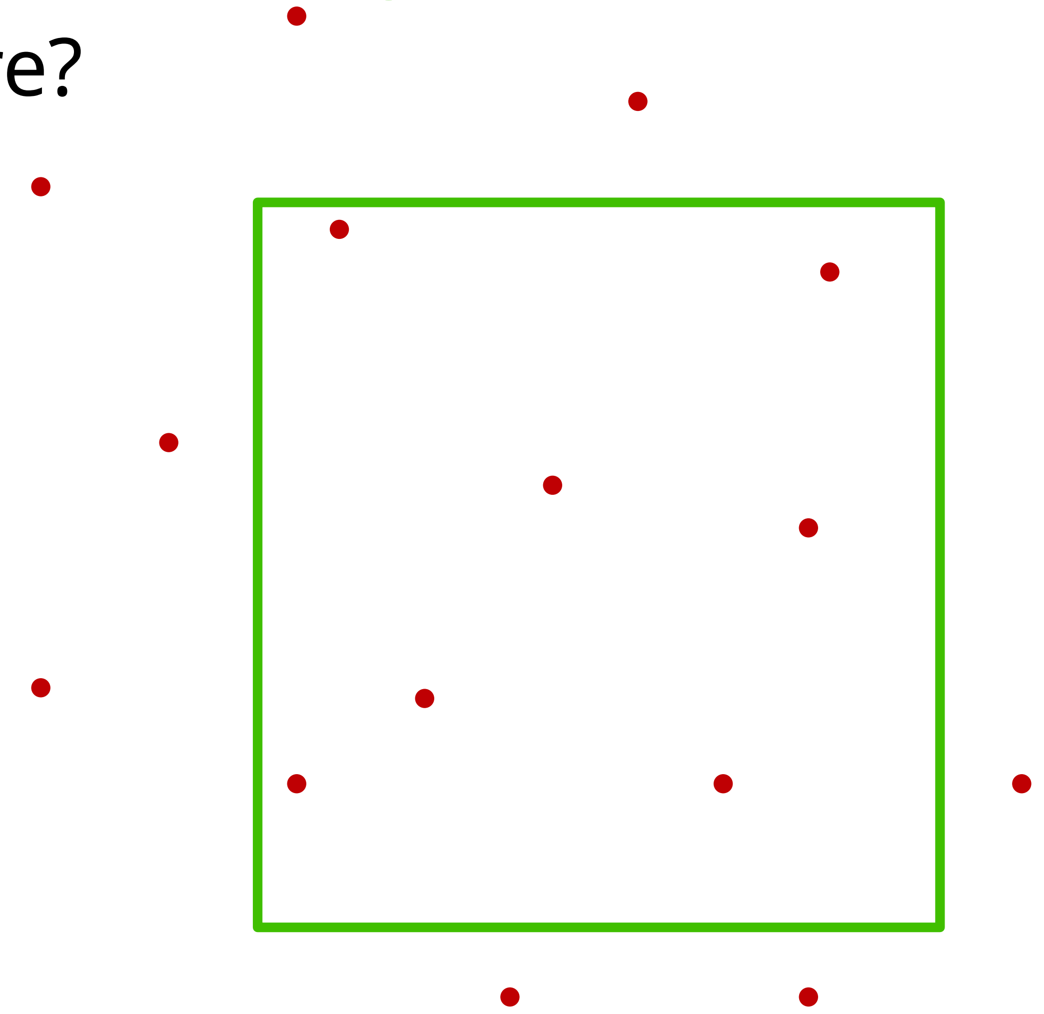


Ideas:

- for 5 points: range with 4 points will contain inner point
- 2^n subsets of P by general ranges but much fewer by rectangles

Quiz

Given point set P of size n and axis-aligned rectangles as ranges, how many sets $P \cap r$ are there?



A $O(n^2)$

B $O(n^3)$

C $O(n^4)$

(we ask for a tight bound)

Quiz

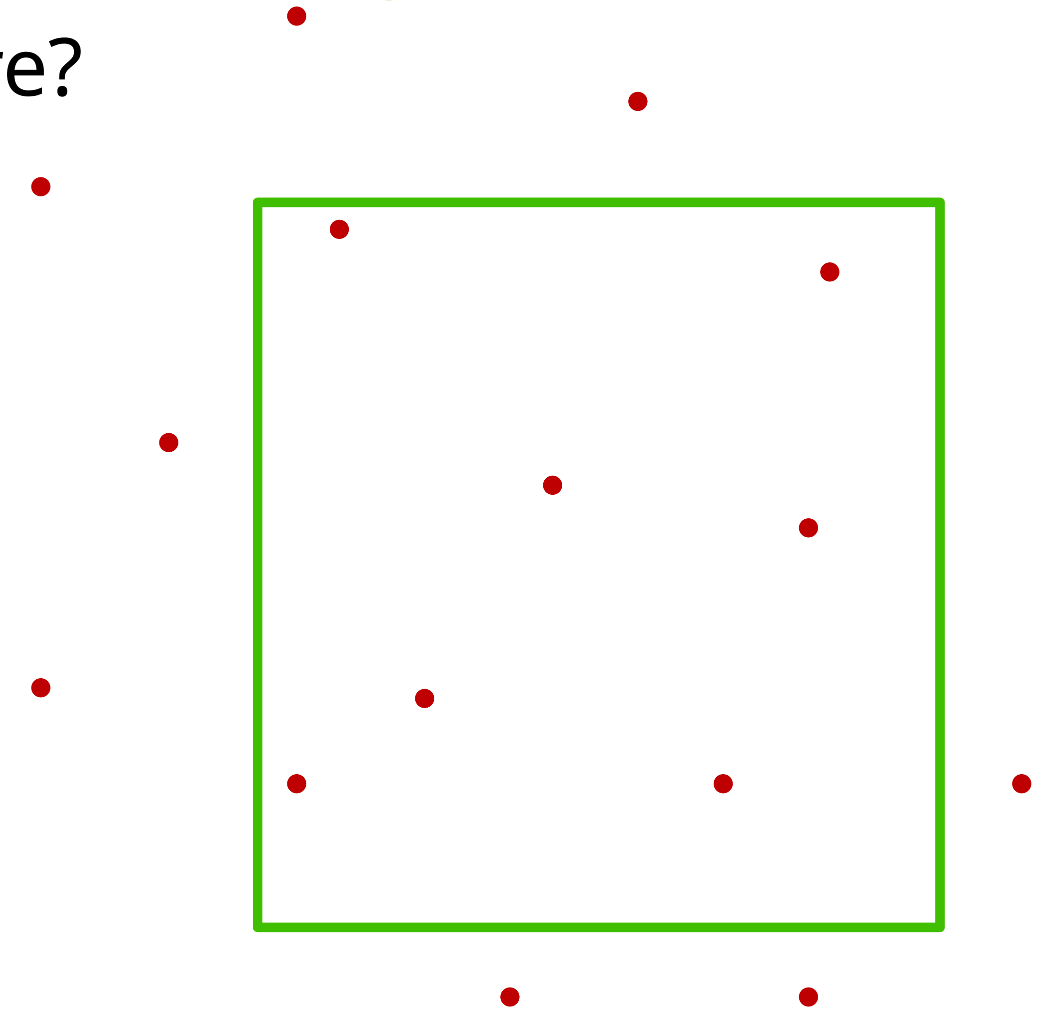
Given point set P of size n and axis-aligned rectangles as ranges, how many sets $P \cap r$ are there?

A $O(n^2)$

B $O(n^3)$

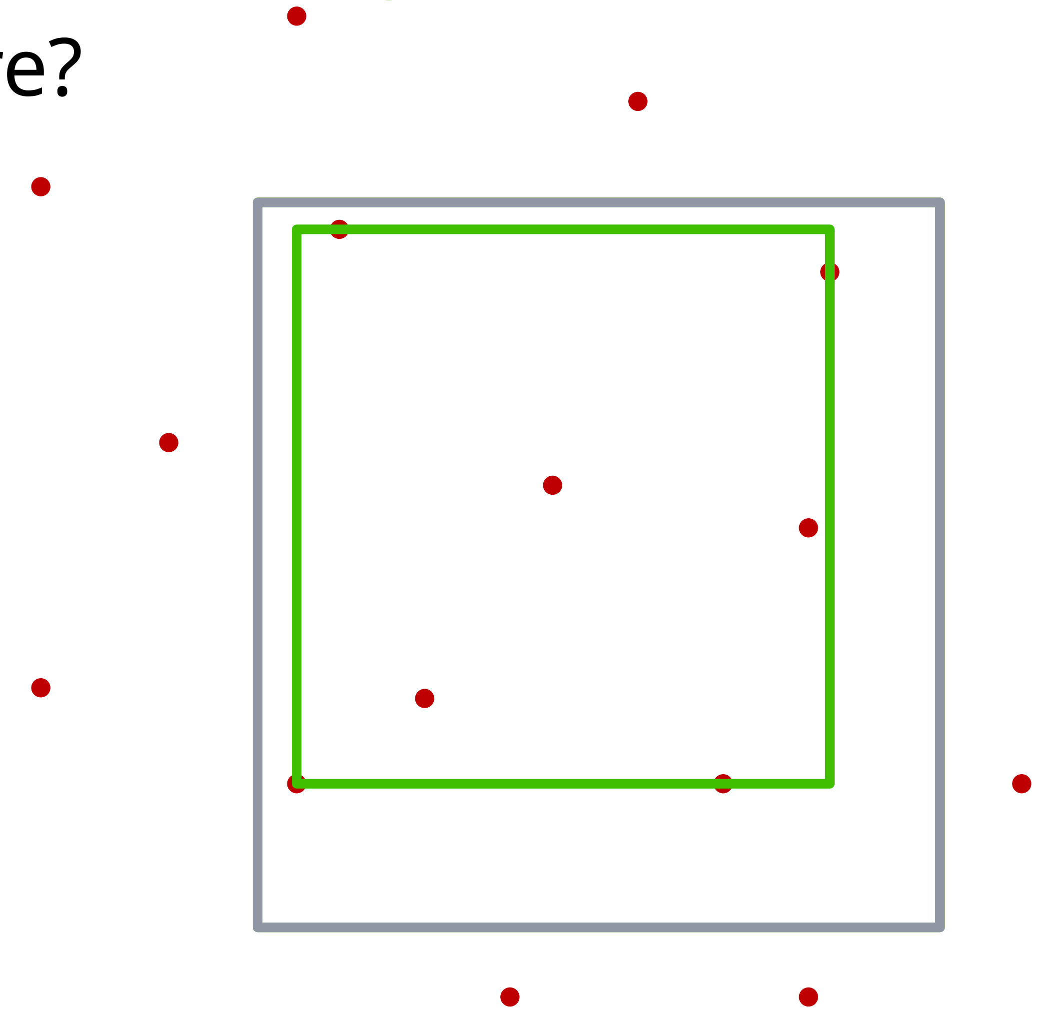
C $O(n^4)$

(we ask for a tight bound)



Quiz

Given point set P of size n and axis-aligned rectangles as ranges, how many sets $P \cap r$ are there?



A $O(n^2)$

B $O(n^3)$

C $O(n^4)$

(we ask for a tight bound)

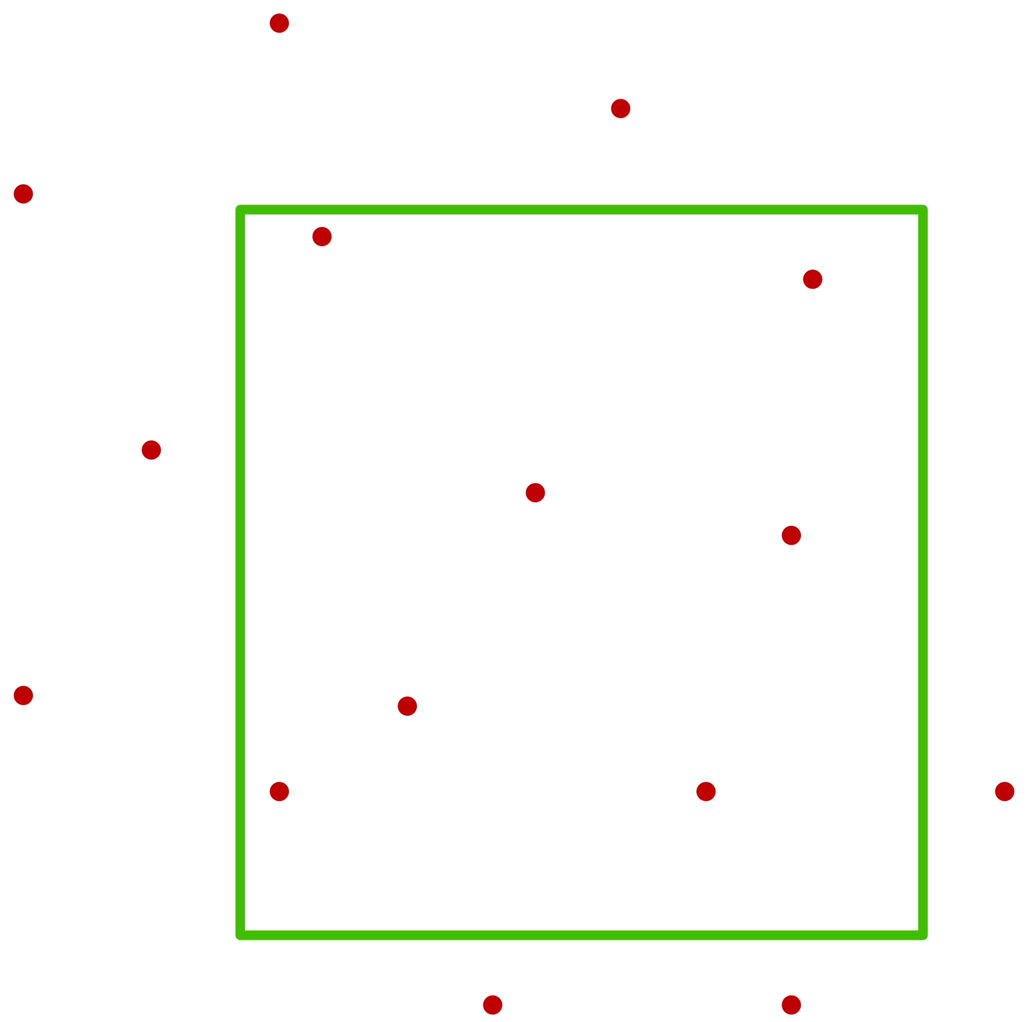
each minimal rectangle defined by left, top, right, bottom point

range spaces and VC-dimension

Range space

range space: pair $(X, \mathcal{R})$

- X is a set
- $\mathcal{R}$ is a subset of power set of X



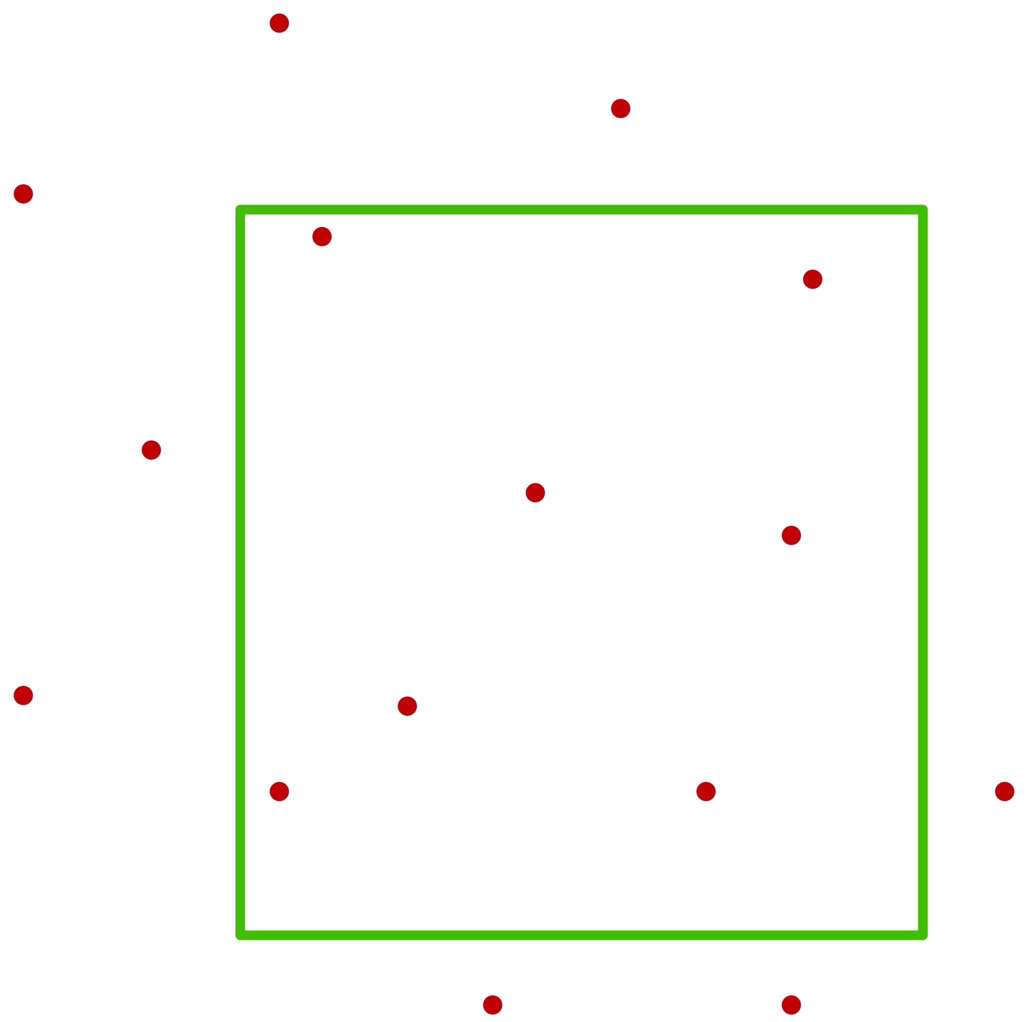
Range space

range space: pair $(X, \mathcal{R})$

- X is a set
- $\mathcal{R}$ is a subset of power set of X

example

- $X = \mathbb{R}^2$
- $\mathcal{R}$: set of axis-aligned rectangles



Range space

range space: pair $(X, \mathcal{R})$

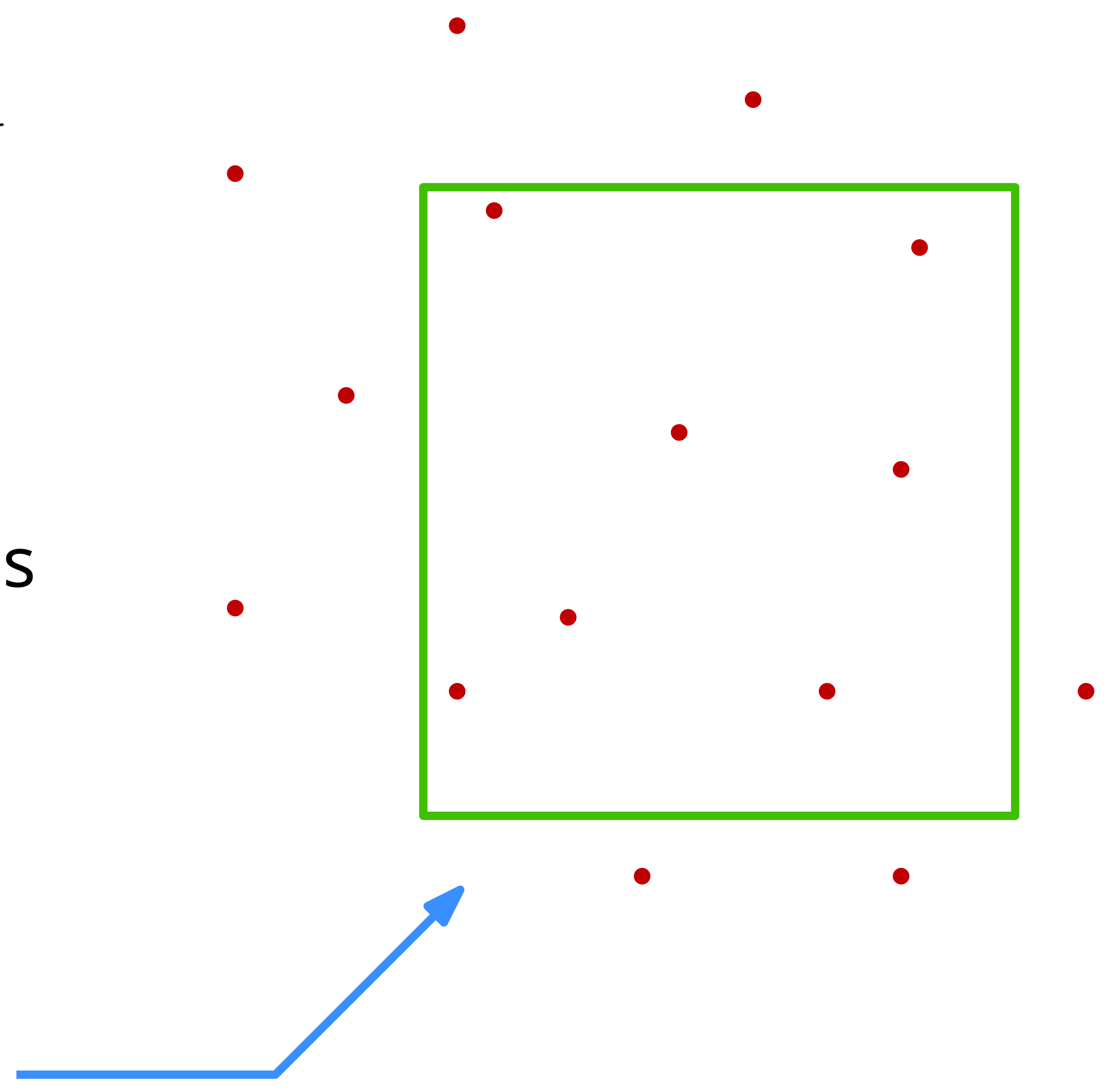
- X is a set
- $\mathcal{R}$ is a subset of power set of X

example

- $X = \mathbb{R}^2$
- $\mathcal{R}$: set of axis-aligned rectangles

restriction $\mathcal{R}_{|P}$

- $P \subseteq X$
- $\mathcal{R}_{|P} := \{r \cap P | r \in \mathcal{R}\}$
- $(P, \mathcal{R}_{|P})$ is a range space, e.g.,



Range space

range space: pair $(X, \mathcal{R})$

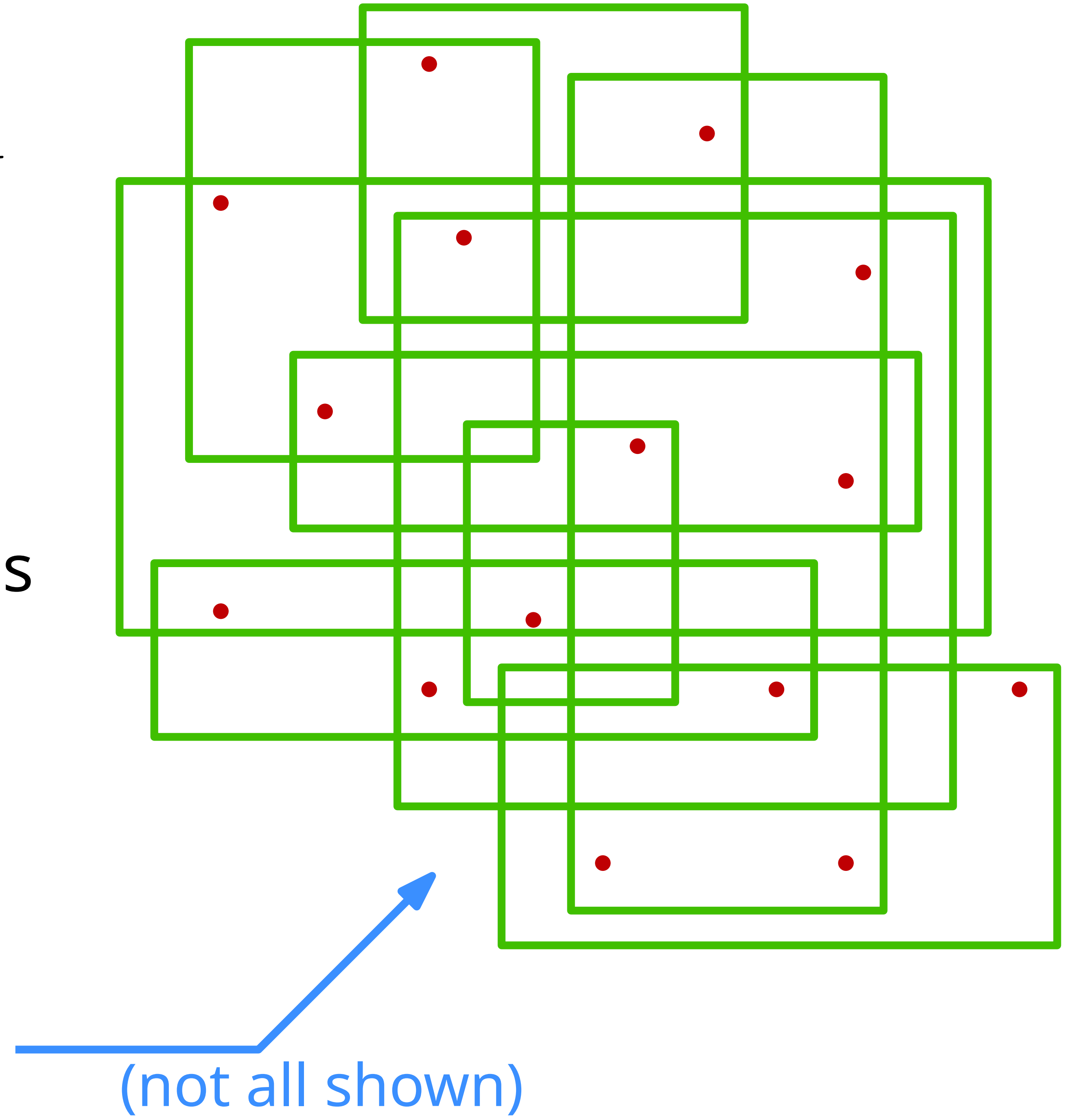
- X is a set
- $\mathcal{R}$ is a subset of power set of X

example

- $X = \mathbb{R}^2$
- $\mathcal{R}$: set of axis-aligned rectangles

restriction $\mathcal{R}_{|P}$

- $P \subseteq X$
- $\mathcal{R}_{|P} := \{r \cap P | r \in \mathcal{R}\}$
- $(P, \mathcal{R}_{|P})$ is a range space, e.g.,



Examples of range spaces

$(\mathbb{R}, \mathcal{I})$, with $\mathcal{I} =$ set of closed intervals

$(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D} =$ set of disks

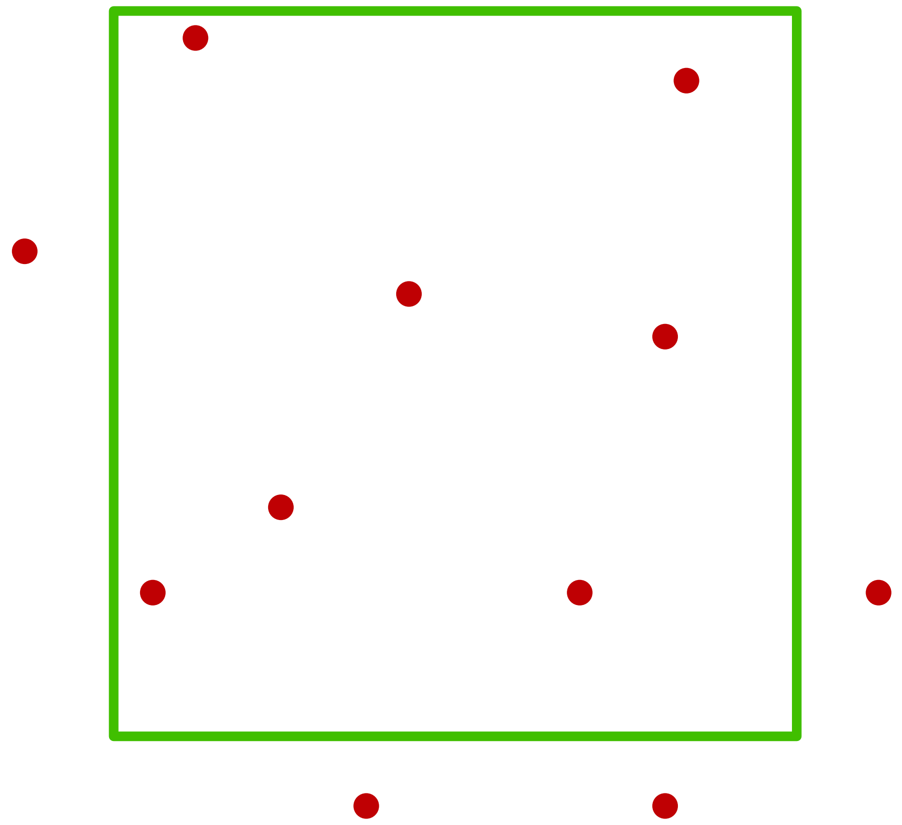
$(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR} =$ set of axis-aligned rectangles

$(\mathbb{R}^2, \mathcal{GR})$, with $\mathcal{GR} =$ set of arbitrary oriented rectangles

$(\mathbb{R}^2, \mathcal{C})$, with $\mathcal{C} =$ set of closed convex sets

VC-dimension

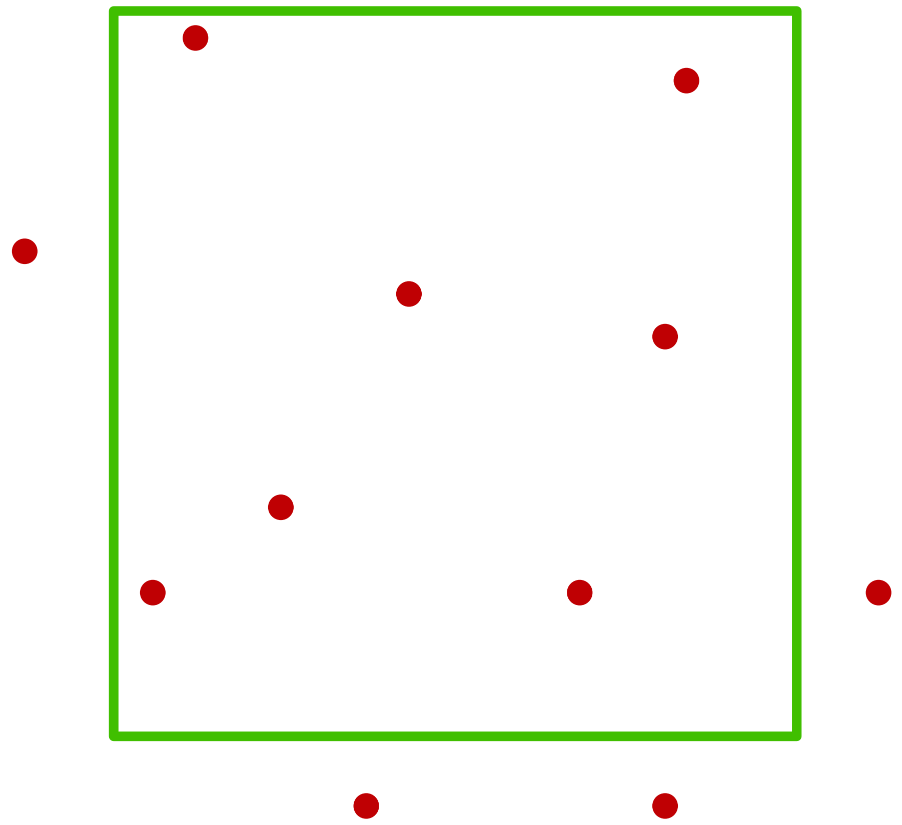
example: $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR} =$ set of axis-aligned rectangles



VC-dimension

example: $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles

want to quantify: range space has "low complexity"

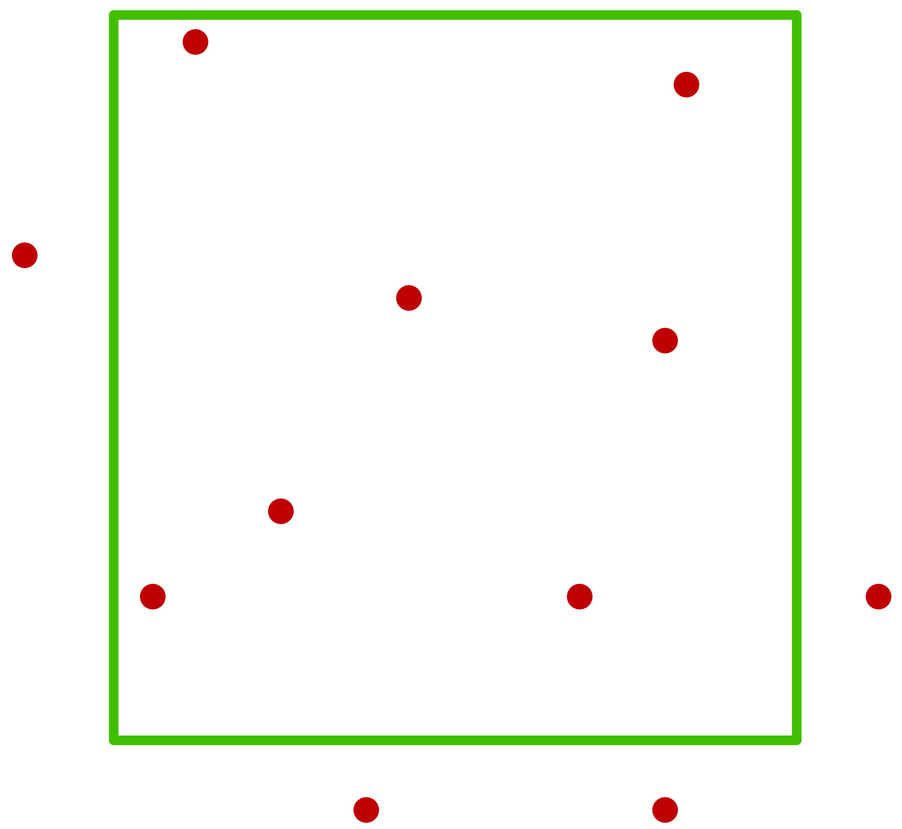


VC-dimension

example: $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR} =$ set of axis-aligned rectangles

want to quantify: range space has "low complexity"

recall: $\mathcal{R}_{|Q} := \{r \cap Q | r \in \mathcal{R}\}$



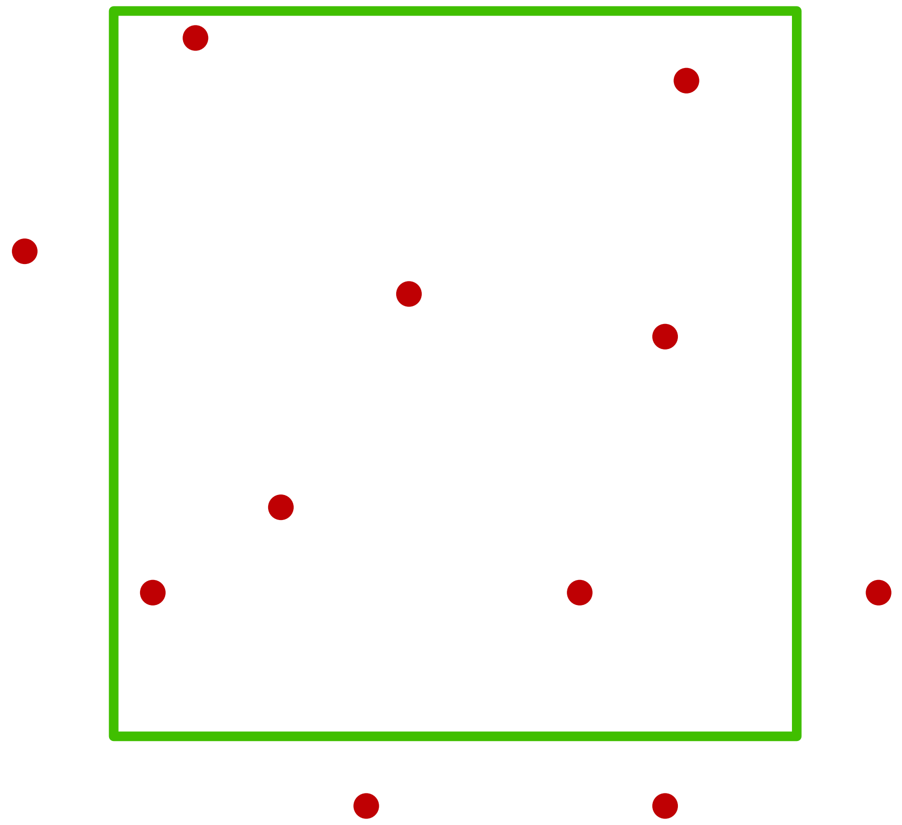
VC-dimension

example: $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles

want to quantify: range space has "low complexity"

recall: $\mathcal{R}_{|Q} := \{r \cap Q | r \in \mathcal{R}\}$

Def: Q is shattered by $\mathcal{R}$ if $R_{|Q}$ is the power set of Q



VC-dimension

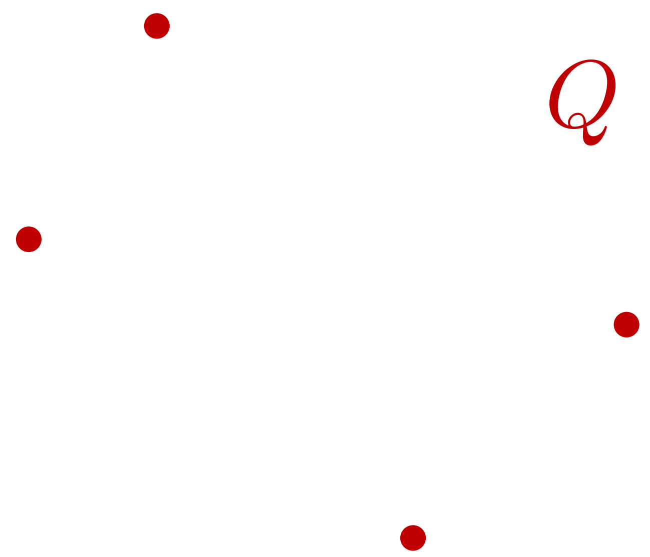
example: $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles

want to quantify: range space has "low complexity"

recall: $\mathcal{R}_{|Q} := \{r \cap Q | r \in \mathcal{R}\}$

Def: Q is shattered by $\mathcal{R}$ if $R_{|Q}$ is the power set of Q

Question: Can Q be shattered by $\mathcal{AR}$?



VC-dimension

example: $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR} =$ set of axis-aligned rectangles

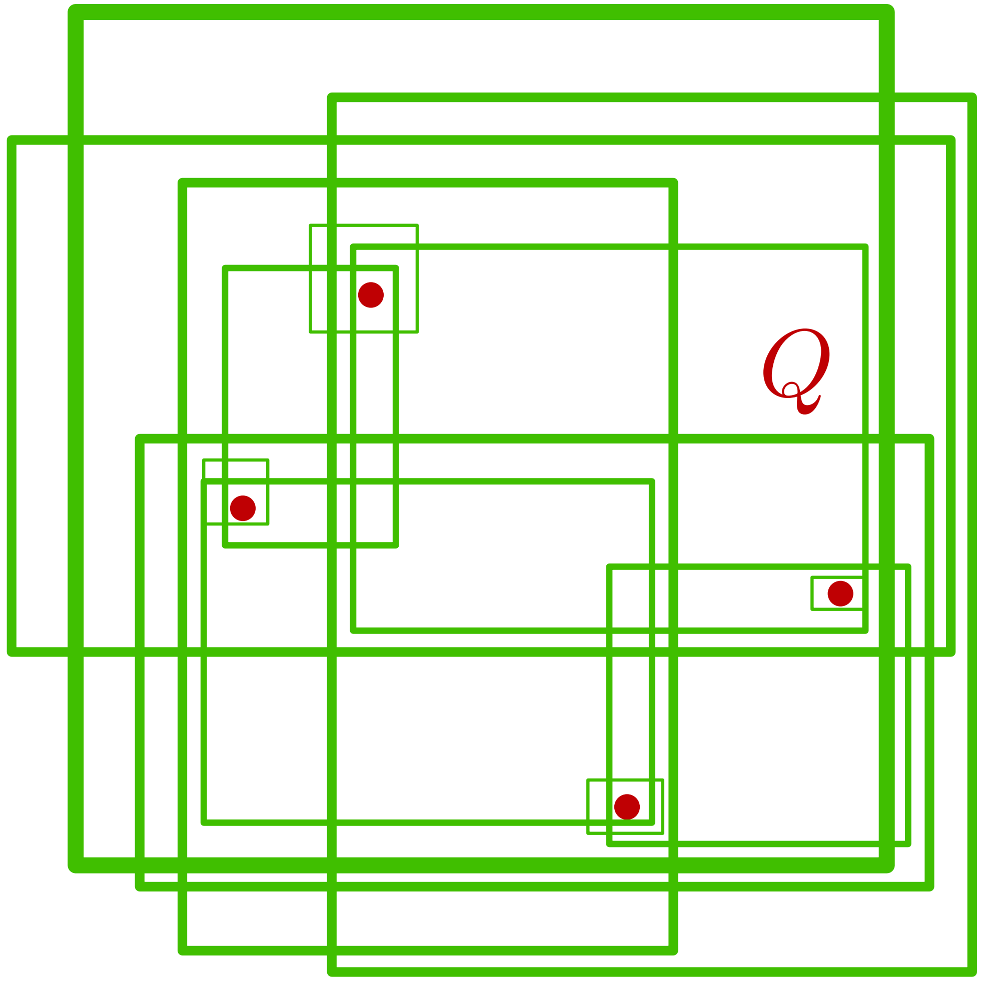
want to quantify: range space has "low complexity"

recall: $\mathcal{R}_{|Q} := \{r \cap Q | r \in \mathcal{R}\}$

Def: Q is shattered by $\mathcal{R}$ if $R_{|Q}$ is the power set of Q

Question: Can Q be shattered by $\mathcal{AR}$?

yes



VC-dimension

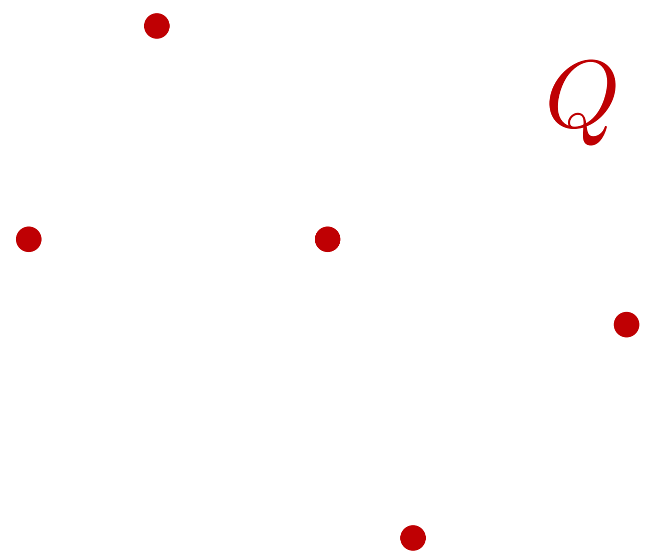
example: $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR} =$ set of axis-aligned rectangles

want to quantify: range space has "low complexity"

recall: $\mathcal{R}_{|Q} := \{r \cap Q | r \in \mathcal{R}\}$

Def: Q is shattered by $\mathcal{R}$ if $R_{|Q}$ is the power set of Q

Question: Can this Q be shattered by $\mathcal{AR}$?



VC-dimension

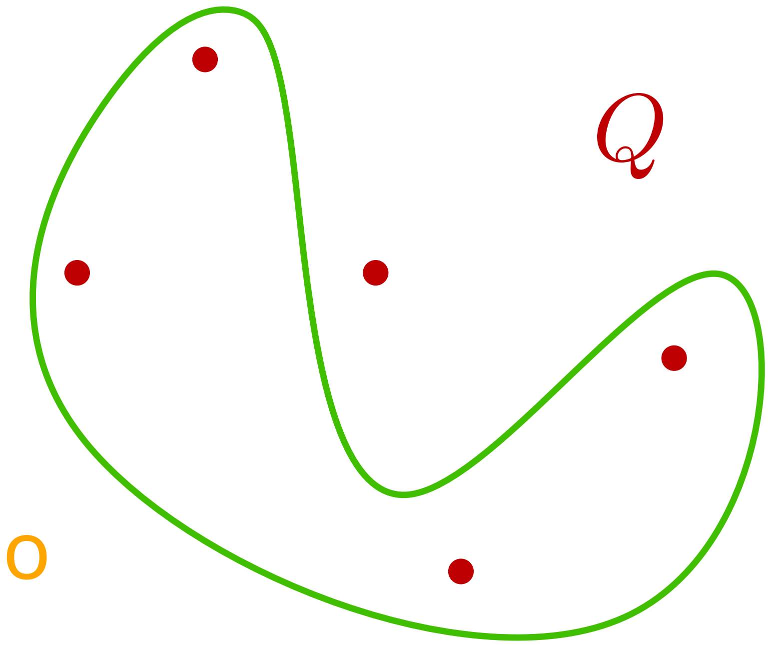
example: $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles

want to quantify: range space has "low complexity"

recall: $\mathcal{R}_{|Q} := \{r \cap Q | r \in \mathcal{R}\}$

Def: Q is shattered by $\mathcal{R}$ if $R_{|Q}$ is the power set of Q

Question: Can this Q be shattered by $\mathcal{AR}$?



no

VC-dimension

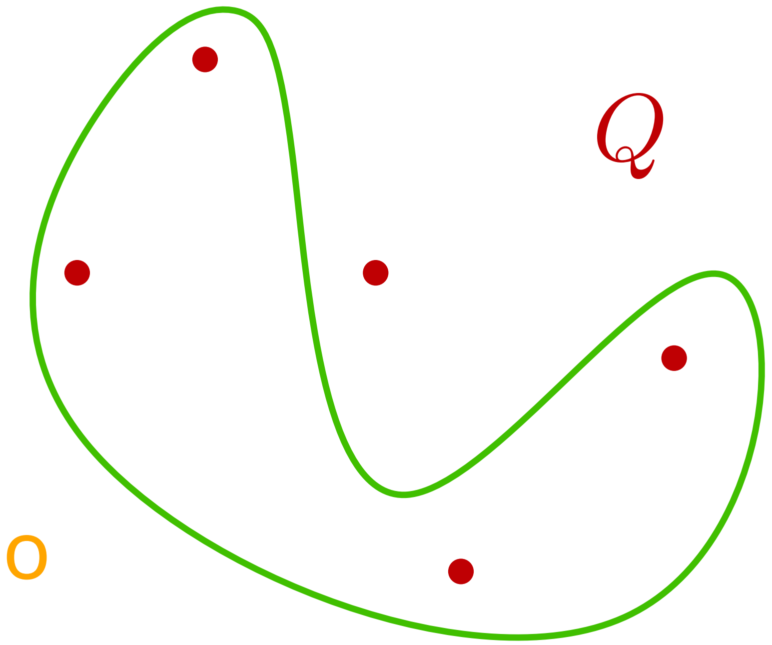
example: $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles

want to quantify: range space has "low complexity"

recall: $\mathcal{R}_{|Q} := \{r \cap Q | r \in \mathcal{R}\}$

Def: Q is shattered by $\mathcal{R}$ if $R_{|Q}$ is the power set of Q

Question: Can this Q be shattered by $\mathcal{AR}$?



no

VC-dimension of a range space:
maximum size of a shattered subset of X

Example $(\mathbb{R}, \mathcal{I})$

No set of 3 or more elements can be shattered.

Example $(\mathbb{R}, \mathcal{I})$

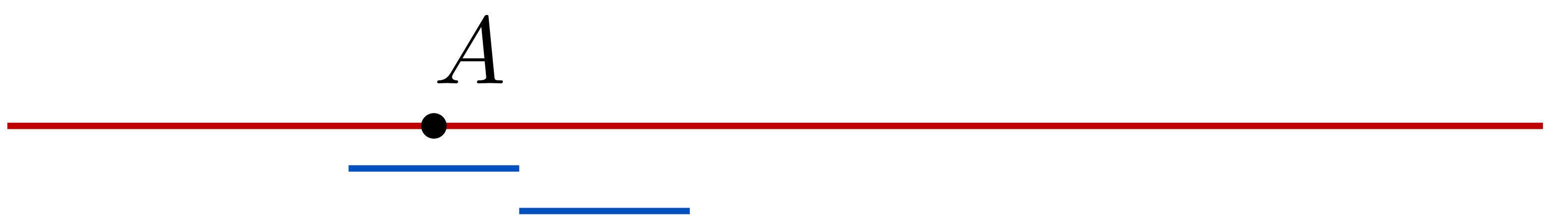
No set of 3 or more elements can be shattered.

Example $(\mathbb{R}, \mathcal{I})$

$$P = \{A\} \subseteq \mathbb{R}$$

No set of 3 or more elements can be shattered.

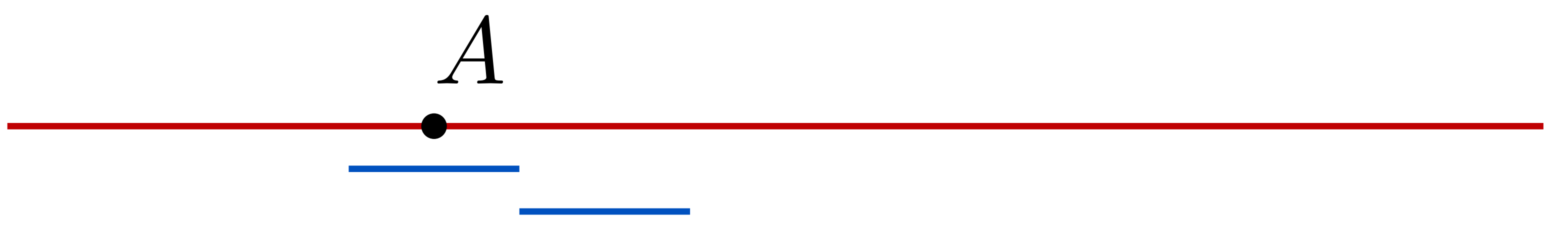
Example $(\mathbb{R}, \mathcal{I})$



$$P = \{A\} \subseteq \mathbb{R}$$

No set of 3 or more elements can be shattered.

Example $(\mathbb{R}, \mathcal{I})$



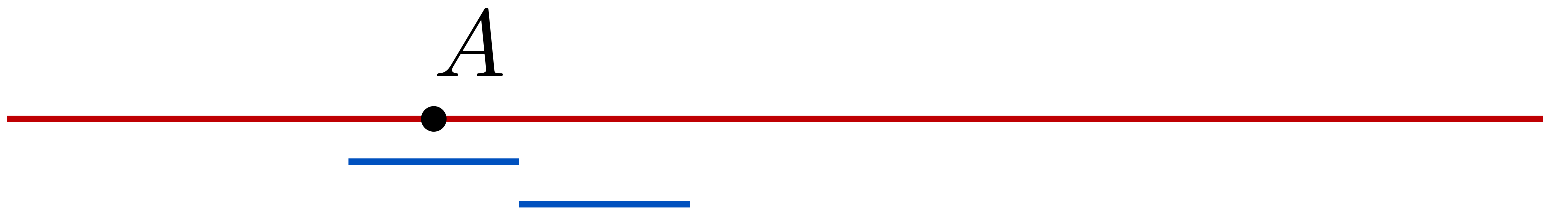
$$P = \{A\} \subseteq \mathbb{R}$$

$$\mathcal{R}_{|P} = \{\varnothing, \{A\}\}$$

$$\left|\mathcal{R}_{|P}\right| = 2 = 2^{|P|}$$

No set of 3 or more elements can be shattered.

Example $(\mathbb{R}, \mathcal{I})$



$$P = \{A\} \subseteq \mathbb{R}$$

$$\mathcal{R}_{|P} = \{\varnothing, \{A\}\}$$

$$\left|\mathcal{R}_{|P}\right| = 2 = 2^{|P|}$$

shattered !

No set of 3 or more elements can be shattered.

Example $(\mathbb{R}, \mathcal{I})$

$$P = \{A, B\} \subseteq \mathbb{R}$$

No set of 3 or more elements can be shattered.

Example $(\mathbb{R}, \mathcal{I})$

$$P = \{A, B\} \subseteq \mathbb{R}$$

$$\varnothing \in \mathcal{R}_{|P}$$

No set of 3 or more elements can be shattered.

Example $(\mathbb{R}, \mathcal{I})$

$$P = \{A, B\} \subseteq \mathbb{R}$$

$$\varnothing \in \mathcal{R}_{|P} \quad \{A\} \in \mathcal{R}_{|P} \quad \{B\} \in \mathcal{R}_{|P}$$

No set of 3 or more elements can be shattered.

Example $(\mathbb{R}, \mathcal{I})$

$P = \{A, B\} \subseteq \mathbb{R}$

$\varnothing \in \mathcal{R}_{|P} \quad \{A\} \in \mathcal{R}_{|P} \quad \{B\} \in \mathcal{R}_{|P} \quad \{A, B\} \in \mathcal{R}_{|P}$

No set of 3 or more elements can be shattered.

Example $(\mathbb{R}, \mathcal{I})$

$$P = \{A, B\} \subseteq \mathbb{R}$$

$$\varnothing \in \mathcal{R}_{|P} \quad \{A\} \in \mathcal{R}_{|P} \quad \{B\} \in \mathcal{R}_{|P} \quad \{A, B\} \in \mathcal{R}_{|P}$$

$$|\mathcal{R}_{|P}| = 2^{|P|}$$

No set of 3 or more elements can be shattered.

Example $(\mathbb{R}, \mathcal{I})$

$$P = \{A, B\} \subseteq \mathbb{R}$$

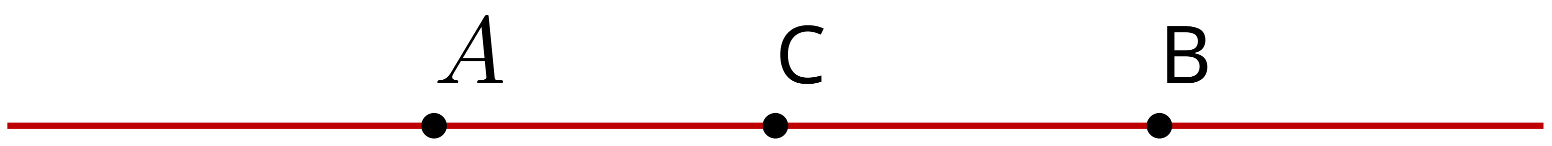
$$\varnothing \in \mathcal{R}_{|P} \quad \{A\} \in \mathcal{R}_{|P} \quad \{B\} \in \mathcal{R}_{|P} \quad \{A, B\} \in \mathcal{R}_{|P}$$

$$|\mathcal{R}_{|P}| = 2^{|P|}$$

shattered !

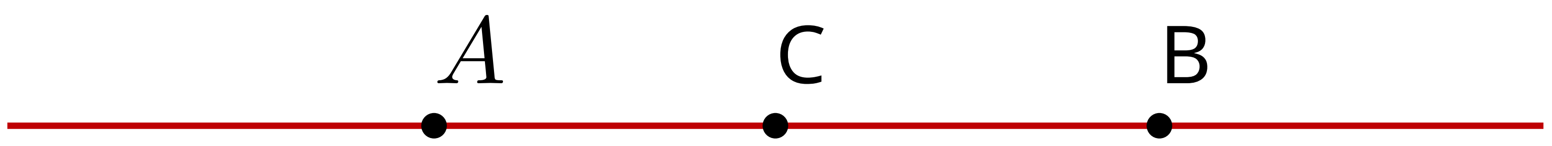
No set of 3 or more elements can be shattered.

Example $(\mathbb{R}, \mathcal{I})$



No set of 3 or more elements can be shattered.

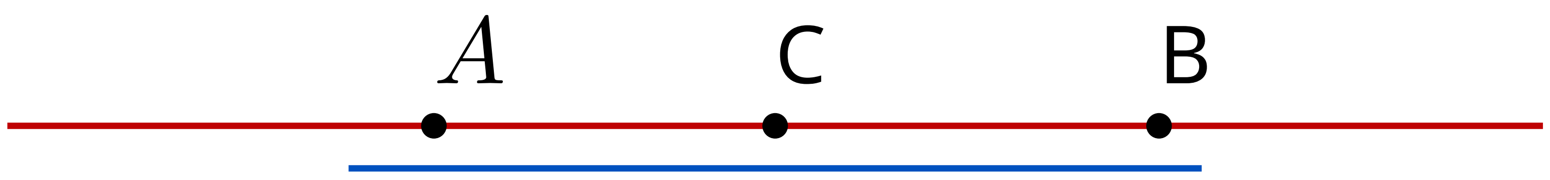
Example $(\mathbb{R}, \mathcal{I})$



$$P = \{A, B, C\} \subseteq \mathbb{R}$$

No set of 3 or more elements can be shattered.

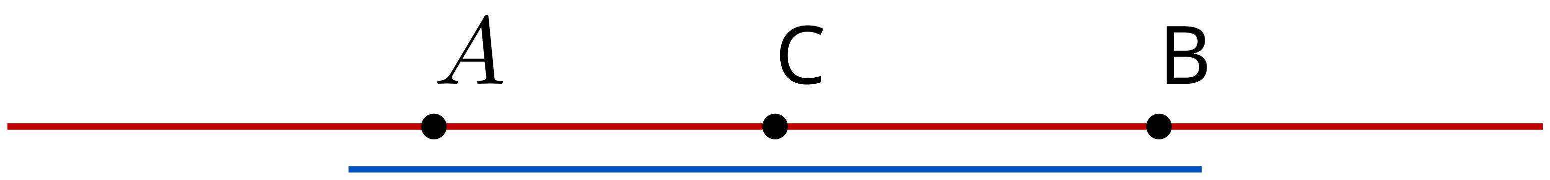
Example $(\mathbb{R}, \mathcal{I})$



$$P = \{A, B, C\} \subseteq \mathbb{R}$$

No set of 3 or more elements can be shattered.

Example $(\mathbb{R}, \mathcal{I})$

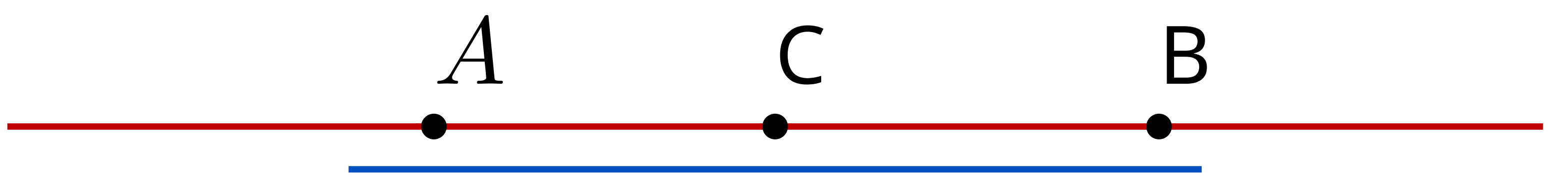


$$P = \{A, B, C\} \subseteq \mathbb{R}$$

$$\{A, B\} \notin \mathcal{R}_{|P}$$

No set of 3 or more elements can be shattered.

Example $(\mathbb{R}, \mathcal{I})$



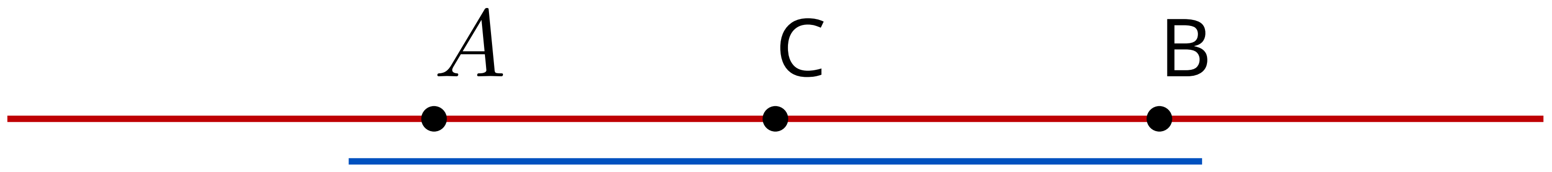
$$P = \{A, B, C\} \subseteq \mathbb{R}$$

$$\{A, B\} \notin \mathcal{R}_{|P}$$

$$|\mathcal{R}_{|P}| < 2^{|P|}$$

No set of 3 or more elements can be shattered.

Example $(\mathbb{R}, \mathcal{I})$



$$P = \{A, B, C\} \subseteq \mathbb{R}$$

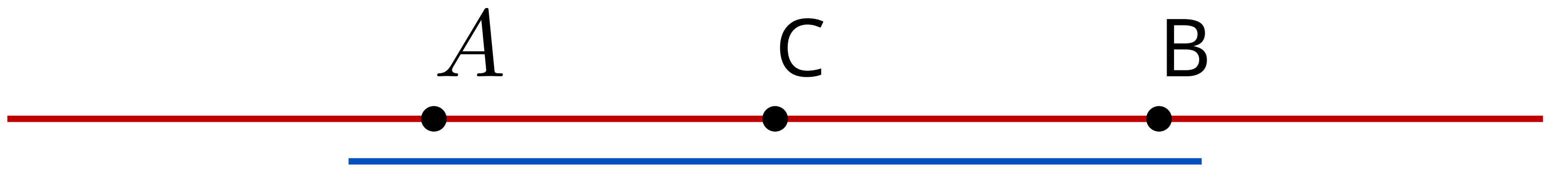
$$\{A, B\} \notin \mathcal{R}_{|P}$$

$$|\mathcal{R}_{|P}| < 2^{|P|}$$

not shattered !

No set of 3 or more elements can be shattered.

Example $(\mathbb{R}, \mathcal{I})$



$$P = \{A, B, C\} \subseteq \mathbb{R}$$

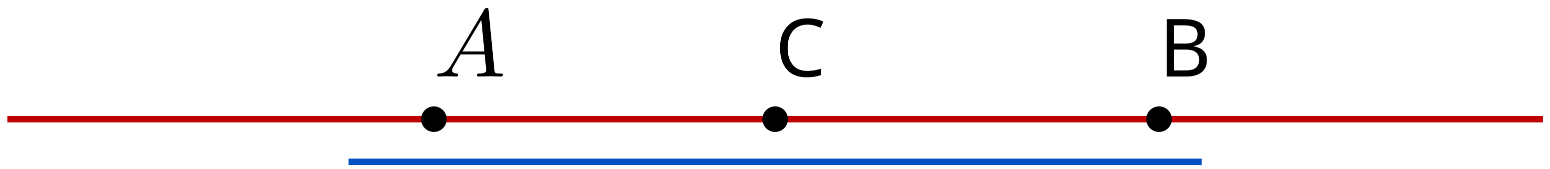
$$\{A, B\} \notin \mathcal{R}_{|P}$$

$$|\mathcal{R}_{|P}| < 2^{|P|}$$

not shattered !

No set of 3 or more elements can be shattered.

Example $(\mathbb{R}, \mathcal{I})$



$$P = \{A, B, C\} \subseteq \mathbb{R}$$

$$\{A, B\} \notin \mathcal{R}_{|P}$$

$$|\mathcal{R}_{|P}| < 2^{|P|}$$

not shattered !

No set of 3 or more elements can be shattered.

VC-dimension $= 2$

Quiz

range space $(\mathbb{R}, \mathcal{I}_{\rightarrow})$ with $\mathcal{I}_{\rightarrow} = \{[a, \infty) | a \in \mathbb{R}\}$

What is the VC-dimension of this space?

A 1

B 2

C 3

Quiz

range space $(\mathbb{R}, \mathcal{I}_{\rightarrow})$ with $\mathcal{I}_{\rightarrow} = \{[a, \infty) | a \in \mathbb{R}\}$

What is the VC-dimension of this space?

A 1

B 2

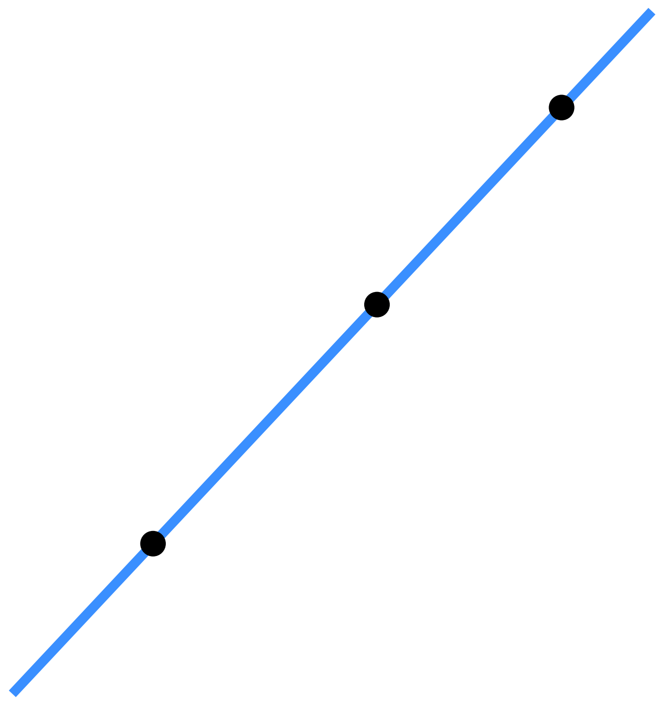
C 3

Example: disks as ranges

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D} =$ set of disks

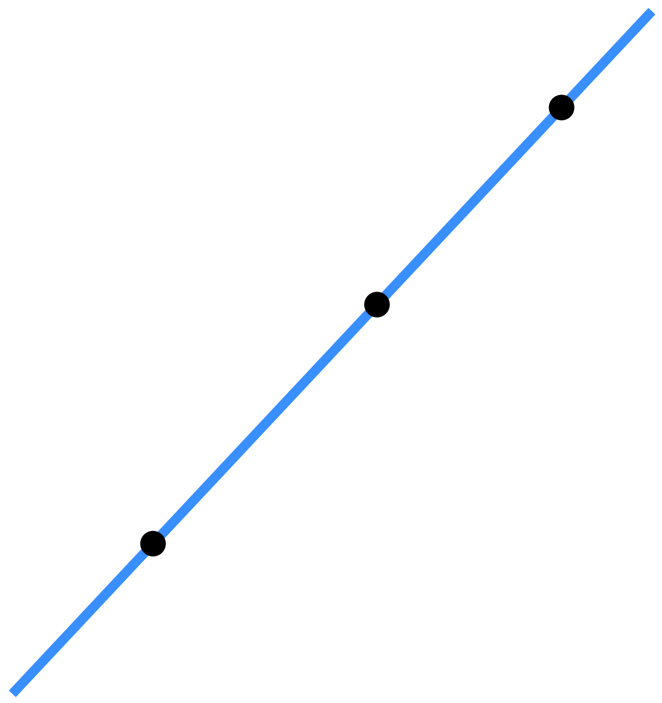
Example: disks as ranges

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D} =$ set of disks



Example: disks as ranges

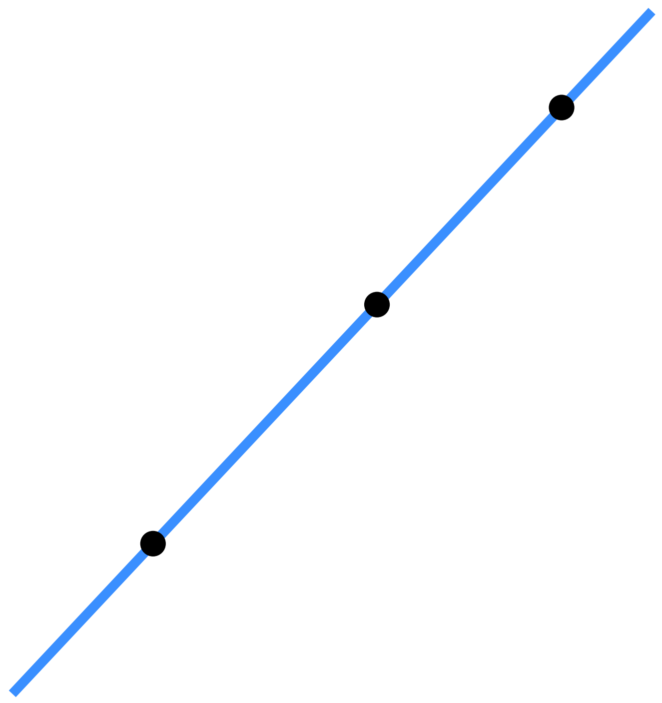
range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D}$ = set of disks



not shatter !

Example: disks as ranges

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D}$ = set of disks

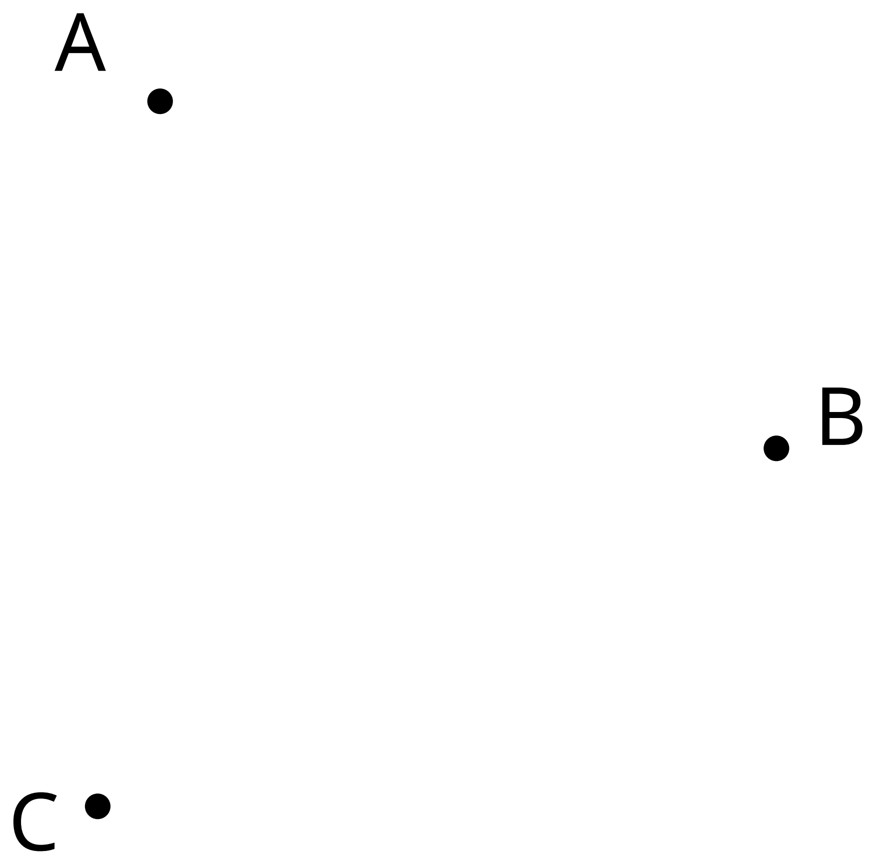


not shatter !

not relevant, since VC-dimension = maximum size of shattered subset

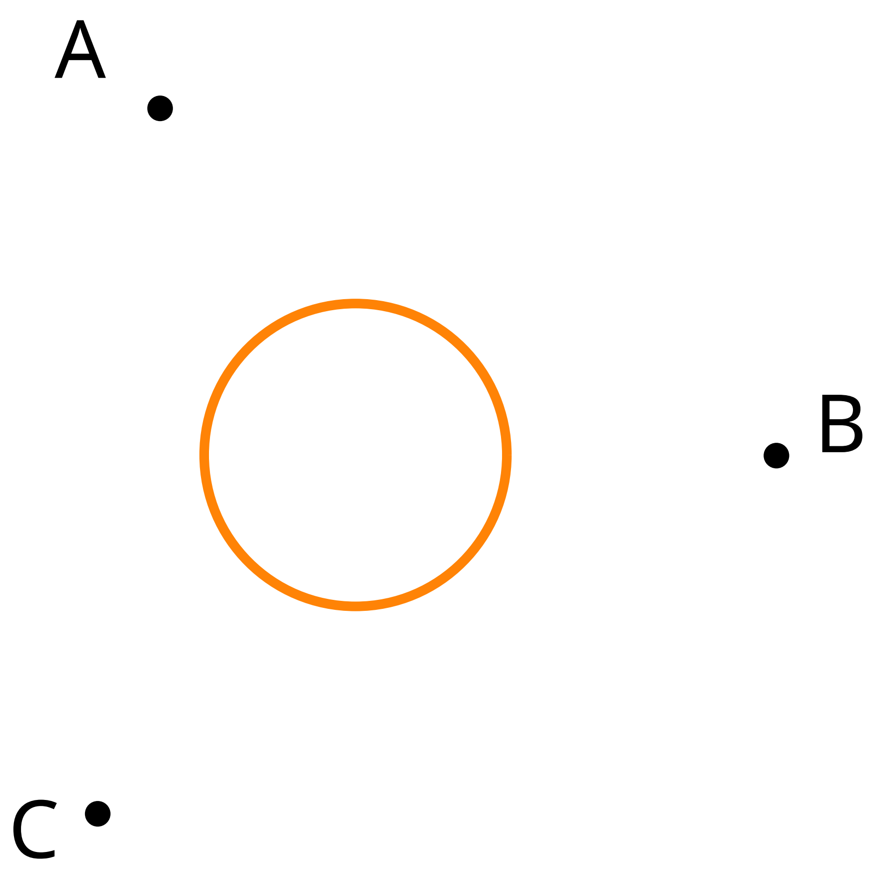
Example: disks as ranges

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D} =$ set of disks



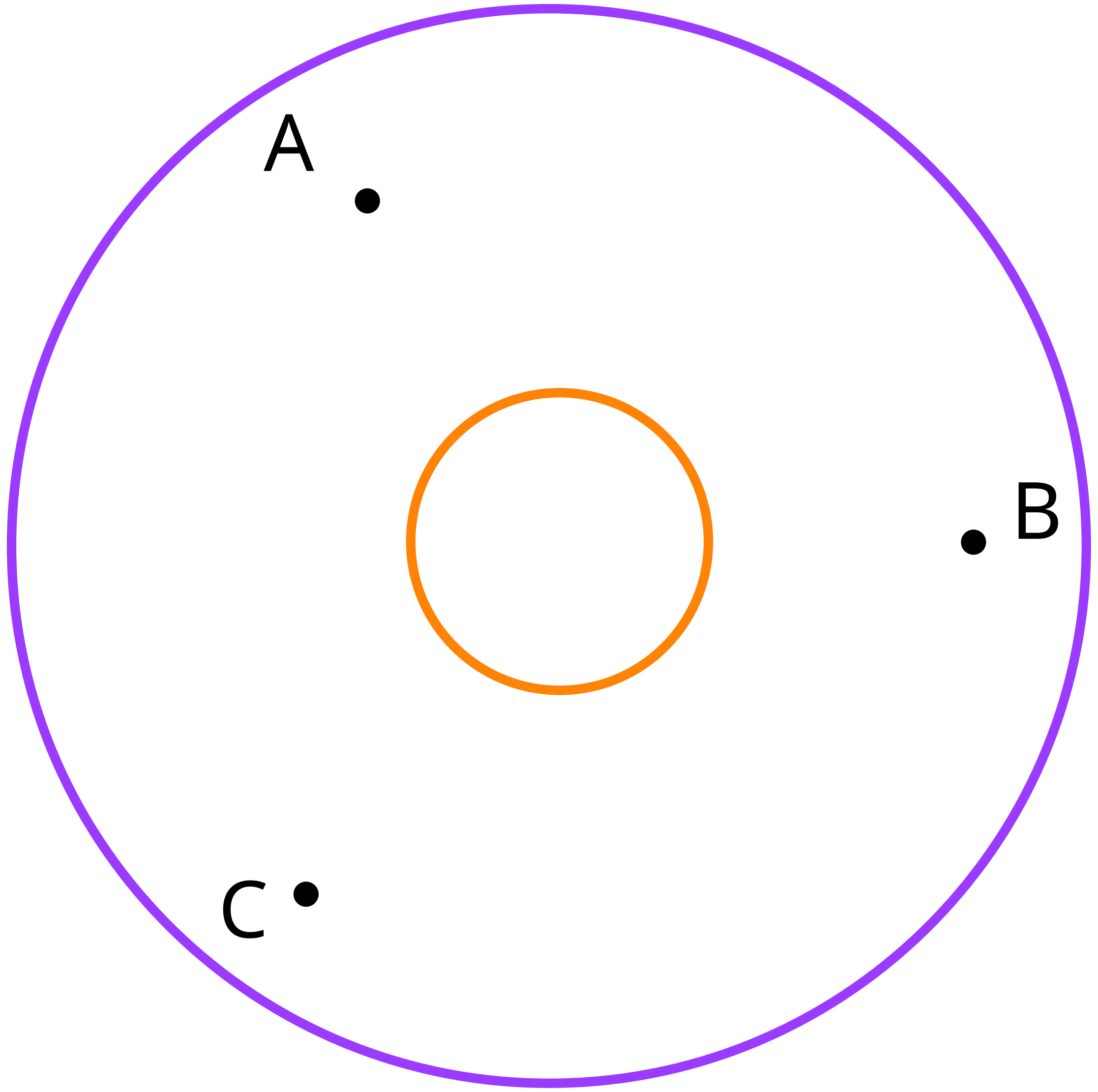
Example: disks as ranges

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D}$ = set of disks



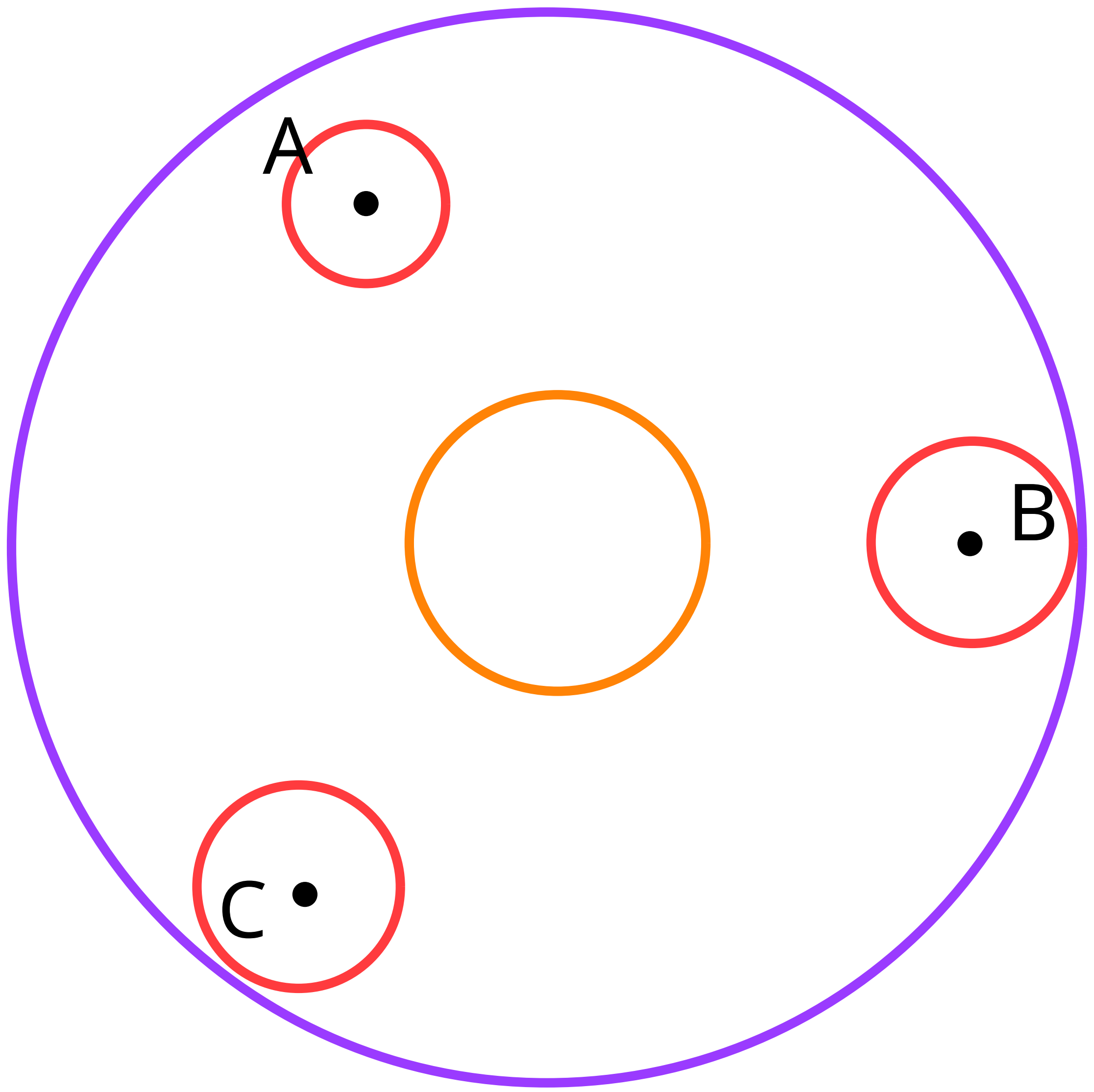
Example: disks as ranges

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D} =$ set of disks



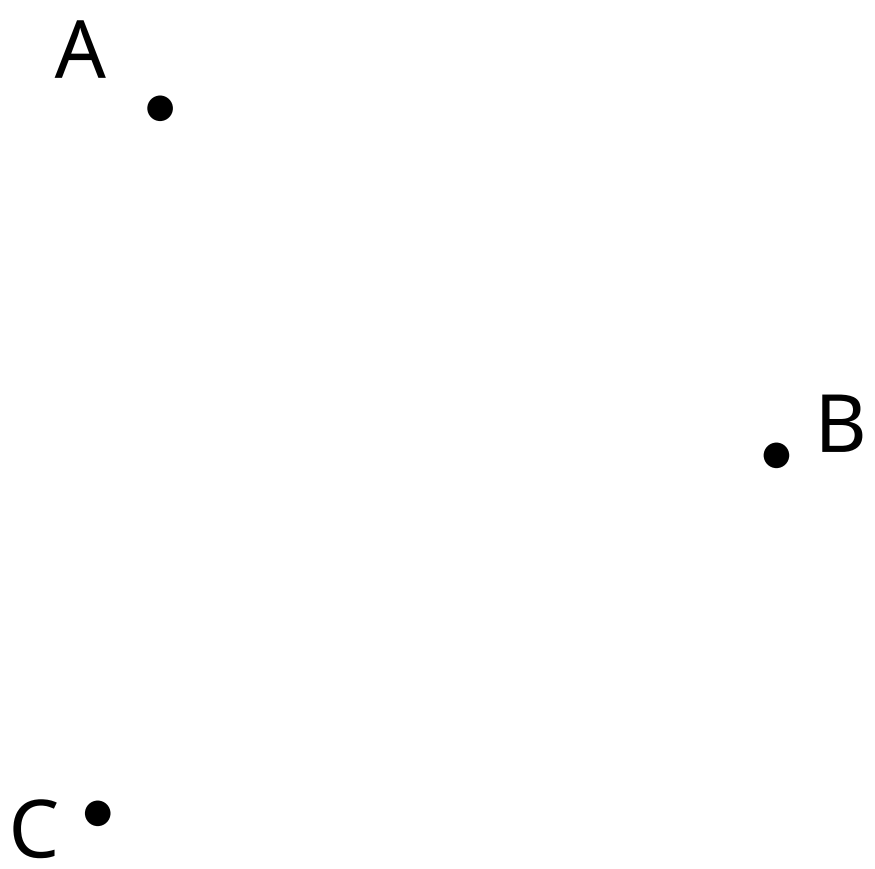
Example: disks as ranges

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D}$ = set of disks



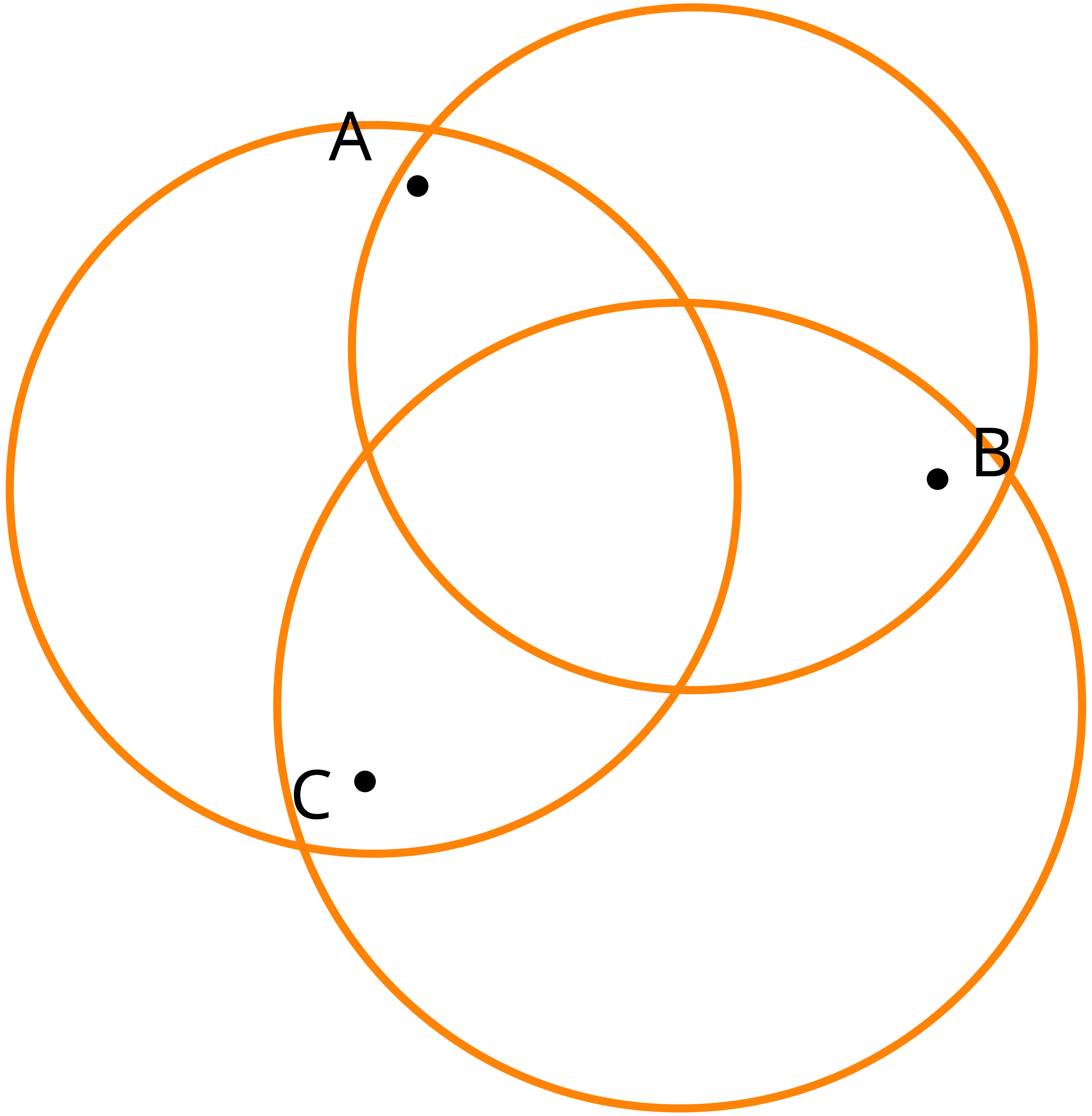
Example: disks as ranges

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D} =$ set of disks



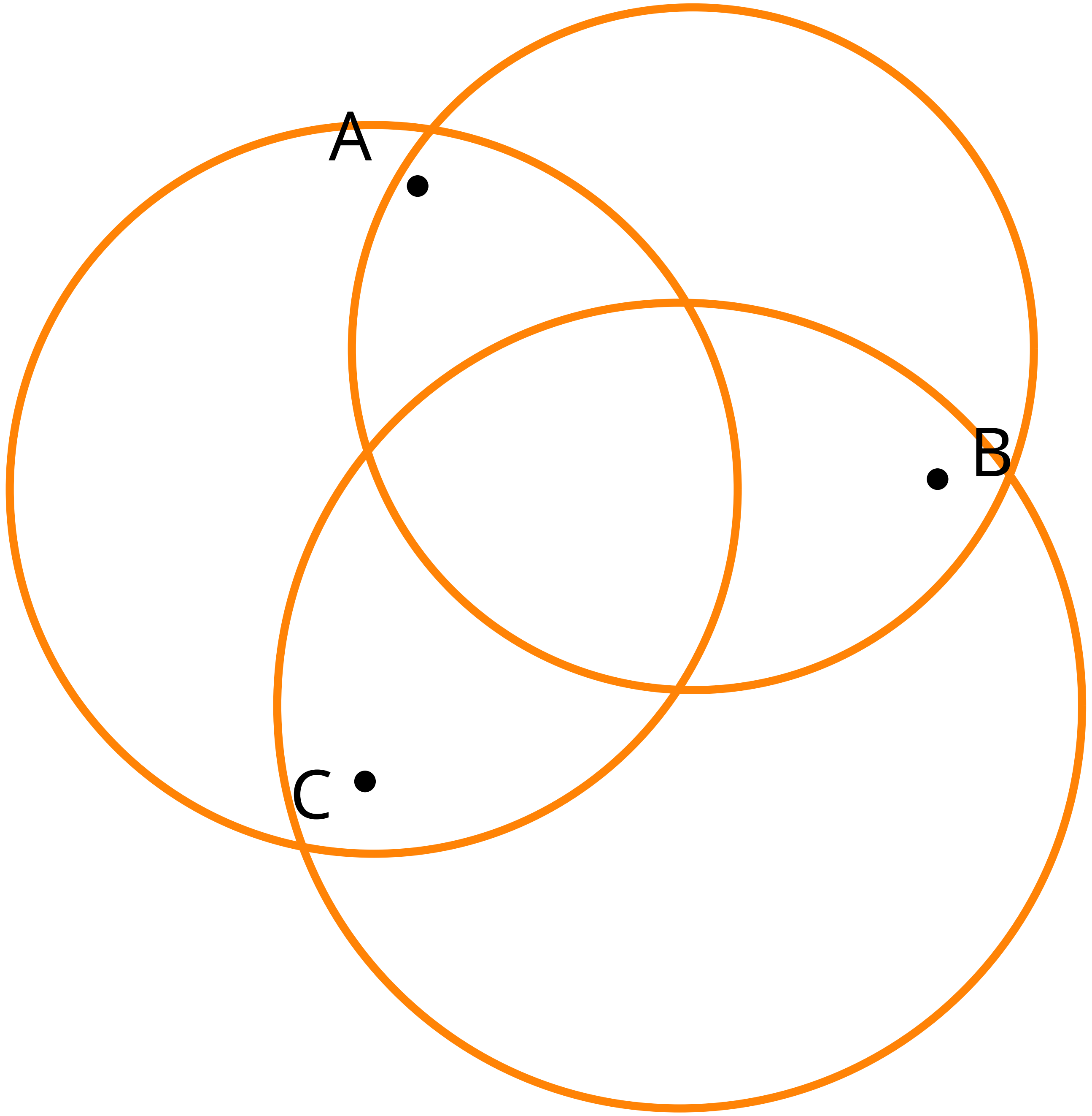
Example: disks as ranges

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D}$ = set of disks



Example: disks as ranges

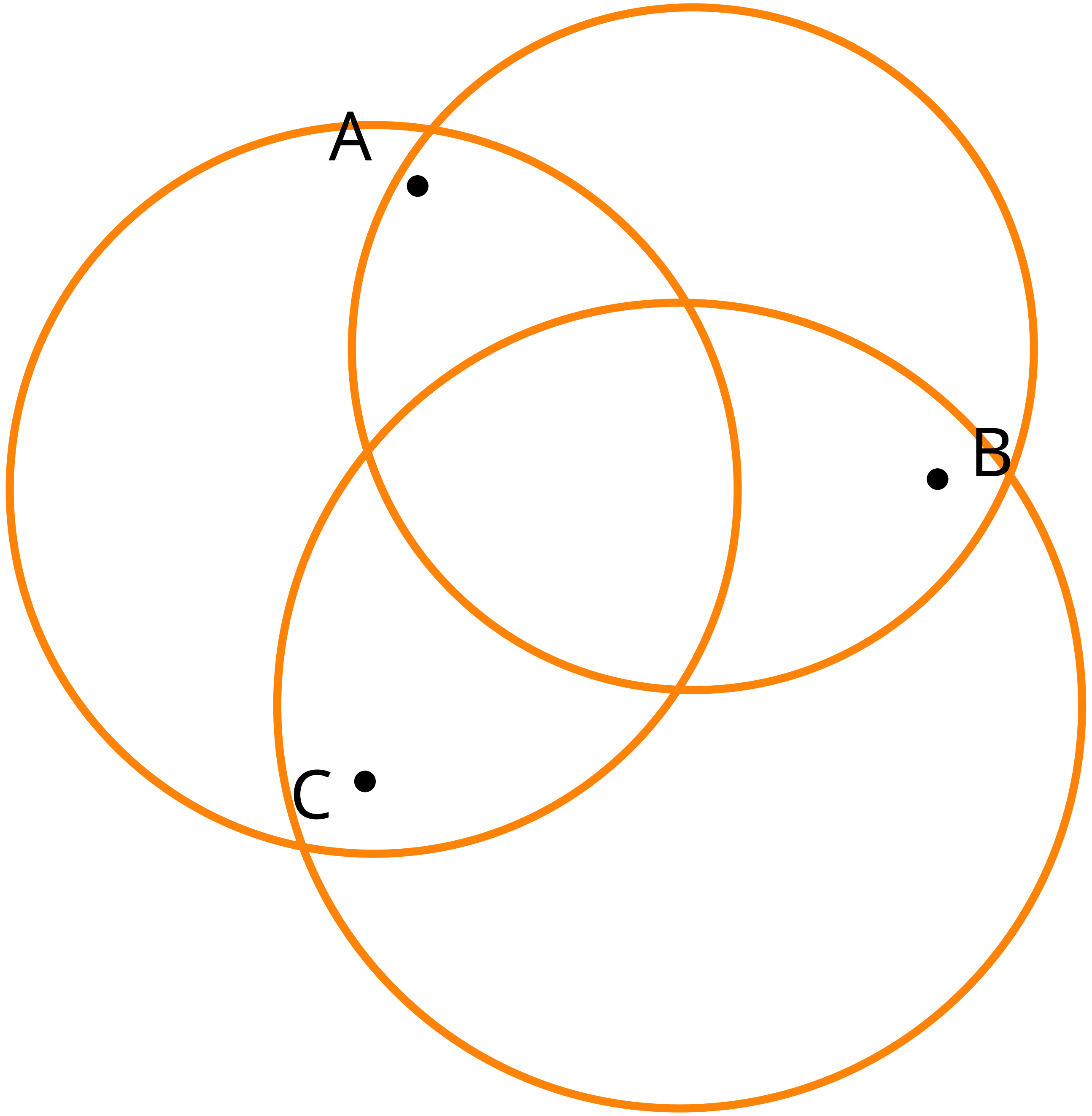
range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D} =$ set of disks



shatter !

Example: disks as ranges

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D} =$ set of disks

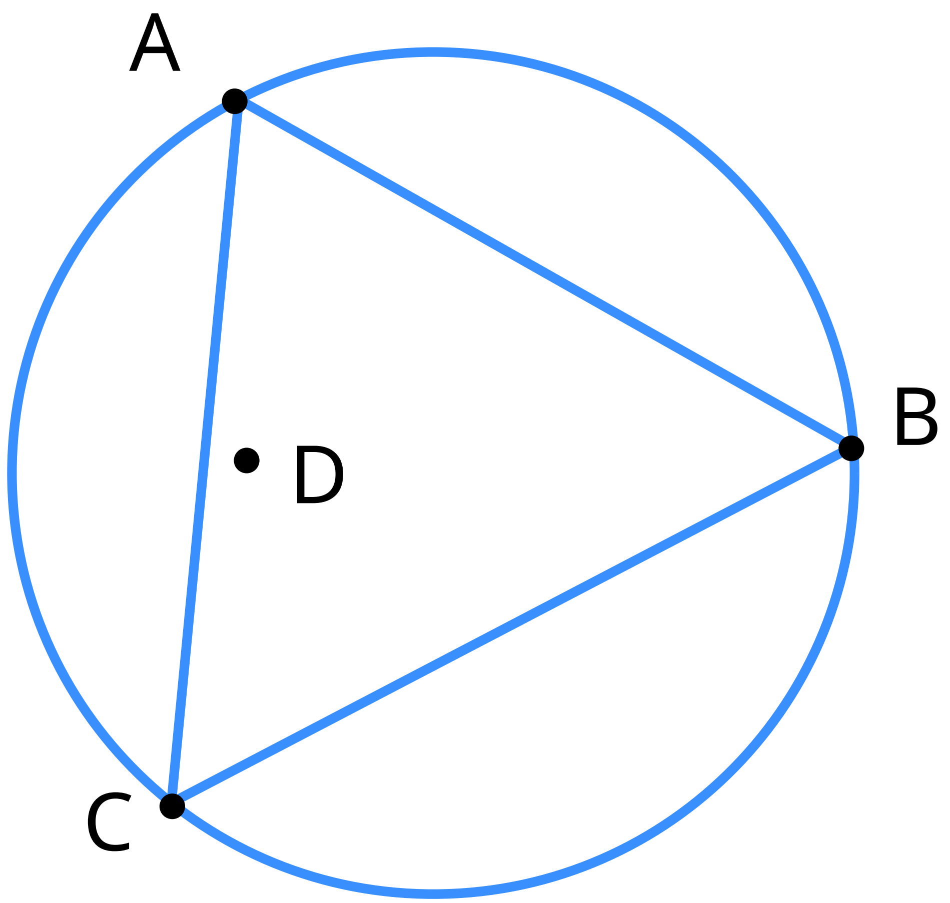


$\Rightarrow$ VC-dimension ≥ 3

shatter !

Example: disks as ranges

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D} =$ set of disks

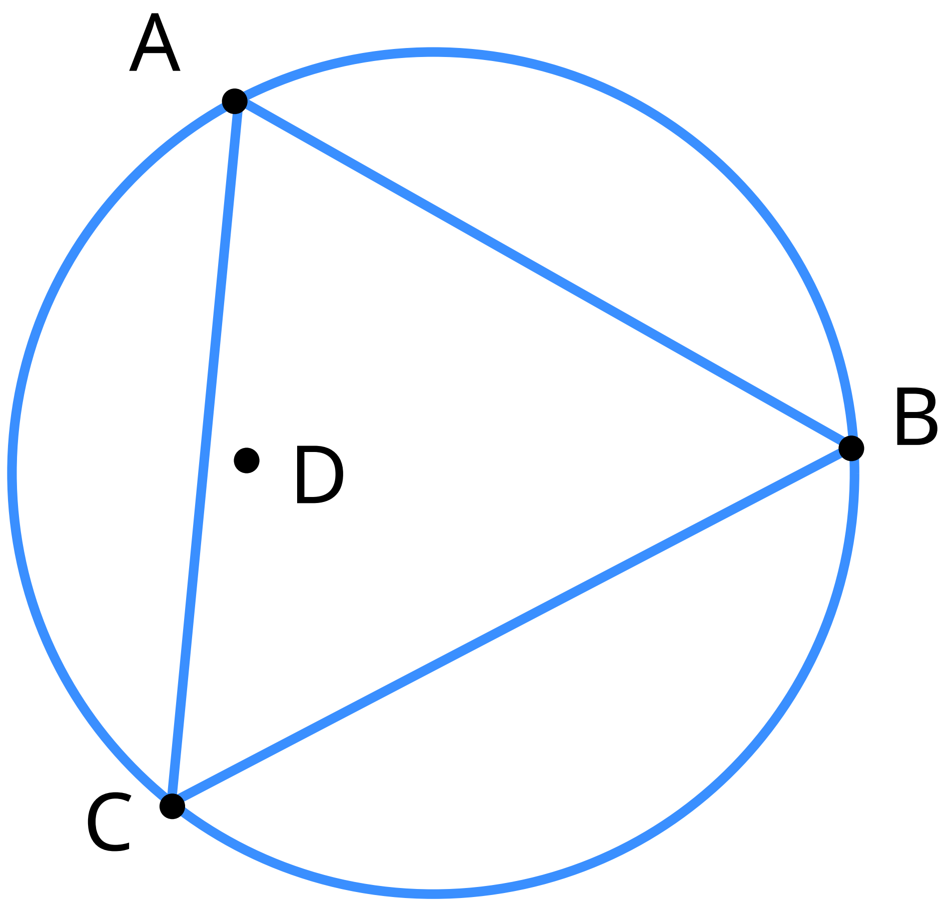


4 points

case 1: $D \in$ triangle(ABC)

Example: disks as ranges

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D} =$ set of disks



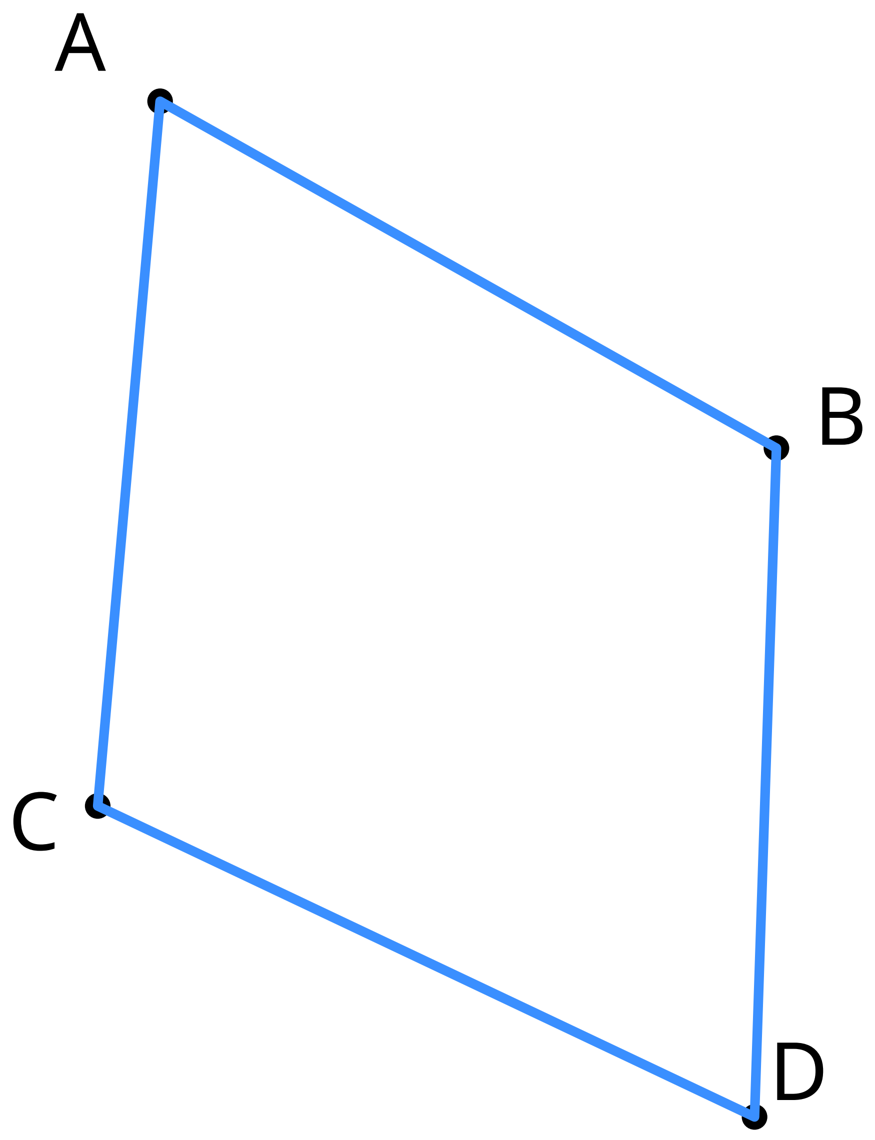
4 points

case 1: $D \in$ triangle(ABC)

not shatter !

Example: disks as ranges

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D} =$ set of disks



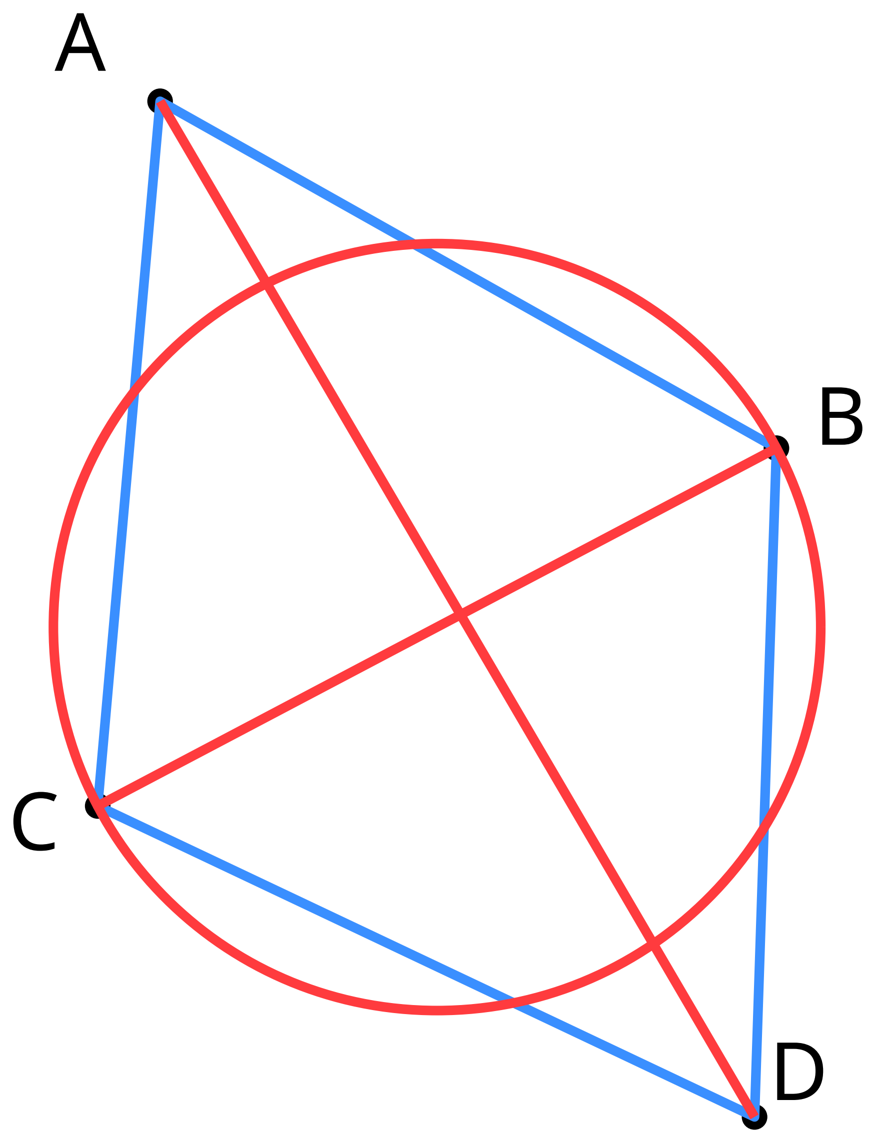
4 points

case 1: $D \in$ triangle(ABC)

case 2: $ABCD$ convex quadrilateral

Example: disks as ranges

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D}$ = set of disks



4 points

case 1: $D \in$ triangle(ABC)

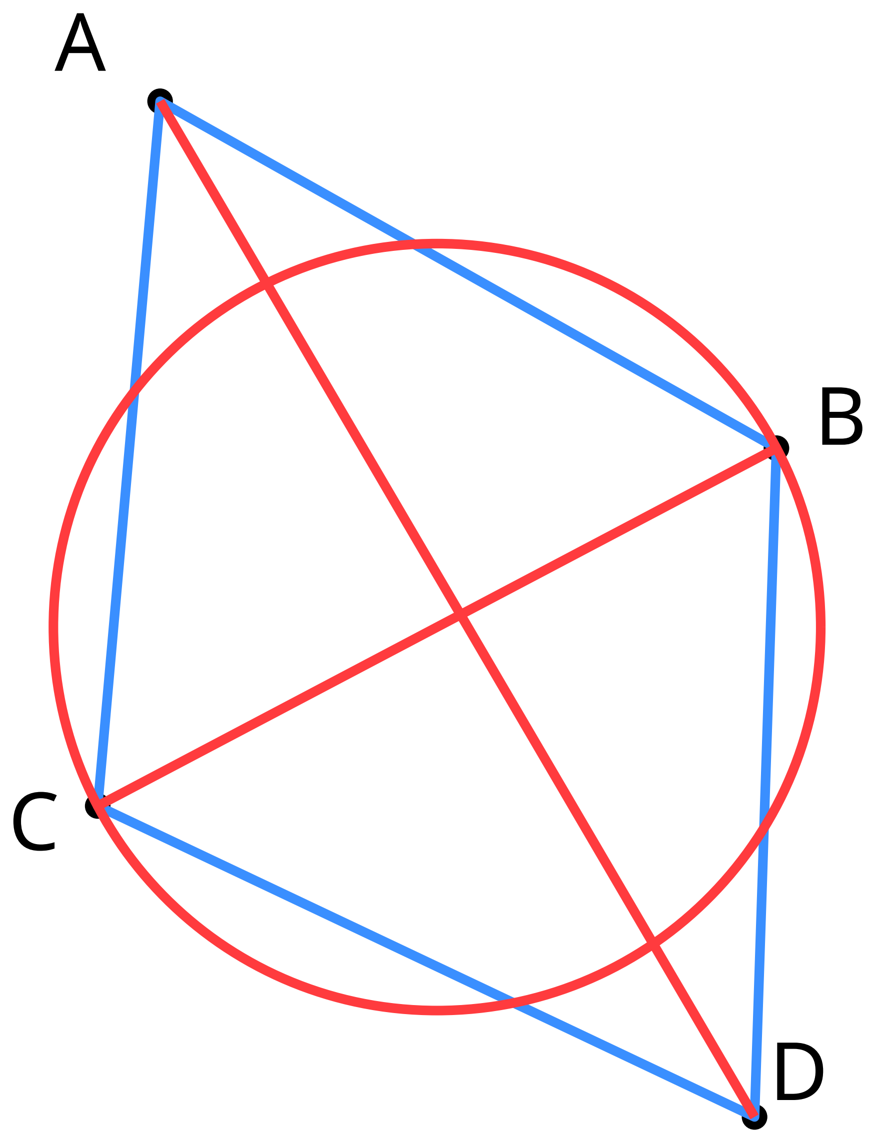
case 2: $ABCD$ convex quadrilateral

without proof:

can't get $\{A, D\}$ and $\{B, C\}$

Example: disks as ranges

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D}$ = set of disks



4 points

case 1: $D \in$ triangle(ABC)

case 2: $ABCD$ convex quadrilateral

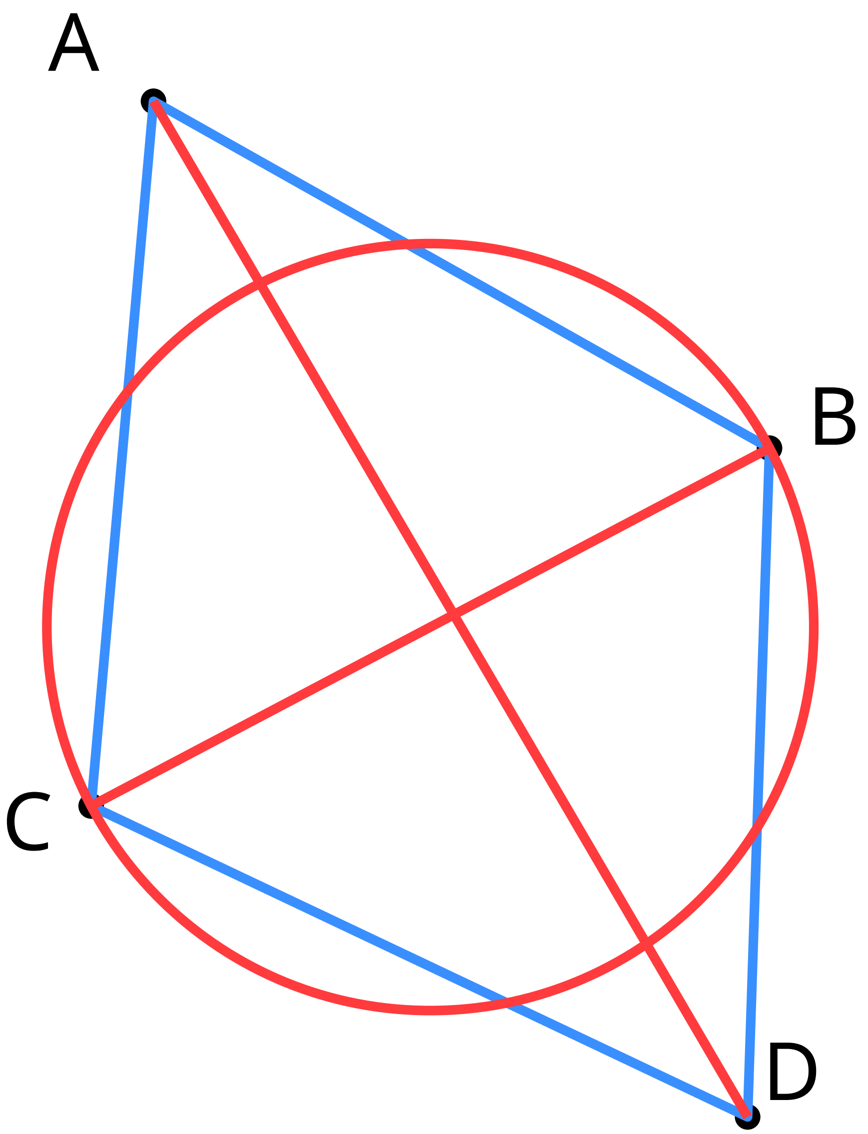
without proof:

can't get $\{A, D\}$ and $\{B, C\}$

not shatter !

Example: disks as ranges

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D}$ = set of disks



4 points

case 1: $D \in$ triangle(ABC)

case 2: $ABCD$ convex quadrilateral

without proof:

can't get $\{A, D\}$ and $\{B, C\}$

$\Rightarrow$ VC-dimension $= 3$

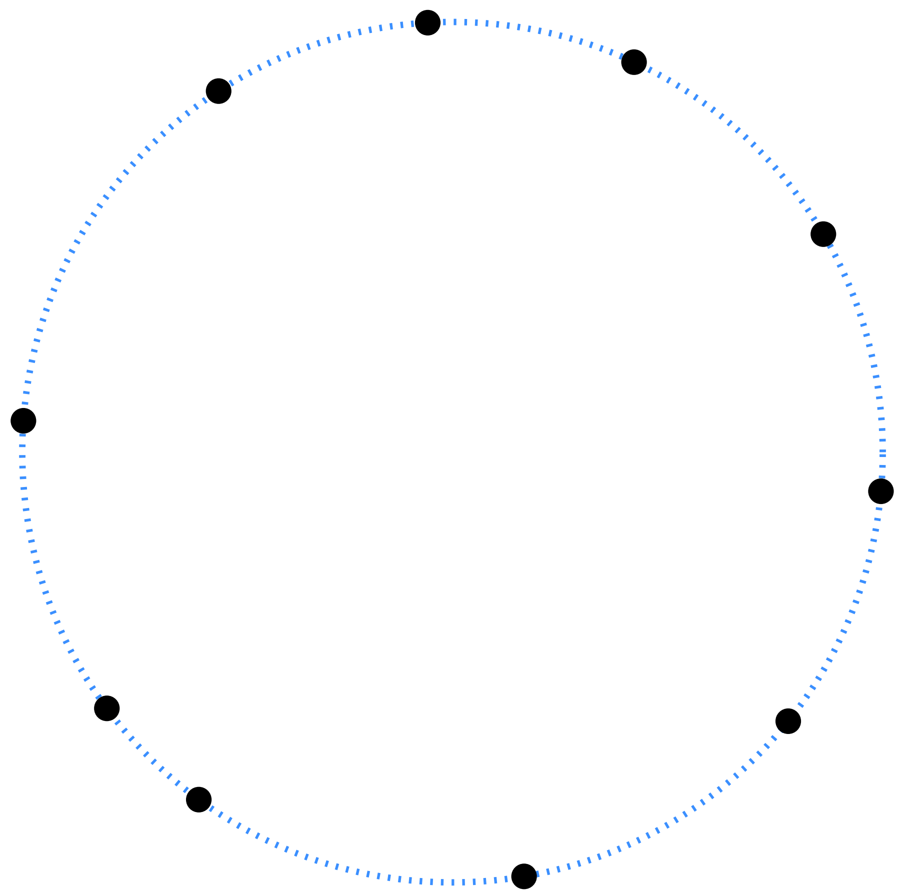
not shatter !

Example: convex sets as ranges

range space $(\mathbb{R}^2, \mathcal{C})$, with $\mathcal{C}$ $=$ set of closed convex sets

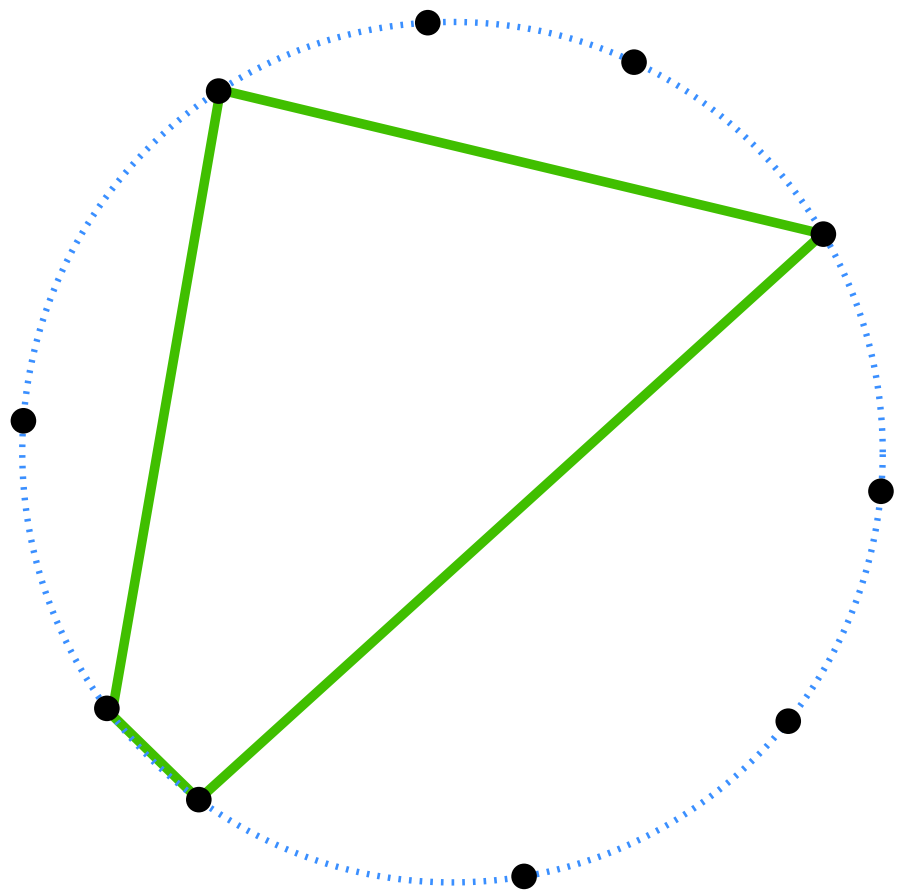
Example: convex sets as ranges

range space $(\mathbb{R}^2, \mathcal{C})$, with $\mathcal{C}$ = set of closed convex sets



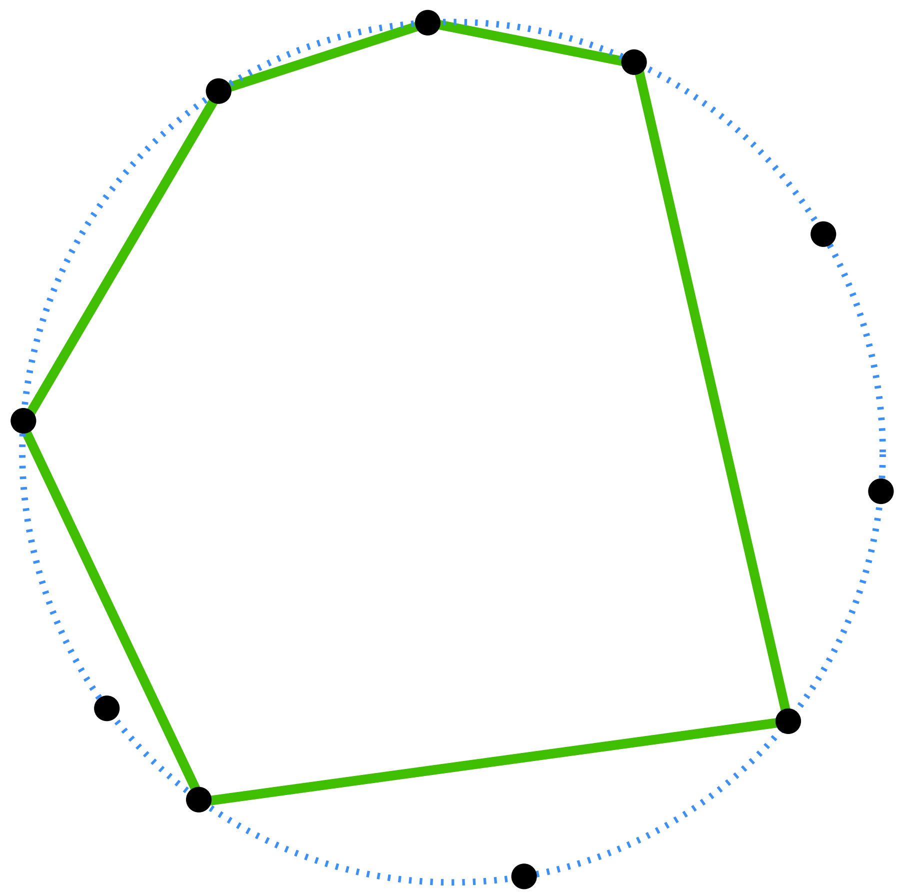
Example: convex sets as ranges

range space $(\mathbb{R}^2, \mathcal{C})$, with $\mathcal{C}$ = set of closed convex sets



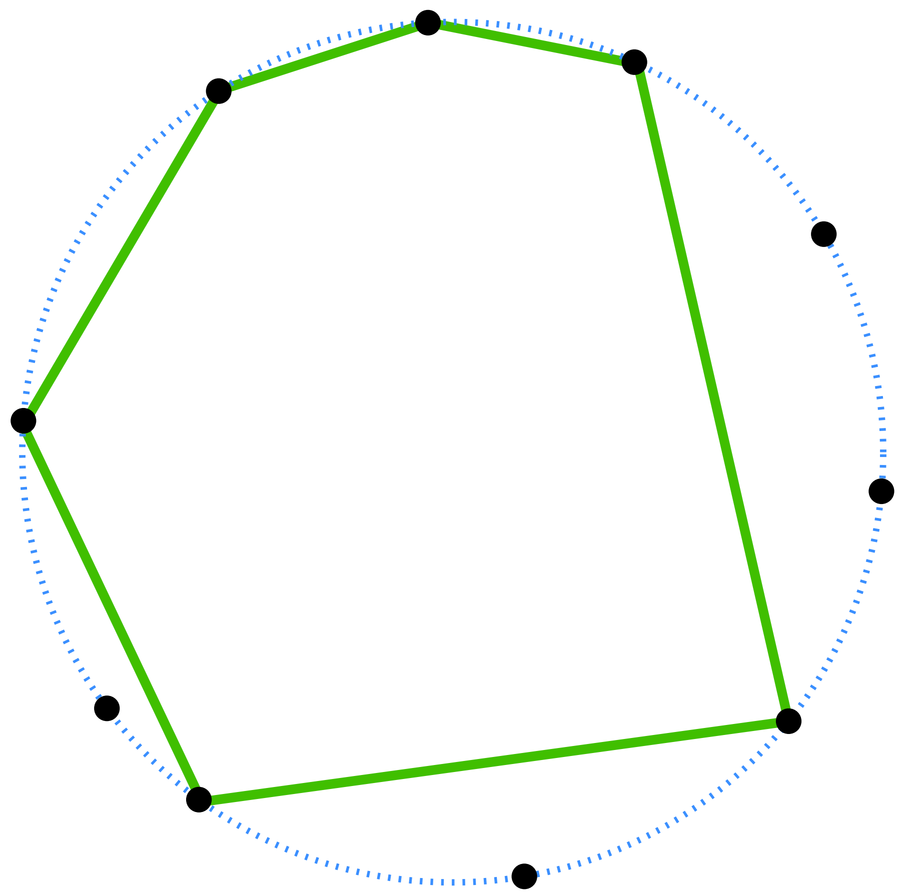
Example: convex sets as ranges

range space $(\mathbb{R}^2, \mathcal{C})$, with $\mathcal{C} =$ set of closed convex sets



Example: convex sets as ranges

range space $(\mathbb{R}^2, \mathcal{C})$, with $\mathcal{C}$ = set of closed convex sets



$\Rightarrow$ VC-dimension $= \infty$

Quiz

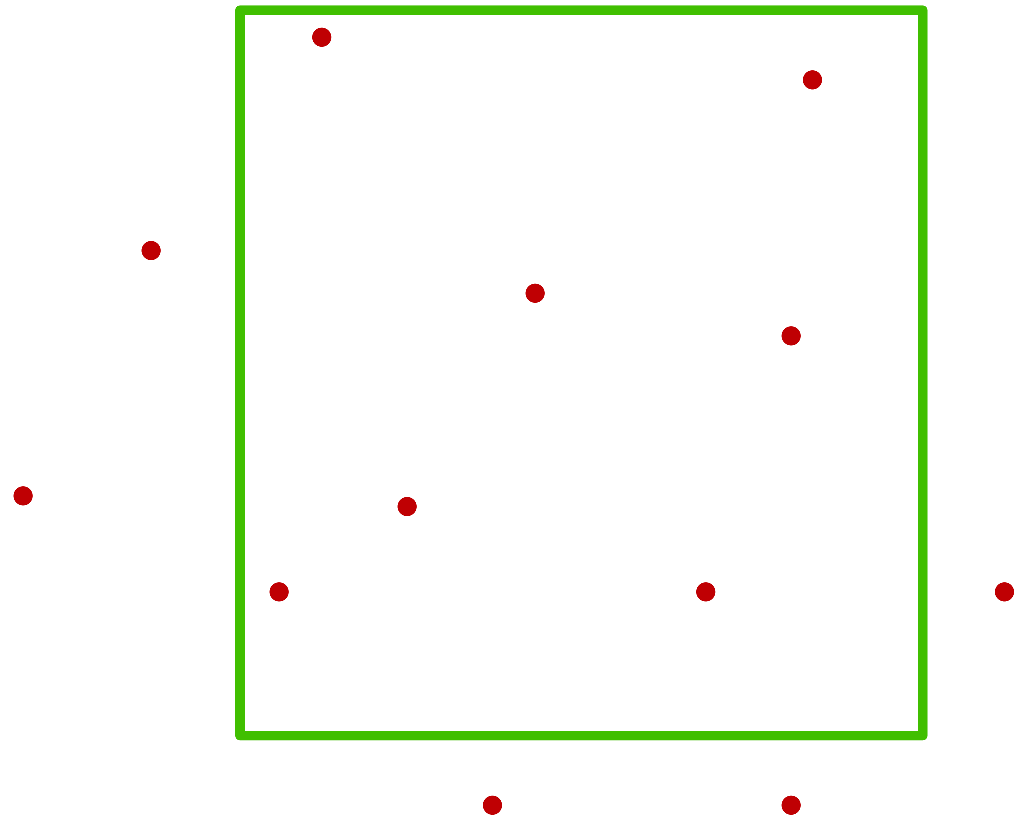
range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles

What is its VC-dimension?

A 4

B 5

C ∞

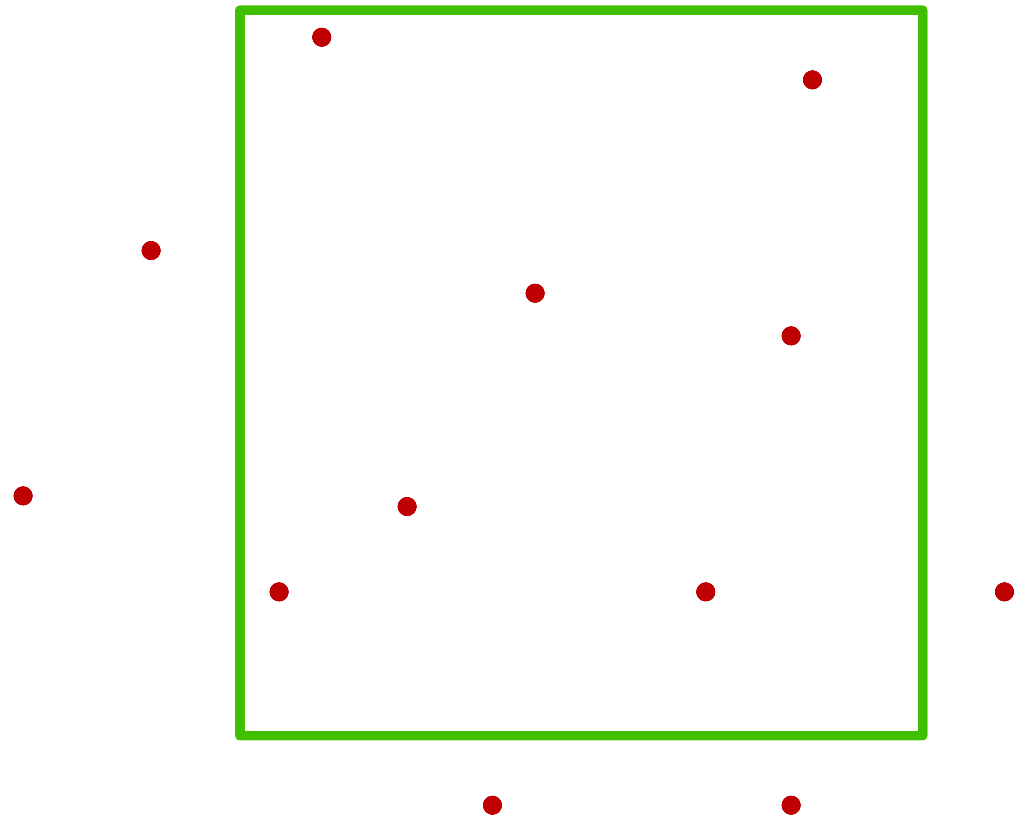


Quiz

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles

What is its VC-dimension?

A	4
B	5
C	∞

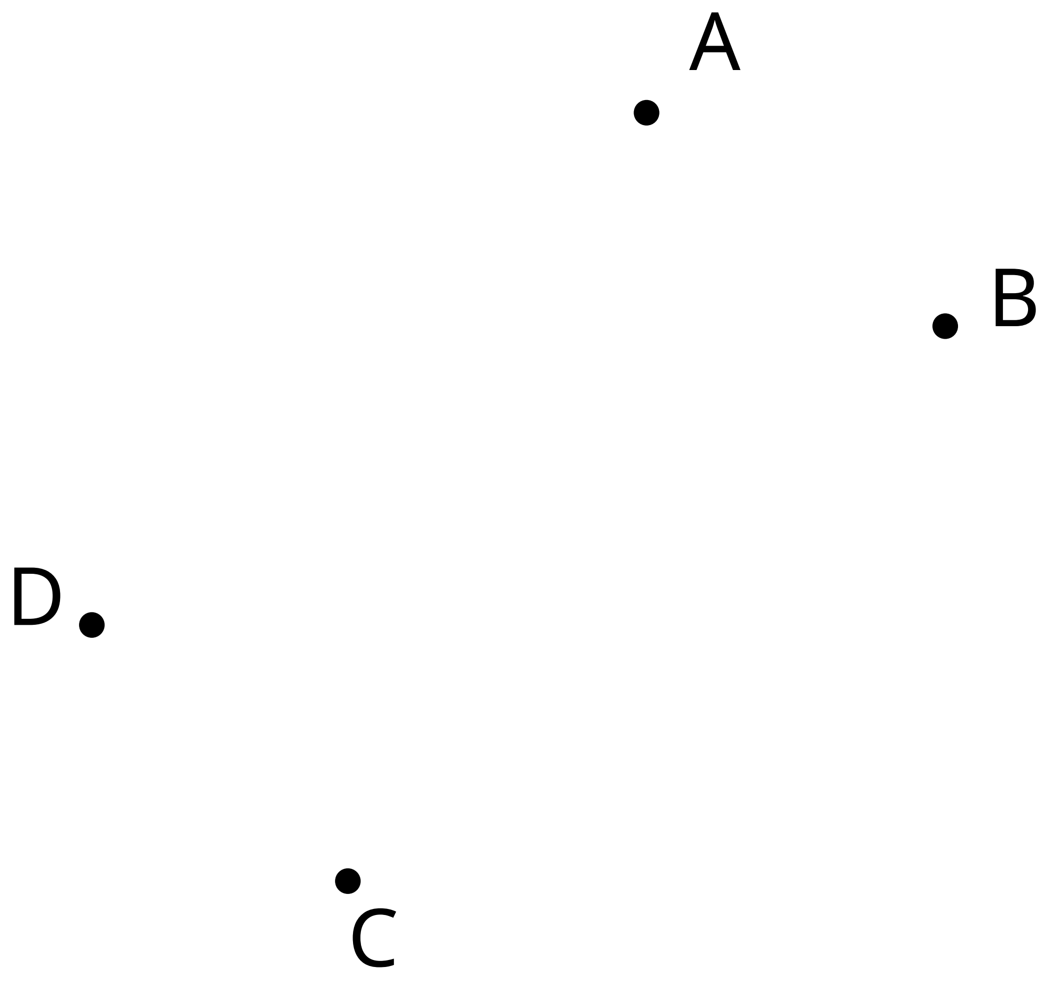


Example: rectangles as ranges

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR} =$ set of axis-aligned rectangles

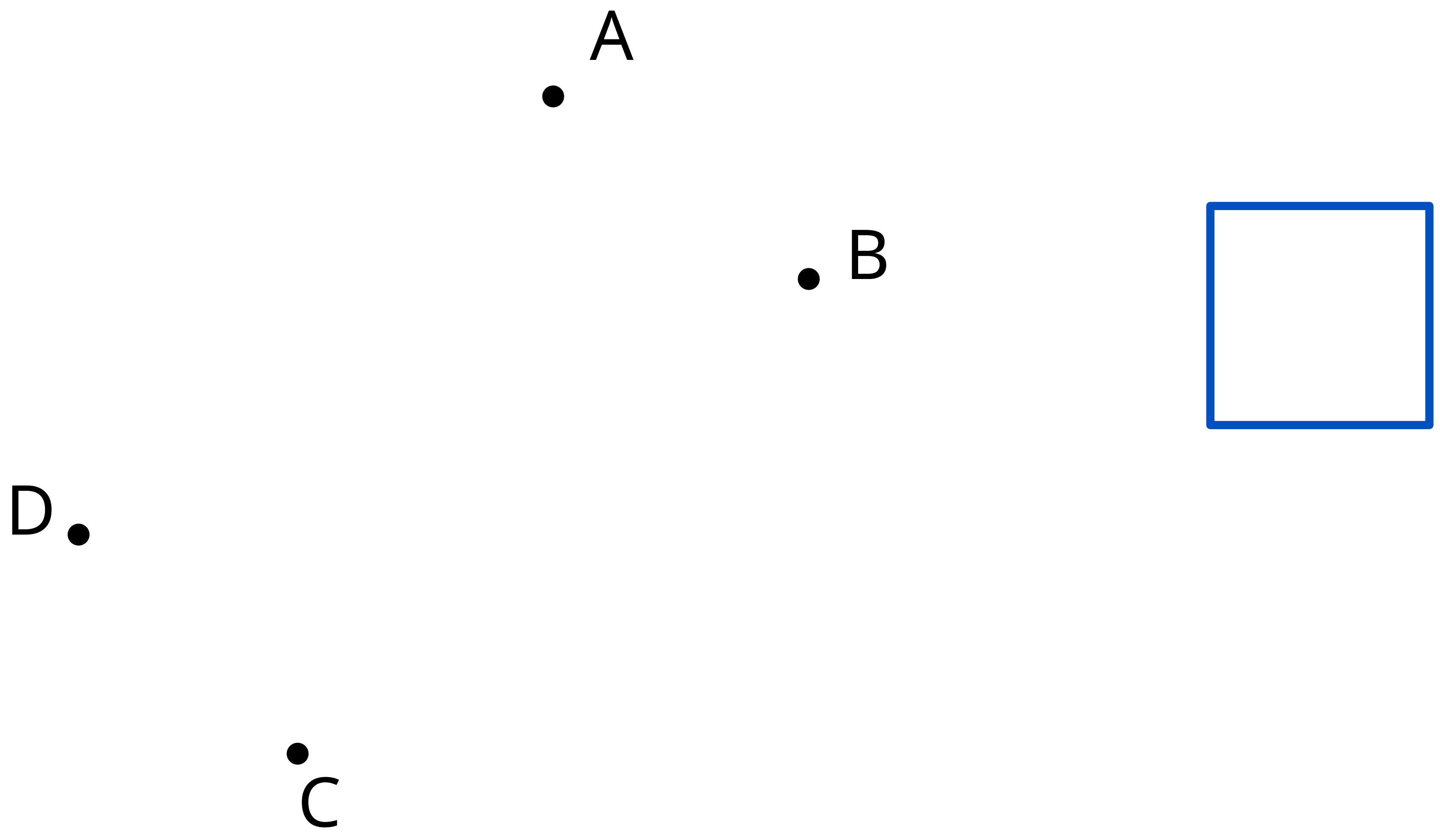
Example: rectangles as ranges

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR} =$ set of axis-aligned rectangles



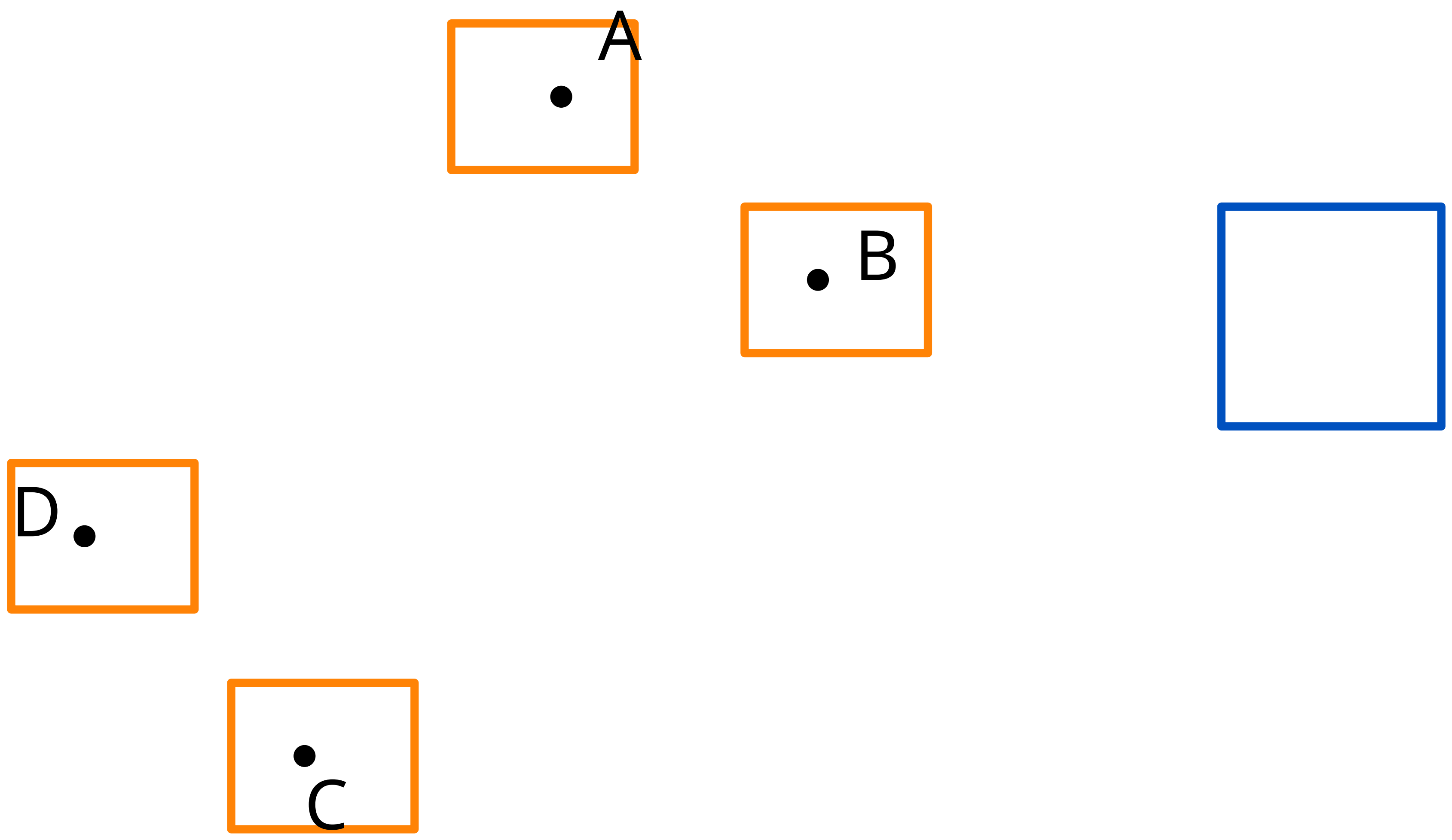
Example: rectangles as ranges

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles



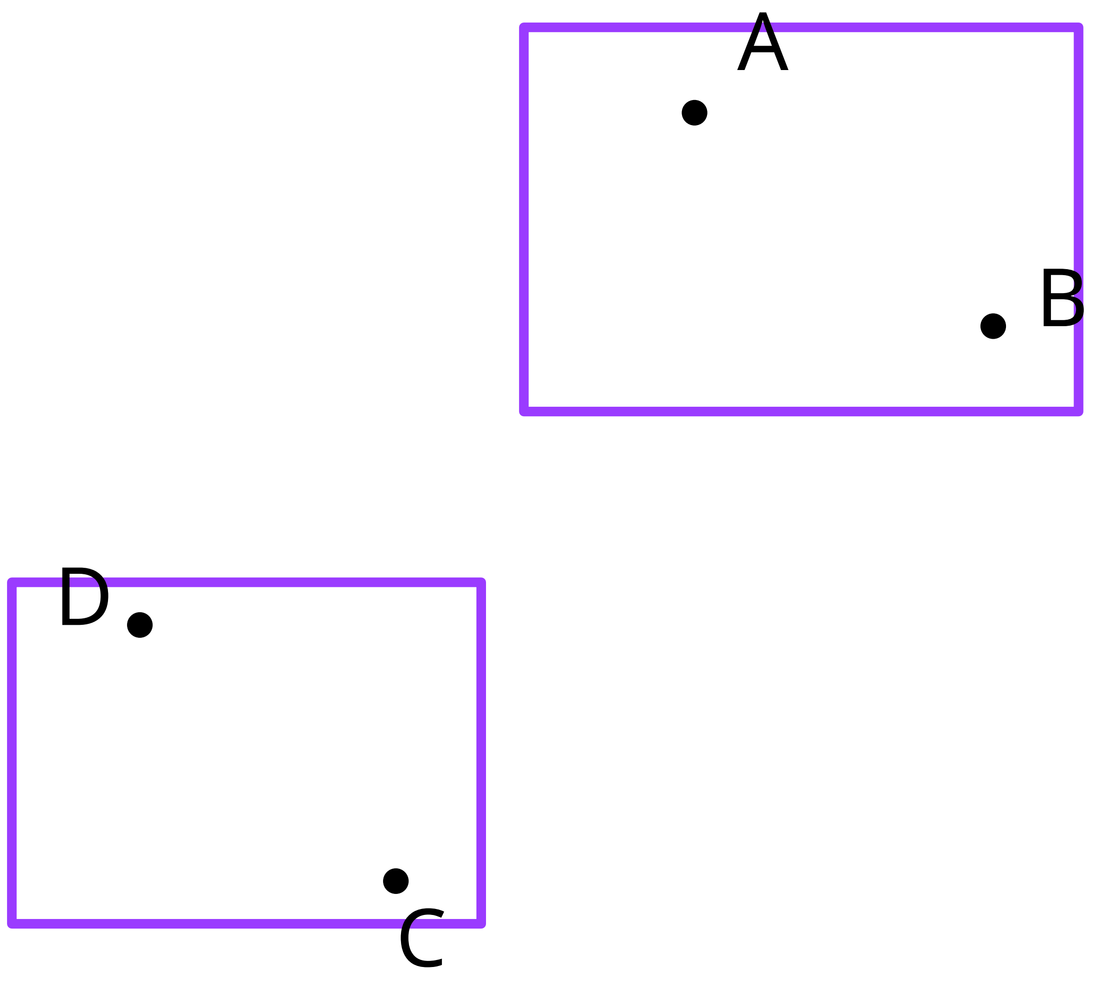
Example: rectangles as ranges

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles



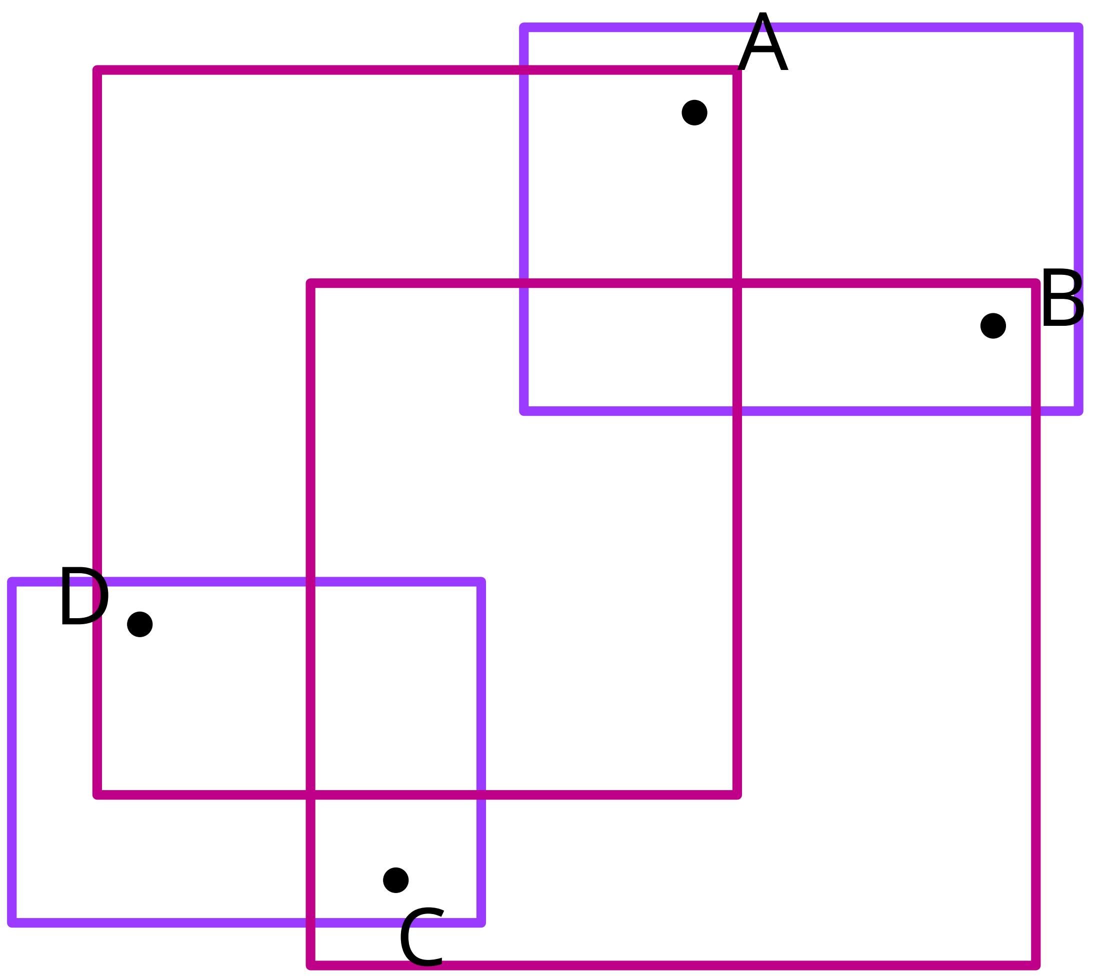
Example: rectangles as ranges

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles



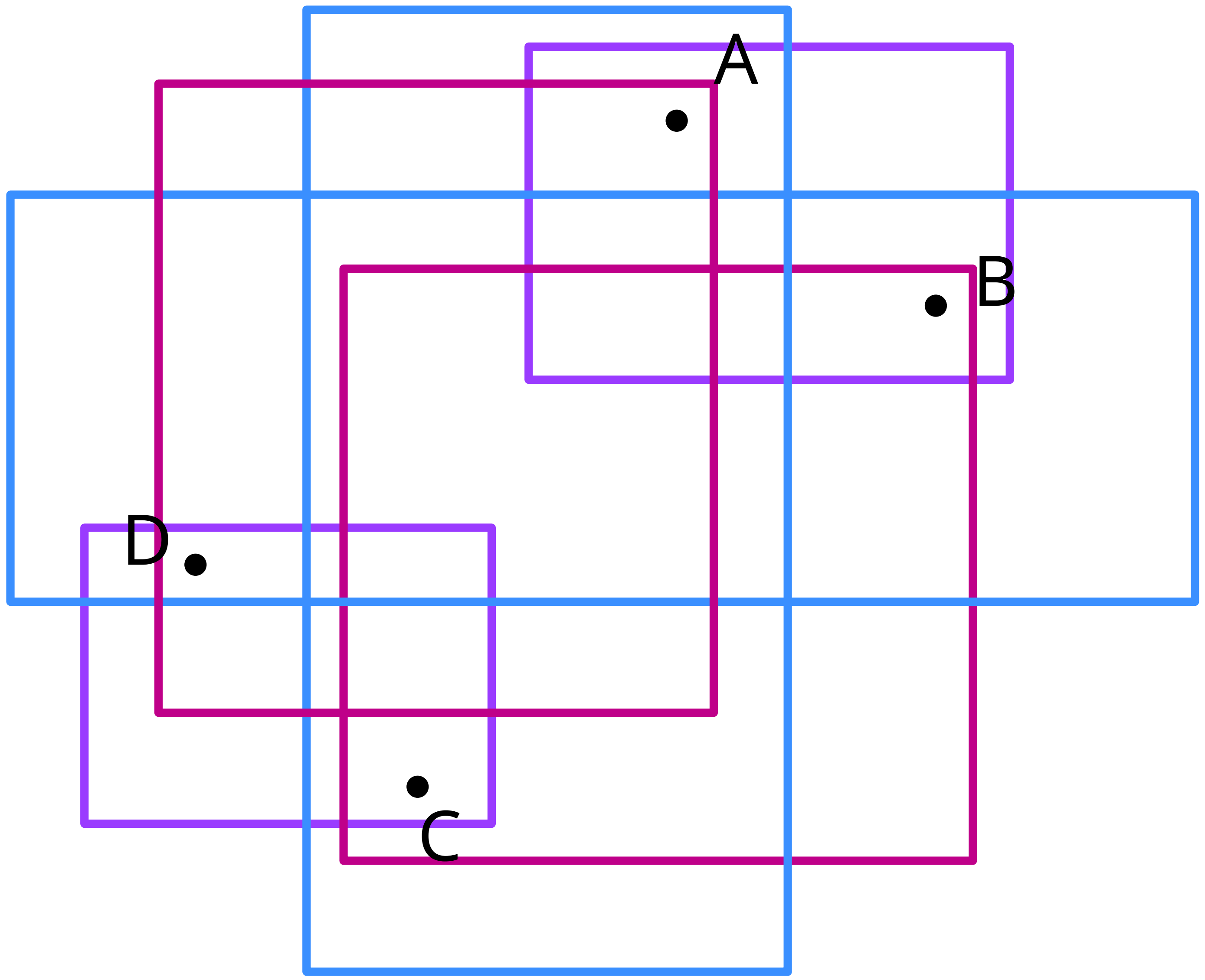
Example: rectangles as ranges

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles



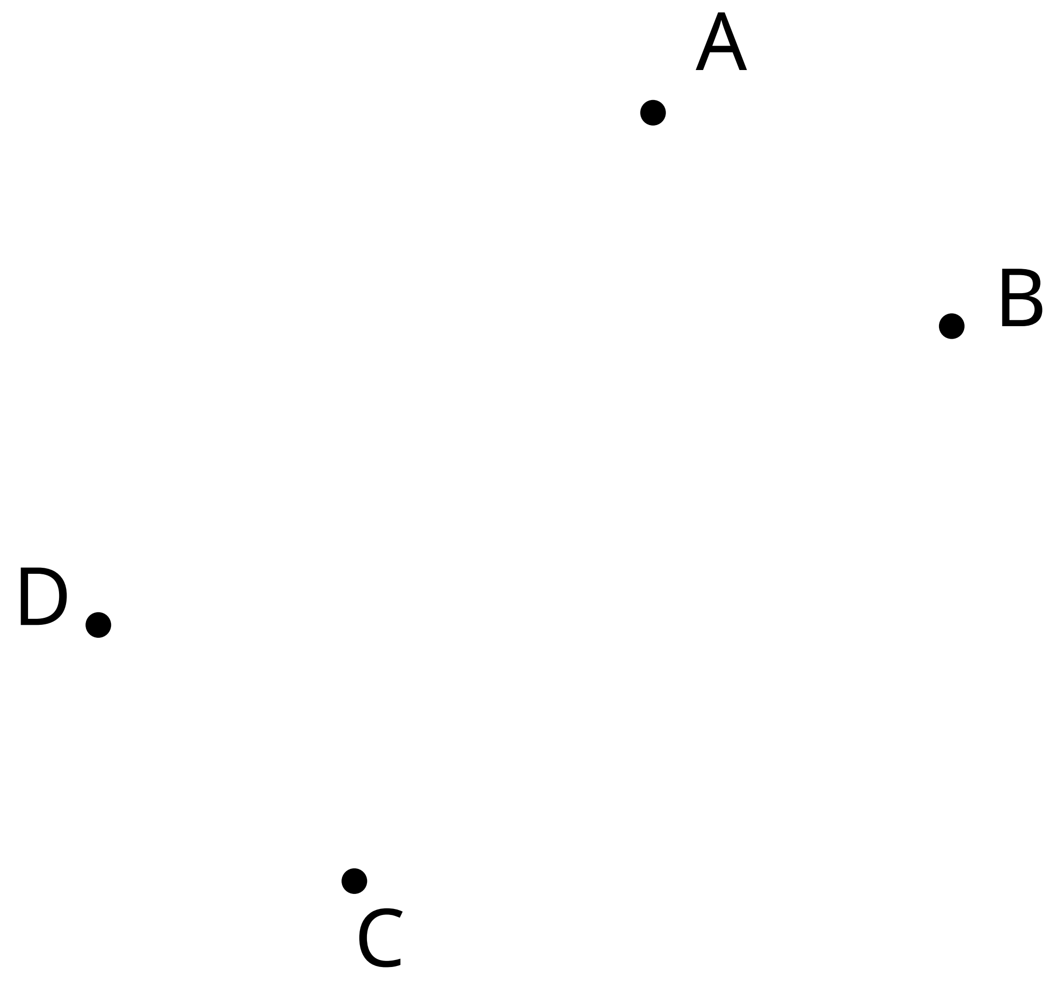
Example: rectangles as ranges

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles



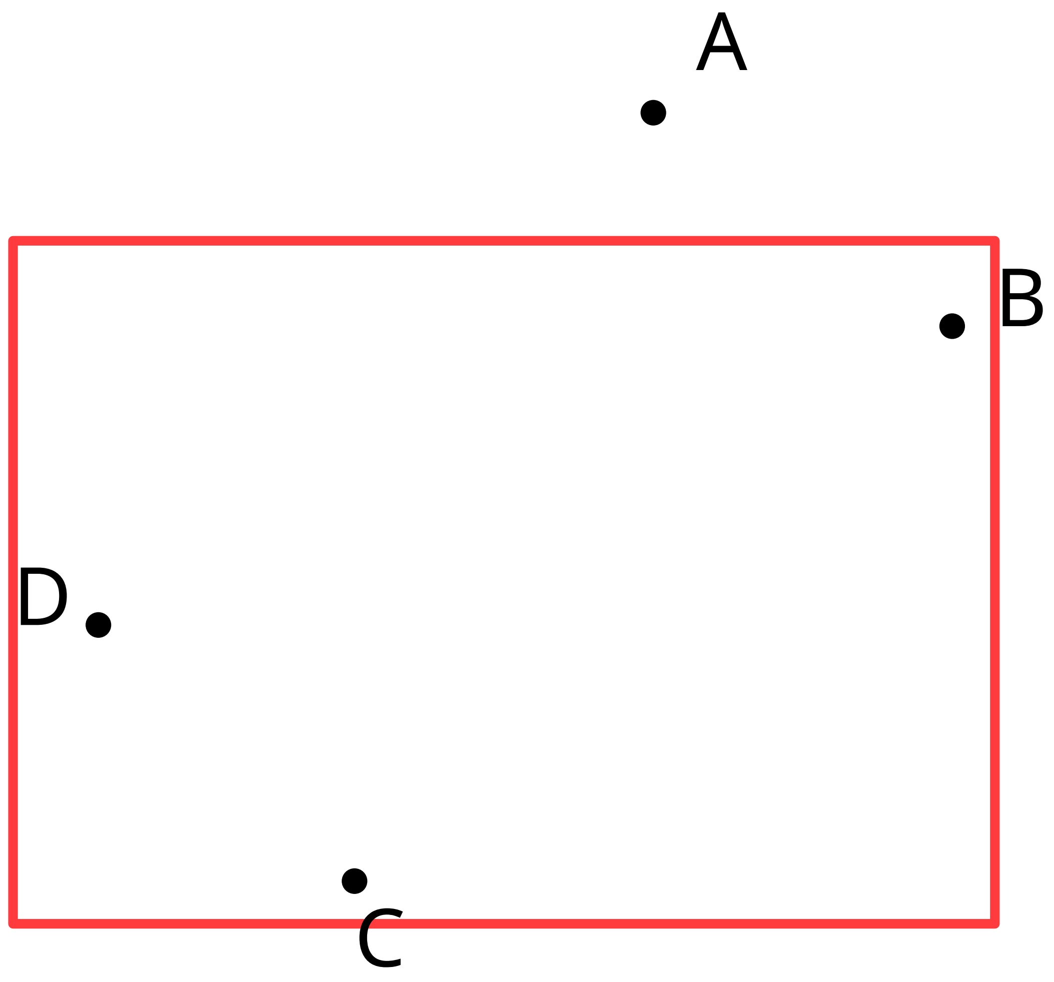
Example: rectangles as ranges

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR} =$ set of axis-aligned rectangles



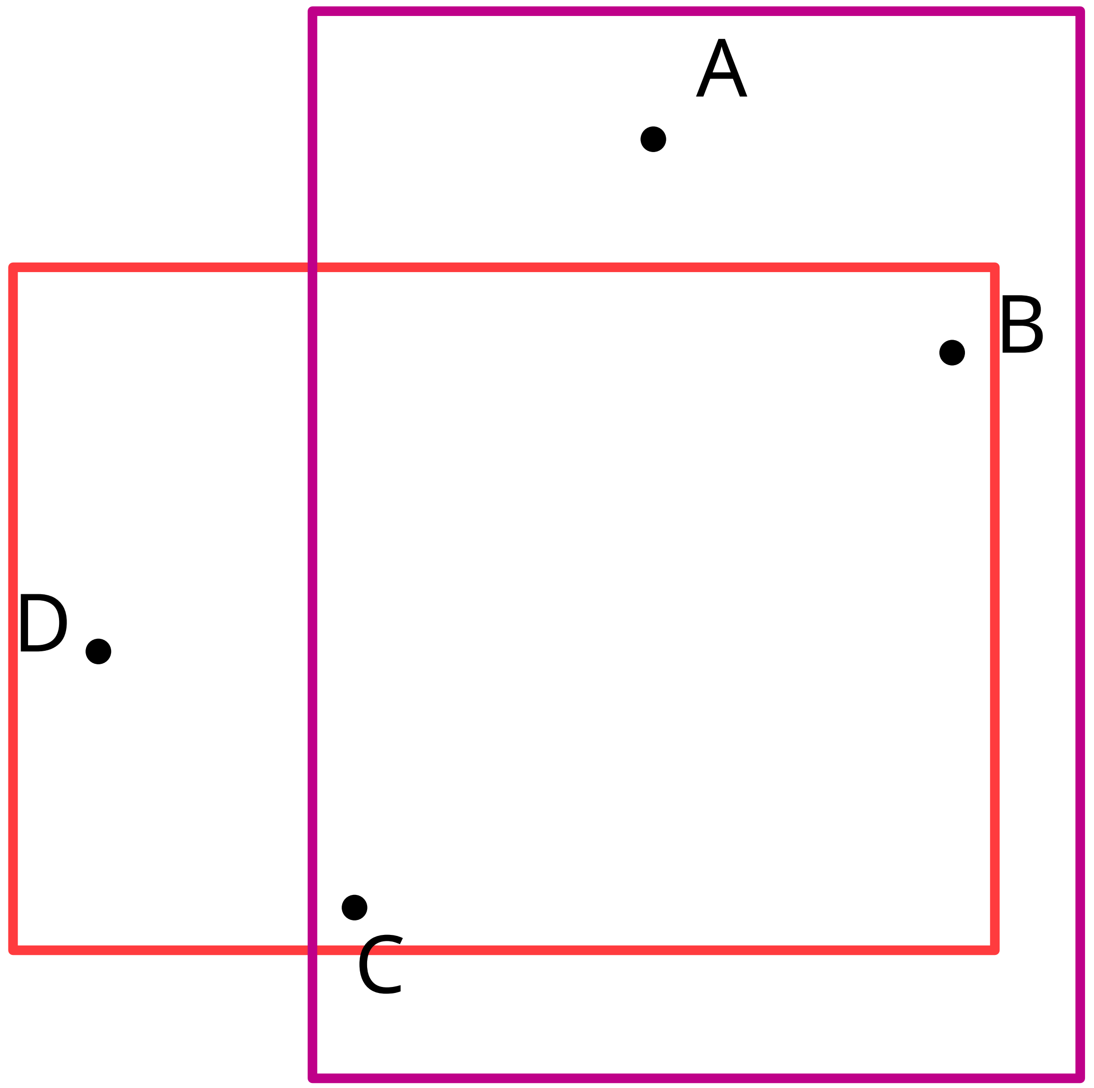
Example: rectangles as ranges

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles



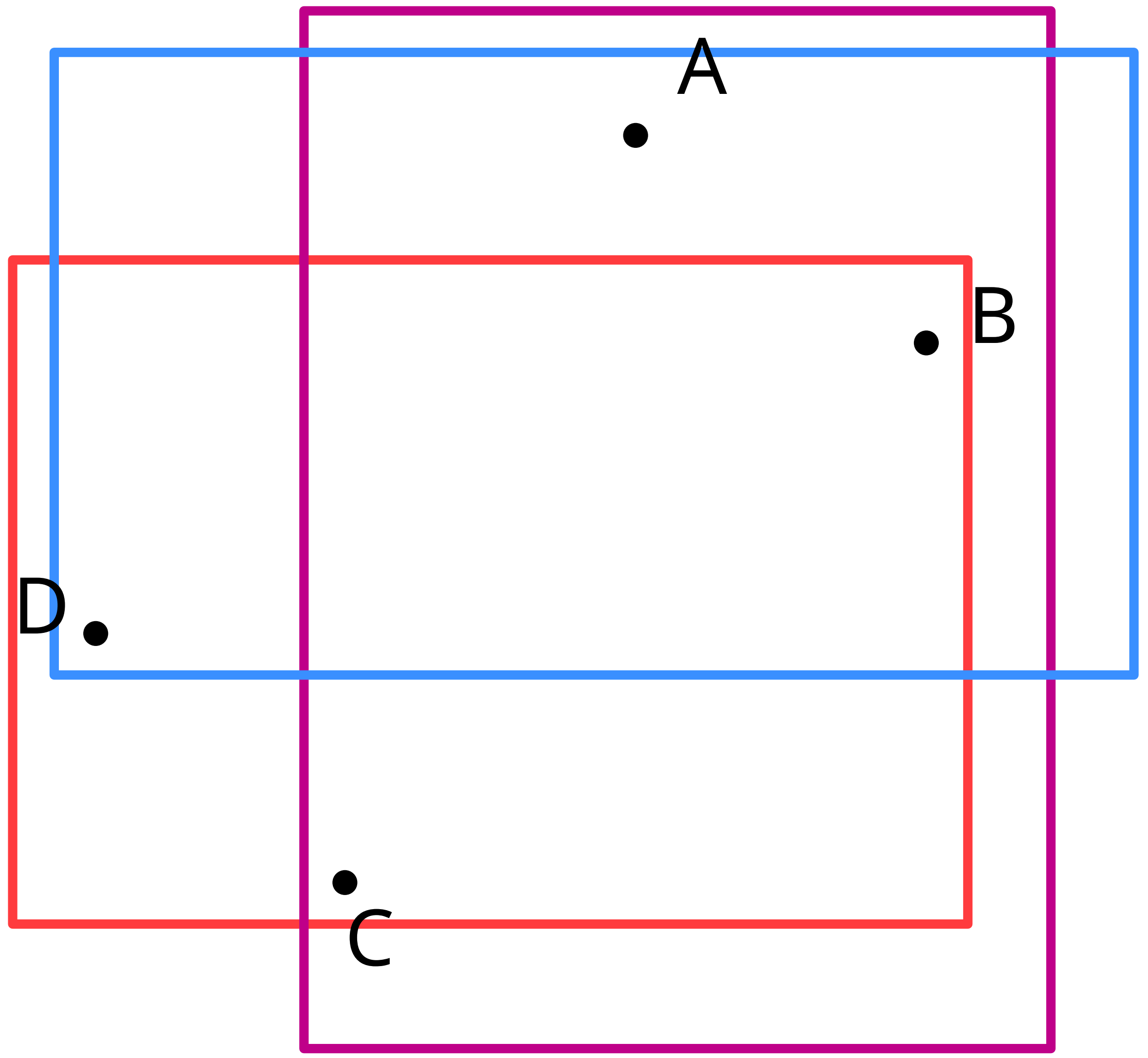
Example: rectangles as ranges

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles



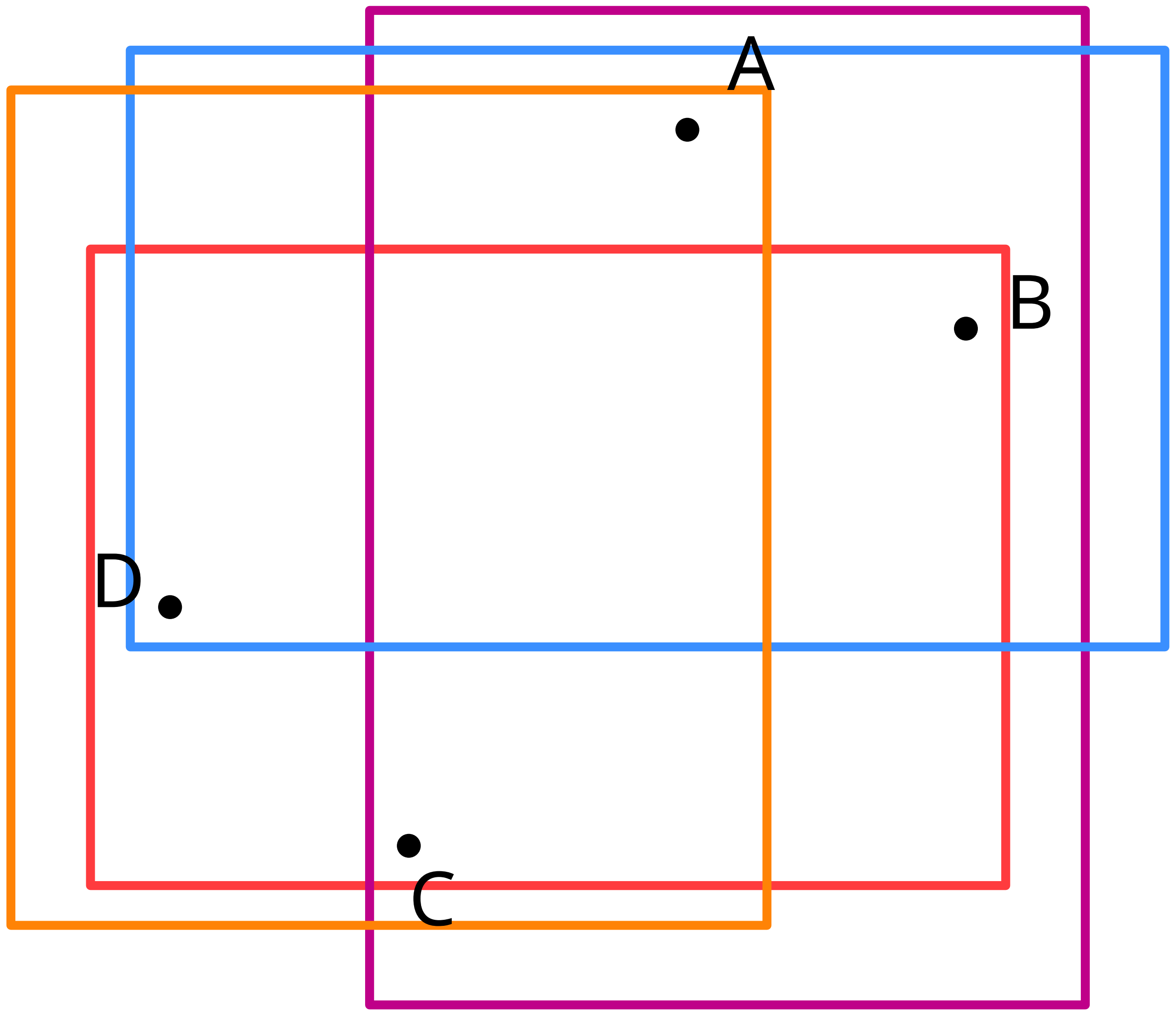
Example: rectangles as ranges

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles



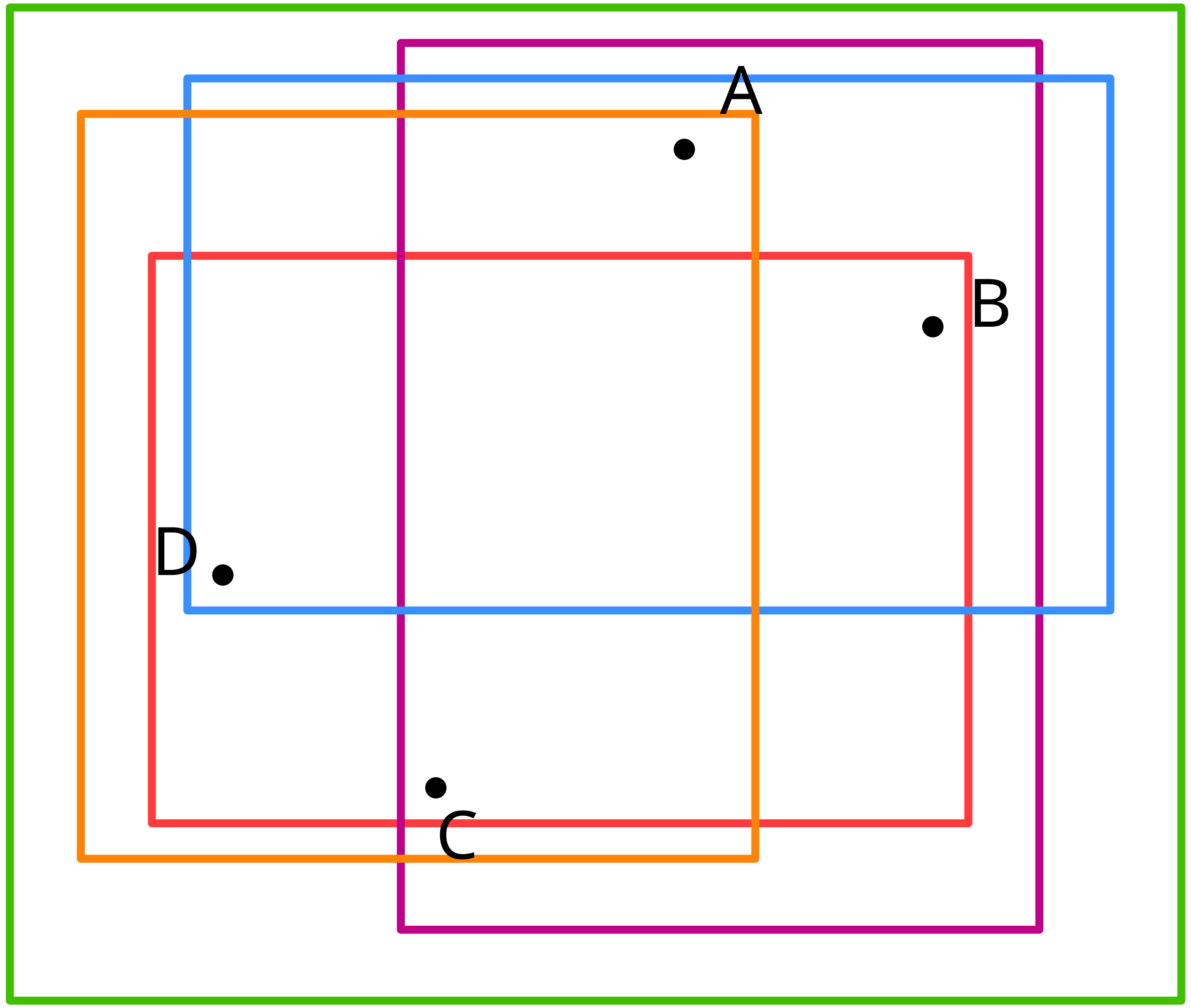
Example: rectangles as ranges

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles



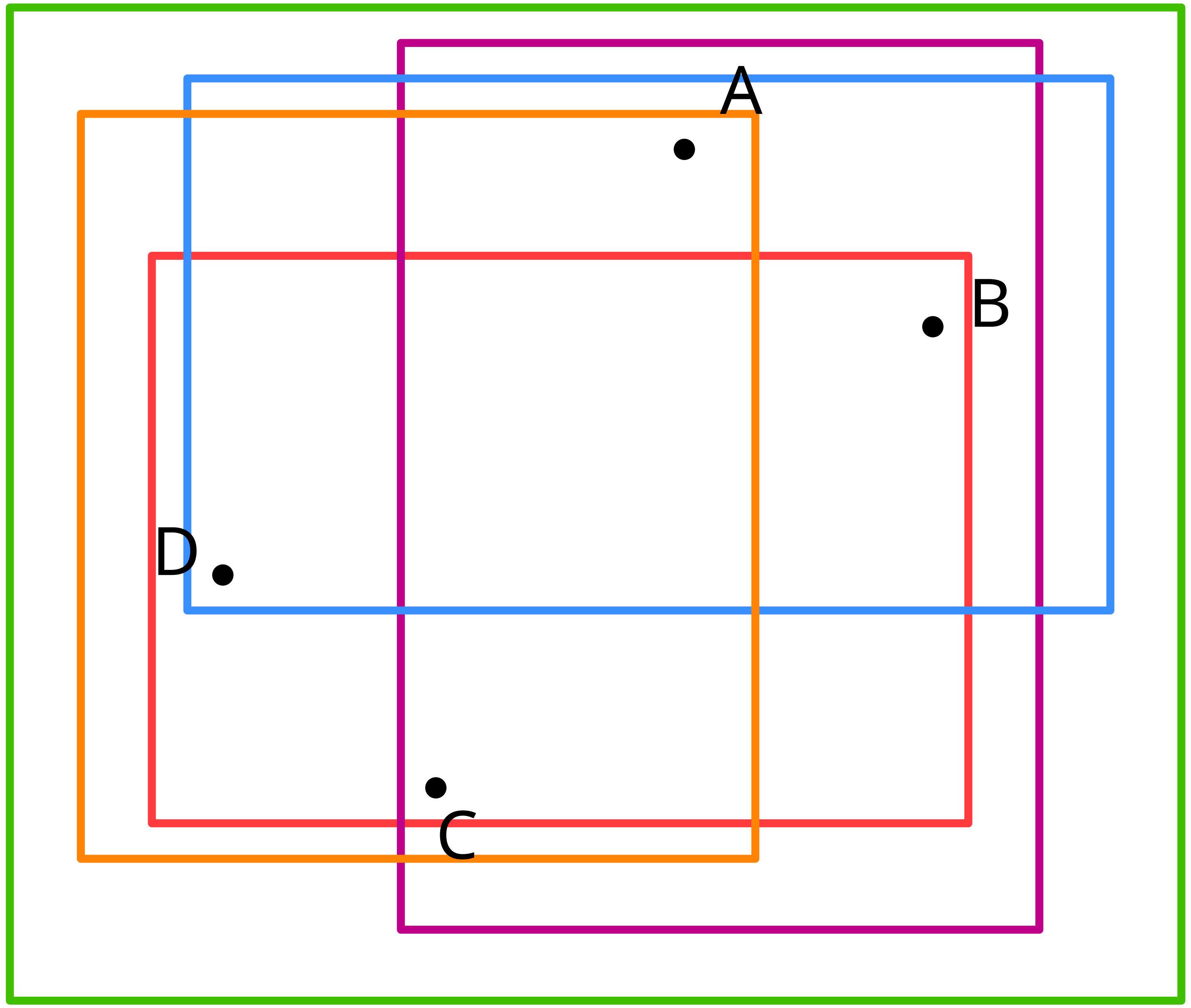
Example: rectangles as ranges

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles



Example: rectangles as ranges

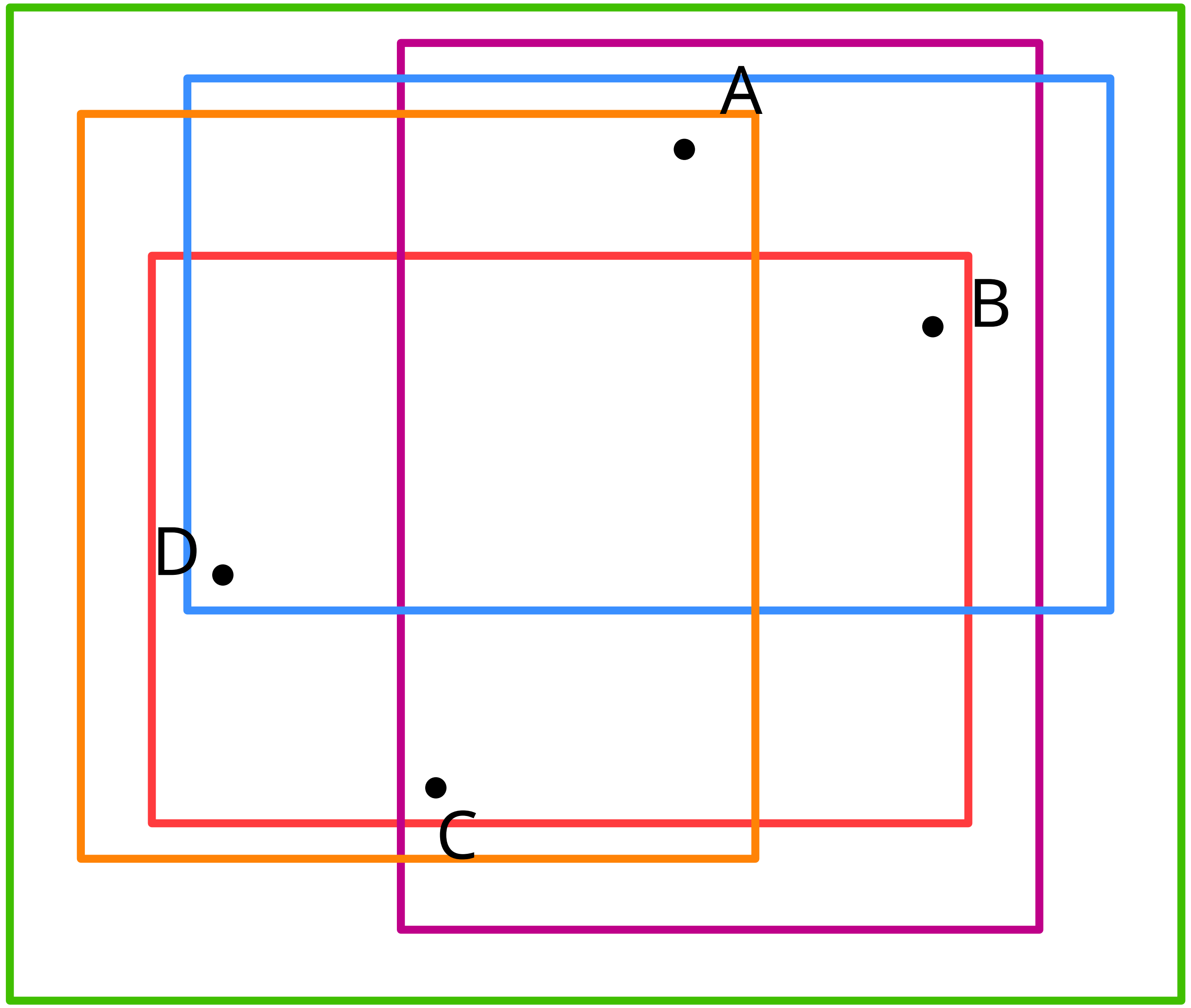
range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles



shattered !

Example: rectangles as ranges

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles

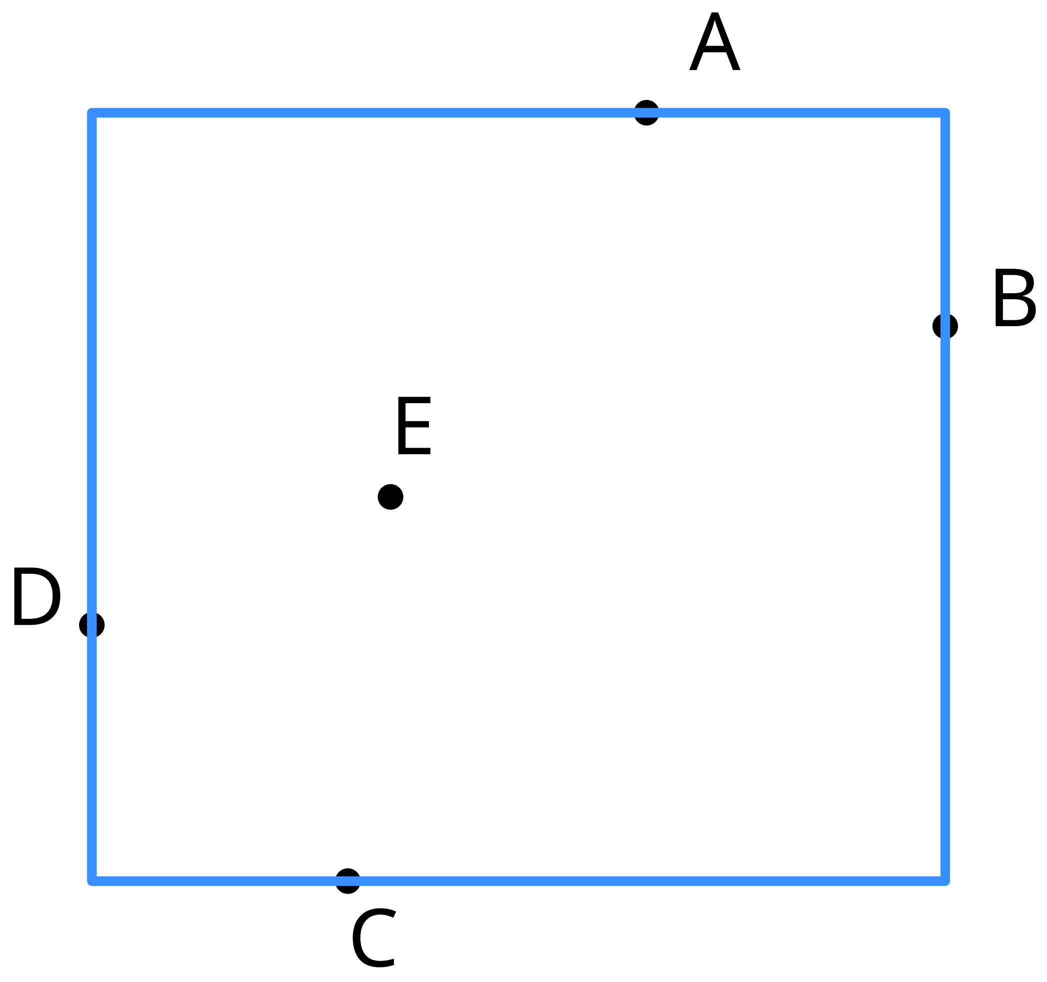


$\Rightarrow$ VC-dimension ≥ 4

shattered !

Example: rectangles as ranges

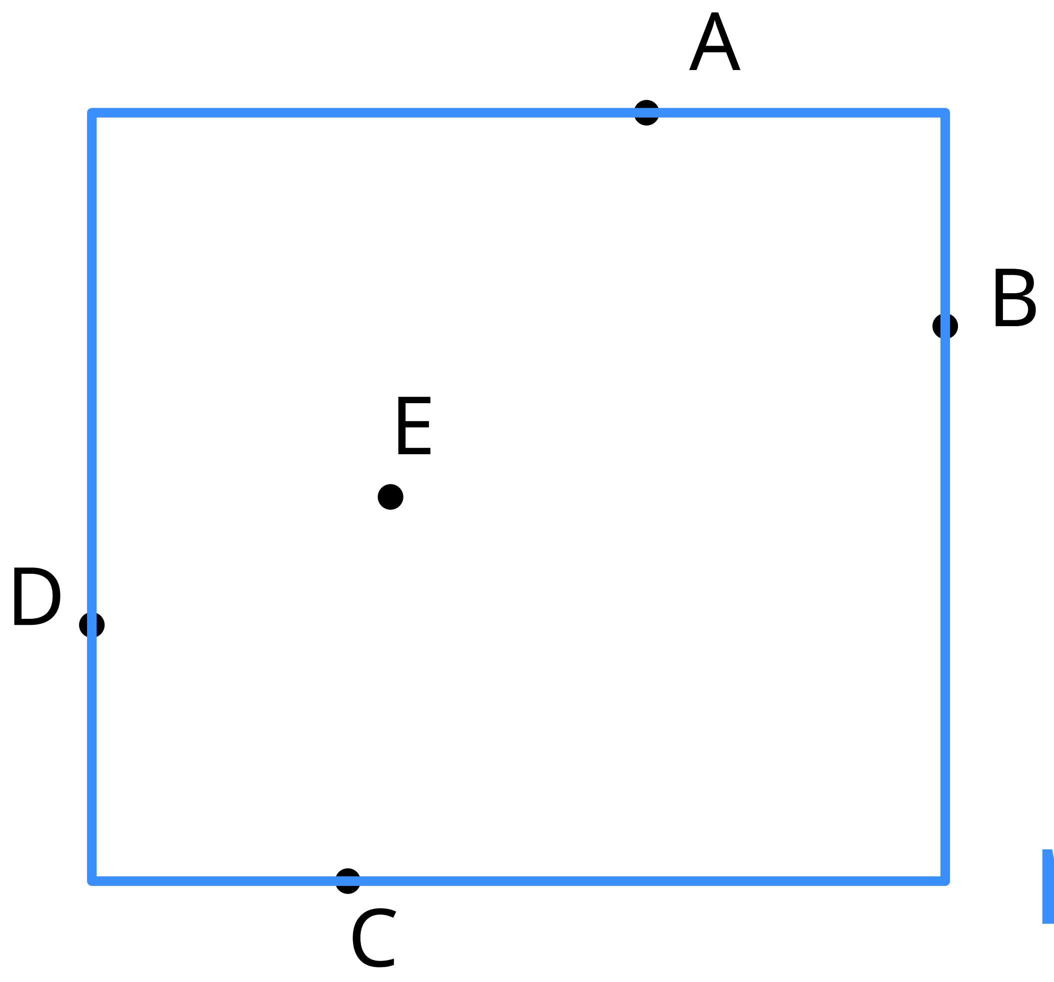
range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles



$\Rightarrow$ VC-dimension ≥ 4

Example: rectangles as ranges

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles

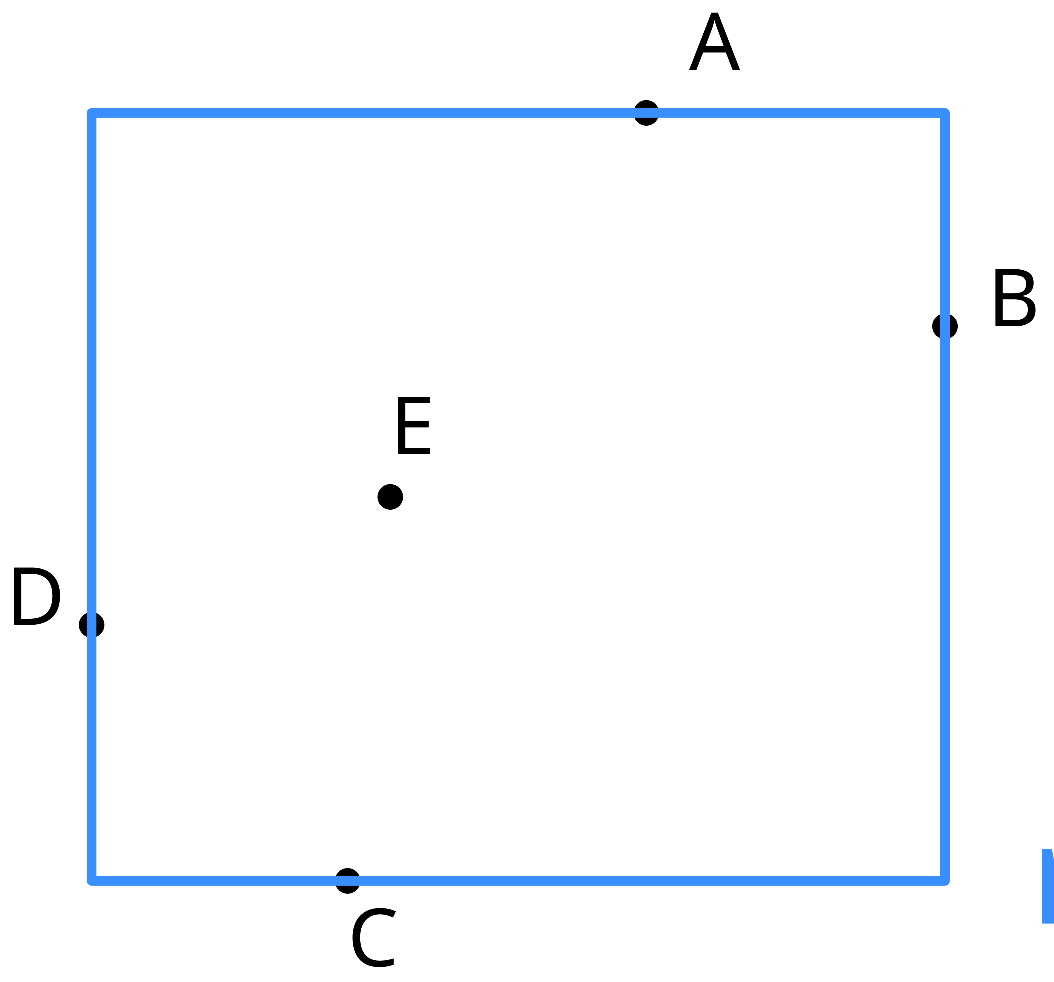


$\Rightarrow$ VC-dimension ≥ 4

not shatter !

Example: rectangles as ranges

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR} =$ set of axis-aligned rectangles



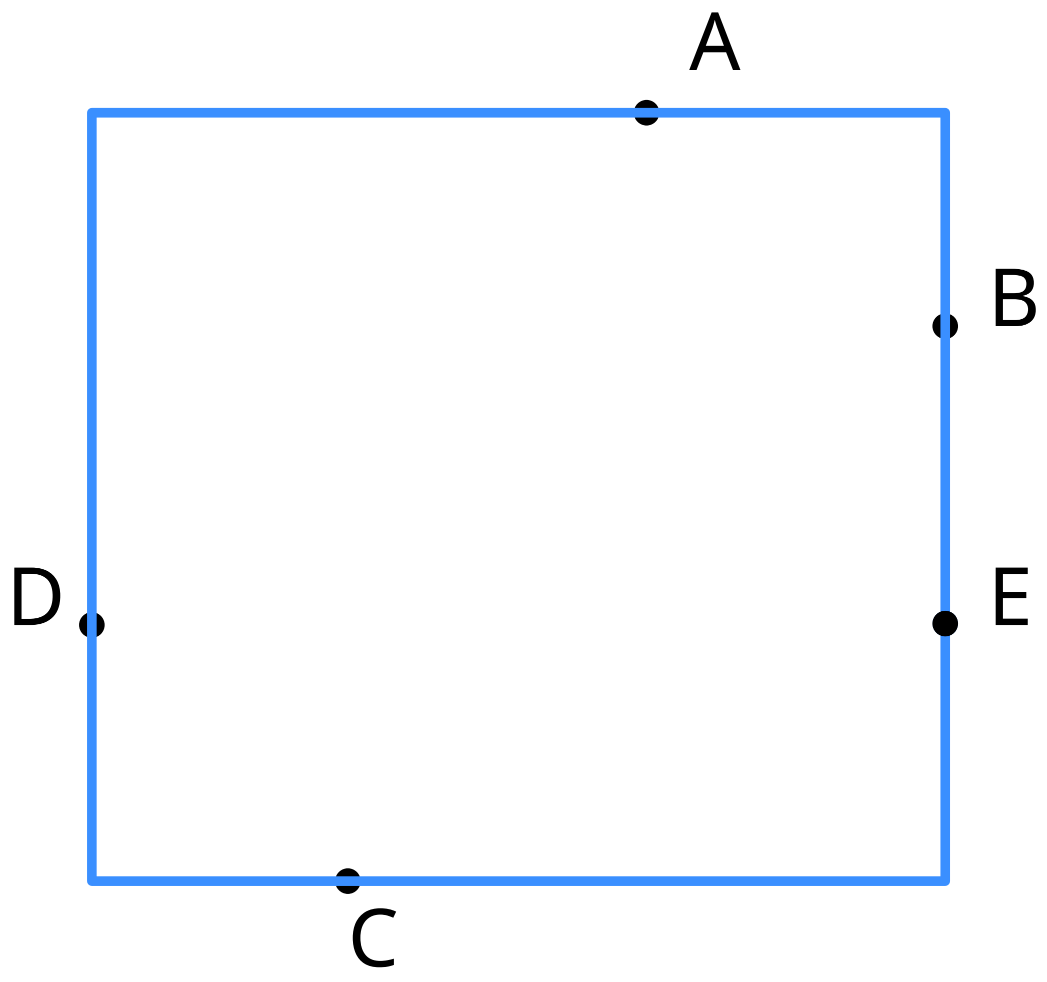
$\Rightarrow$ VC-dimension ≥ 4

case 1: ≥ 1 point inside bounding rectangle

not shatter !

Example: rectangles as ranges

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR} =$ set of axis-aligned rectangles



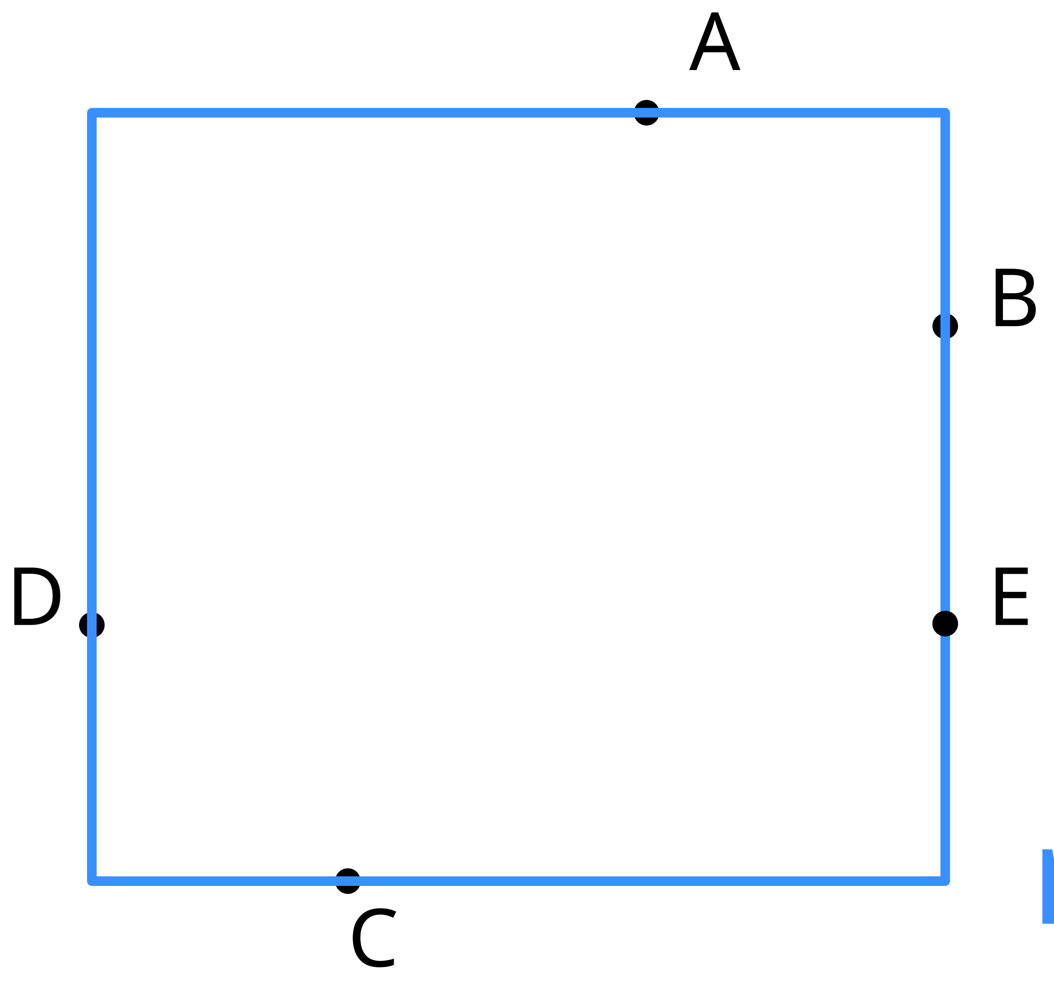
$\Rightarrow$ VC-dimension ≥ 4

case 1: ≥ 1 point inside bounding rectangle

case 2: all points on bounding rectangle

Example: rectangles as ranges

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles



$\Rightarrow$ VC-dimension ≥ 4

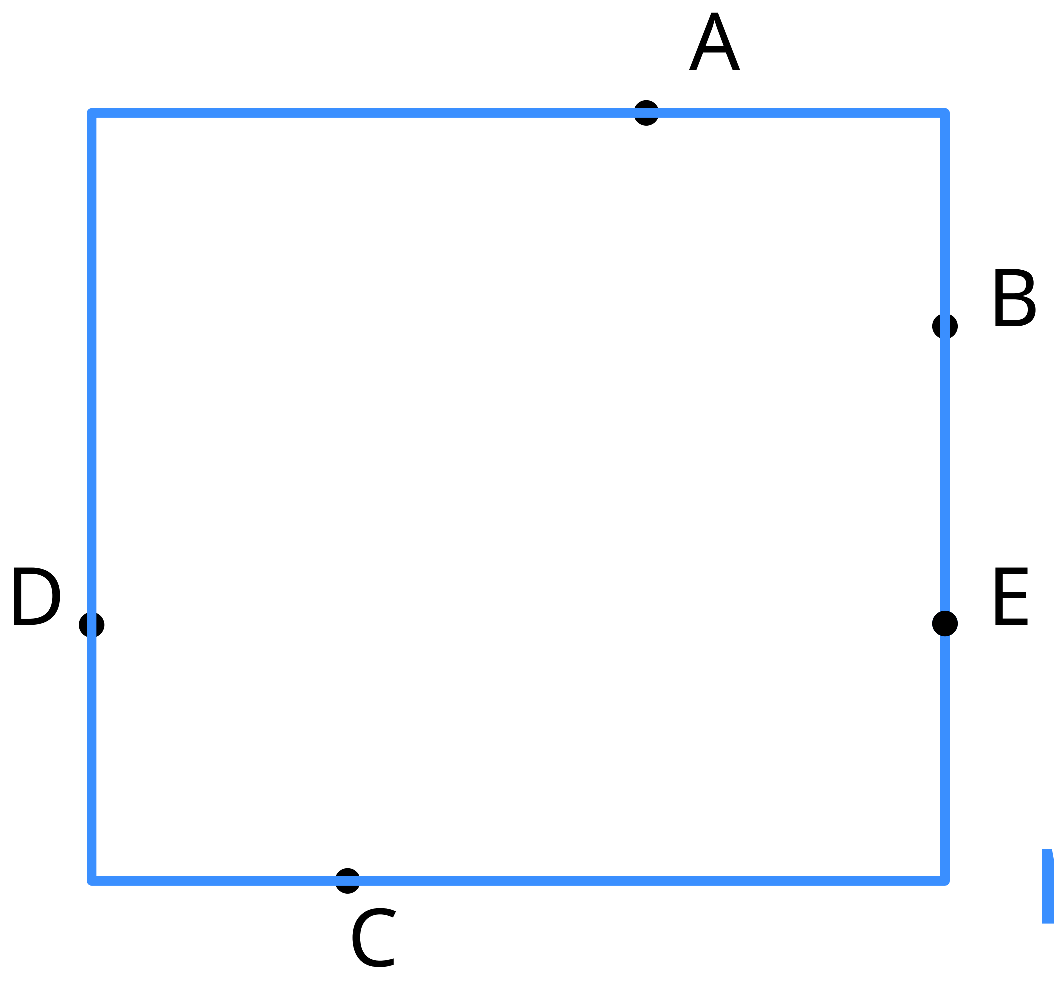
case 1: ≥ 1 point inside bounding rectangle

case 2: all points on bounding rectangle

not shatter !

Example: rectangles as ranges

range space $(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR} =$ set of axis-aligned rectangles



$\Rightarrow$ VC-dimension ≥ 4

case 1: ≥ 1 point inside bounding rectangle

case 2: all points on bounding rectangle

$\Rightarrow$ VC-dimension $= 4$

not shatter !

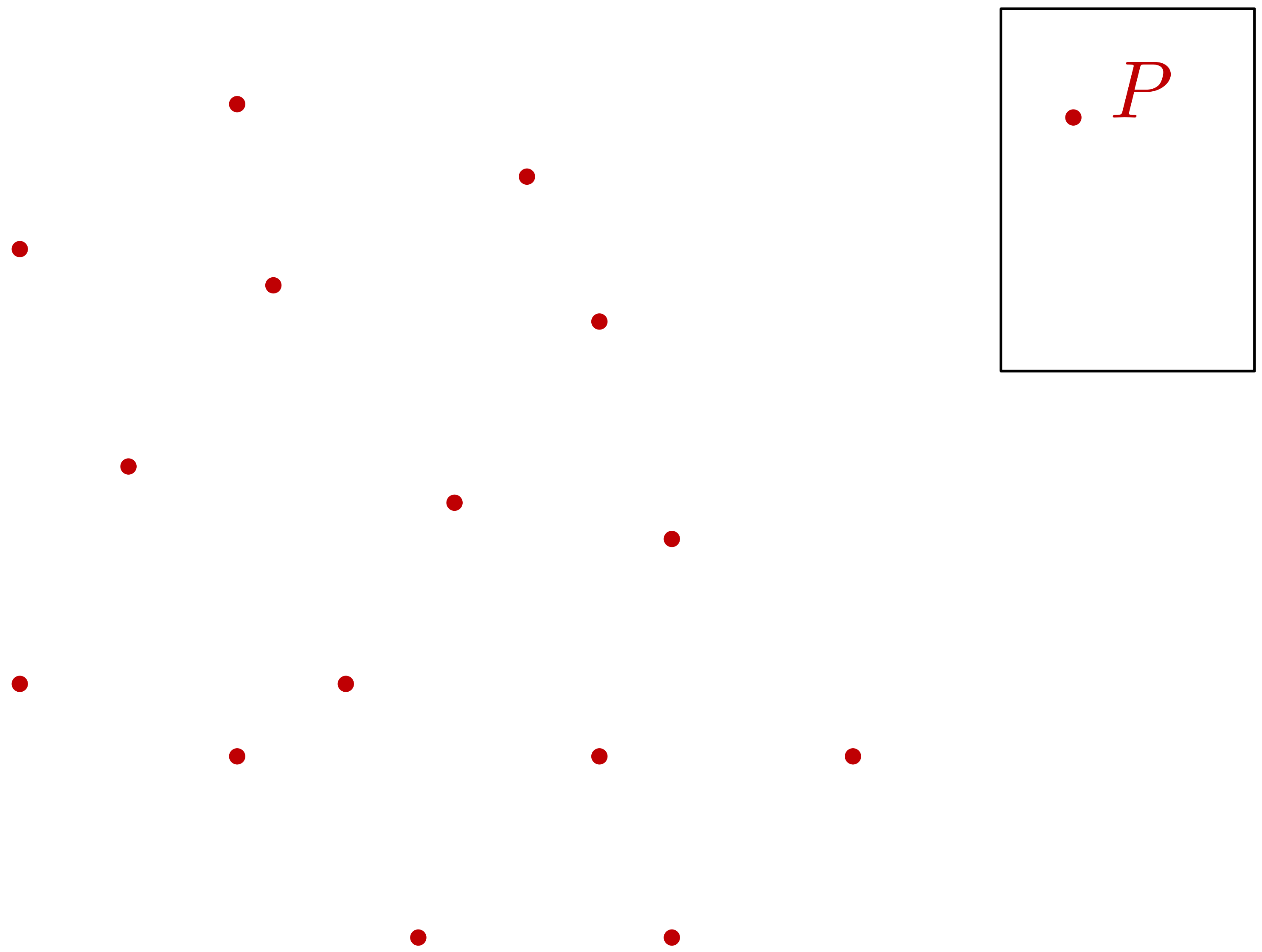
Summary: VC-dimension of geometric range spaces

range space	VC-dimension
$(\mathbb{R}, \mathcal{I})$, with $\mathcal{I}$ = set of closed intervals	2
$(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D}$ = set of disks	3
$(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles	4
$(\mathbb{R}^2, \mathcal{GR})$, with $\mathcal{GR}$ = set of arbitrary oriented rectangles	? ≥ 4
$(\mathbb{R}^2, \mathcal{C})$, with $\mathcal{C}$ = set of closed convex sets	∞

ε-samples

Measure and Estimate

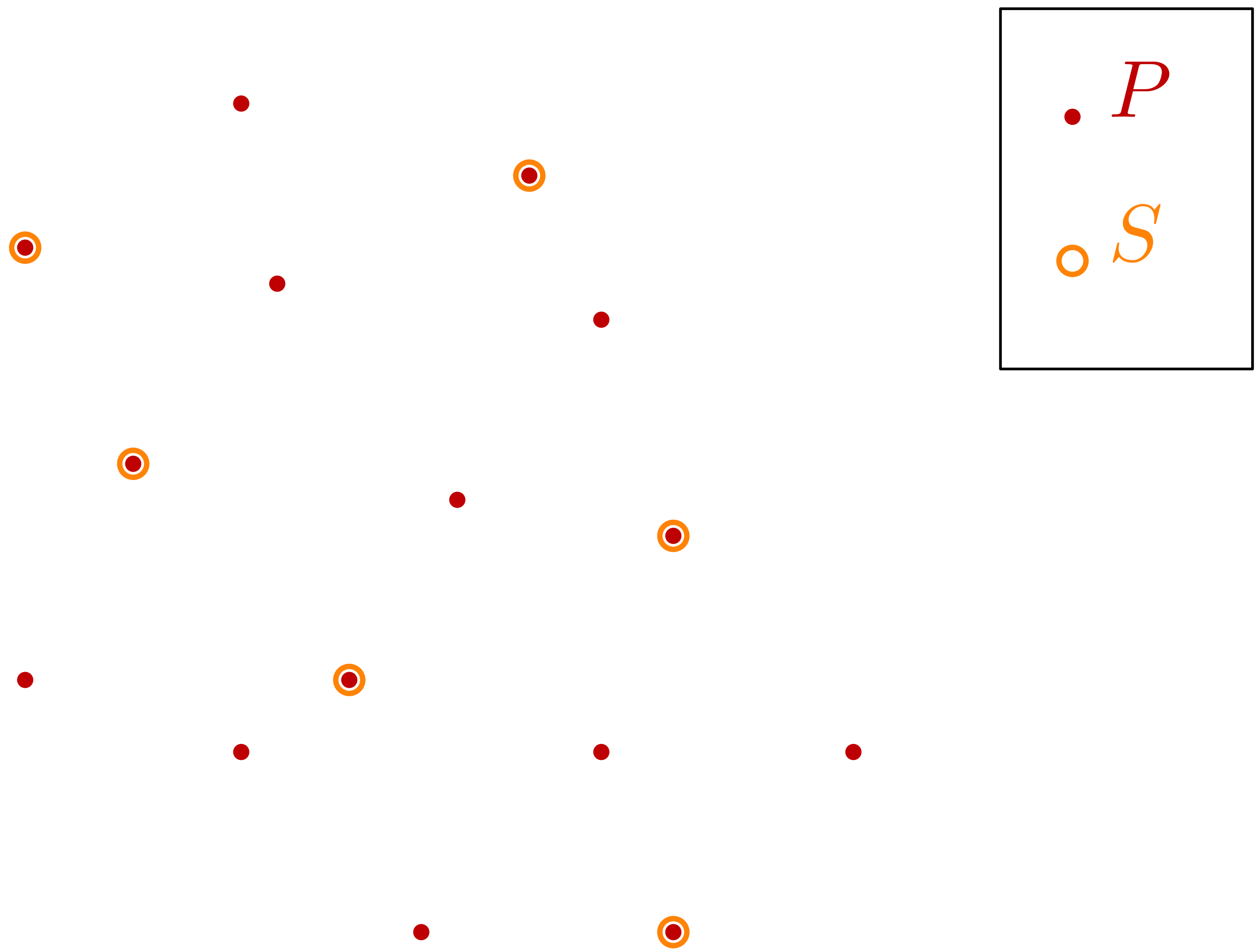
Measure: $\mu(r) = \frac{|r \cap P|}{|P|}$



Measure and Estimate

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

Estimate: $\hat{\mu}(r) = \frac{|r \cap S|}{|S|}$



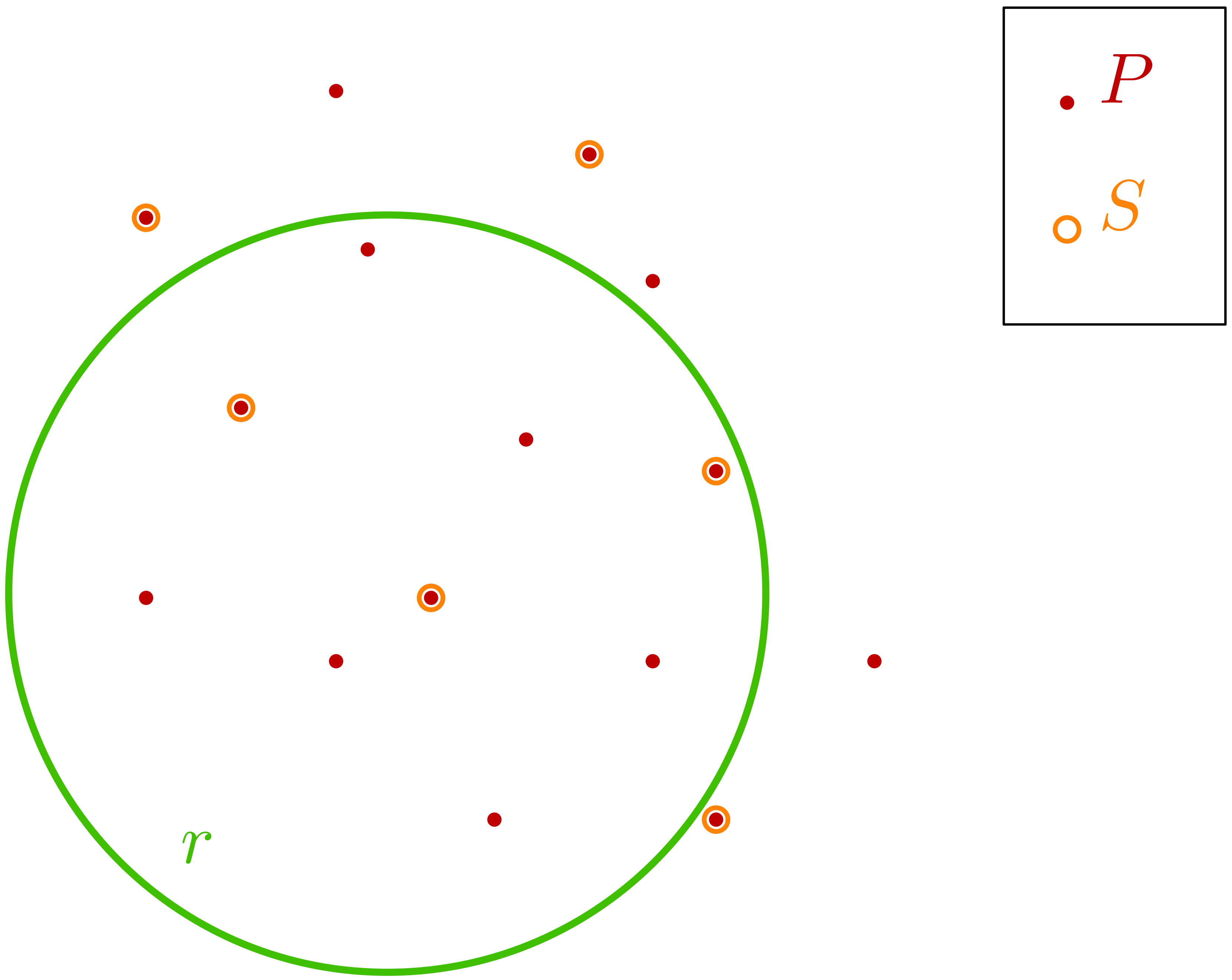
Measure and Estimate

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

$\mu(Q) = \frac{9}{15} = 0.6$

Estimate: $\hat{\mu}(r) = \frac{|r \cap S|}{|S|}$

$\hat{\mu}(Q) = \frac{3}{6} = 0.5$



Measure and Estimate

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

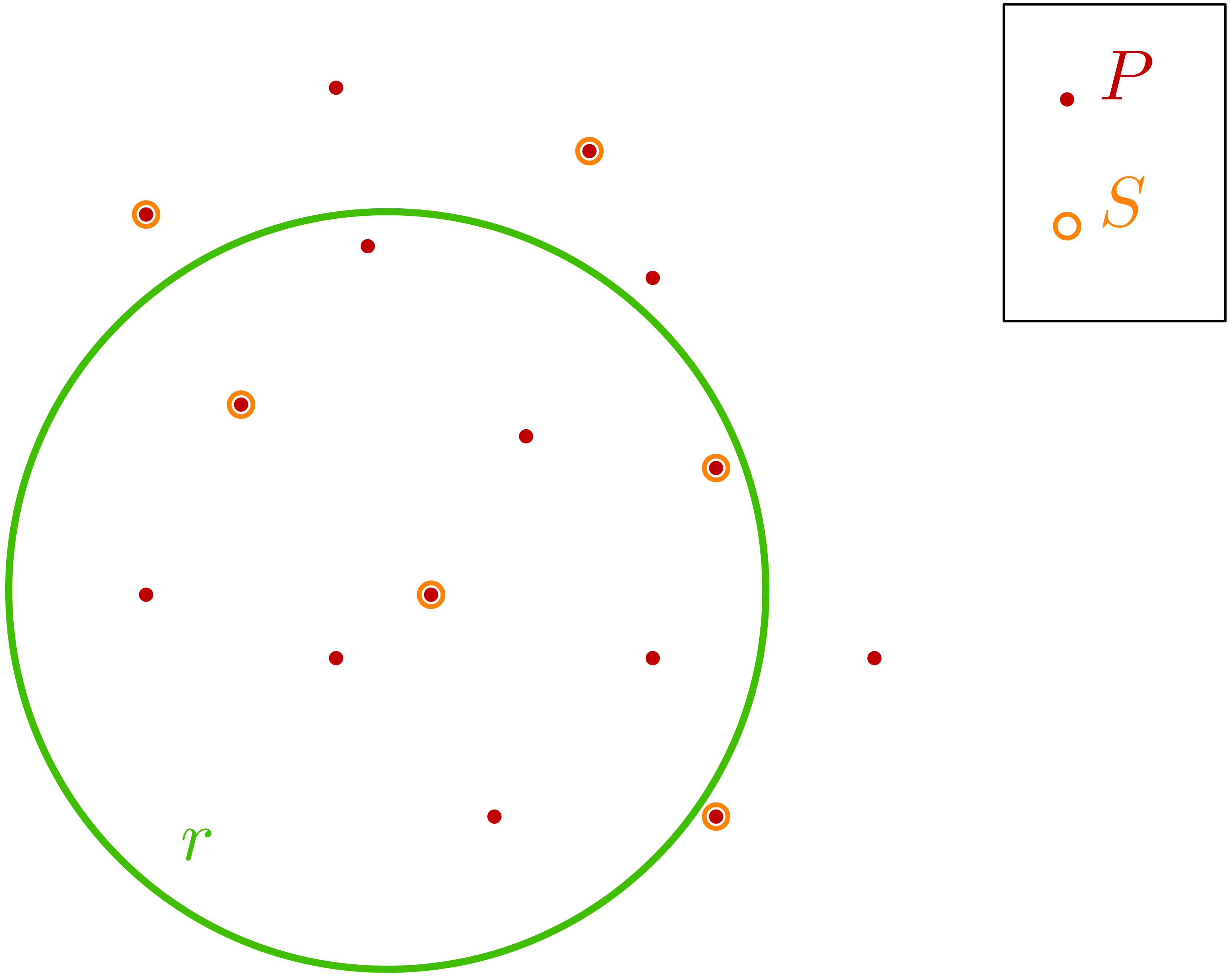
$\mu(Q) = \frac{9}{15} = 0.6$

Estimate: $\hat{\mu}(r) = \frac{|r \cap S|}{|S|}$

$\hat{\mu}(Q) = \frac{3}{6} = 0.5$

Good Sample S:

for all $r \in \mathcal{R}$, $\hat{\mu}(r) \approx \mu(r)$



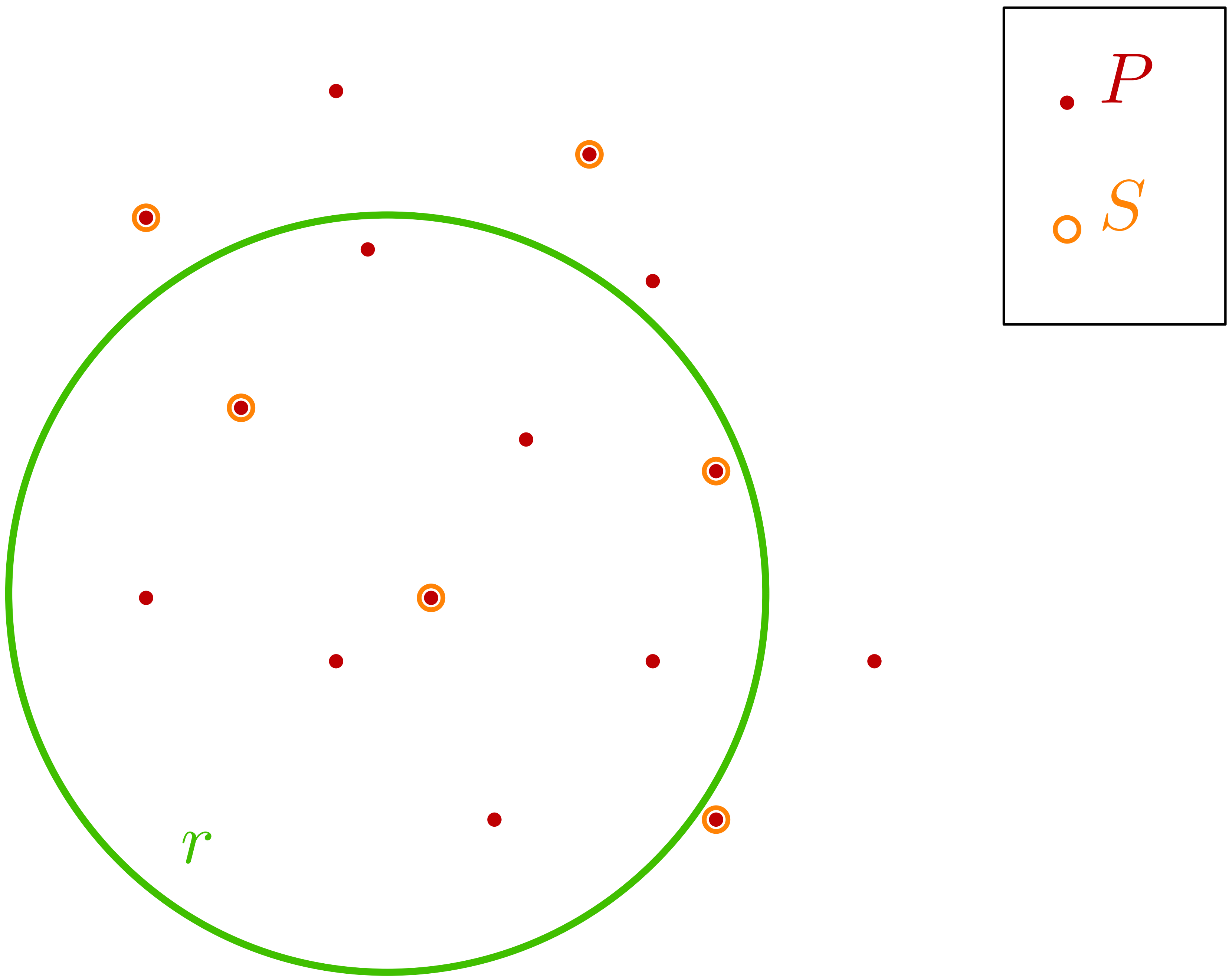
ε-samples

ε-sample S:

for all $r \in \mathcal{R}$ and any

$0 \leq \varepsilon \leq 1$

$|\mu(r) - \hat{\mu}(r)| \leq \varepsilon$



ε-samples

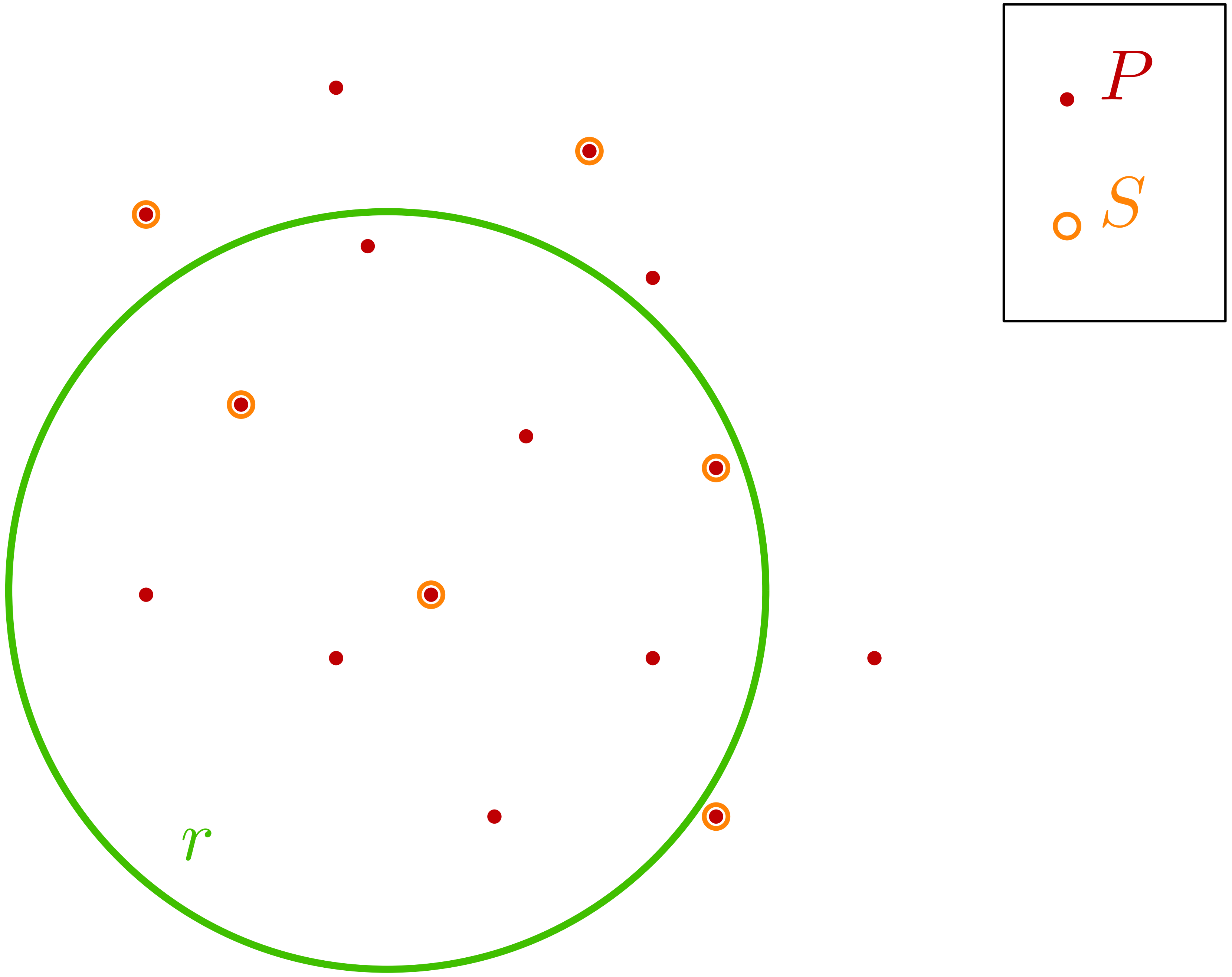
ε-sample S:

for all $r \in \mathcal{R}$ and any

$0 \leq \varepsilon \leq 1$

$|\mu(r) - \hat{\mu}(r)| \leq \varepsilon$

$$
\begin{aligned}
|\mu(r) - \hat{\mu}(r)| &= |9/15 - 3/6| \\
&= 0.1
\end{aligned}
$$



Quiz

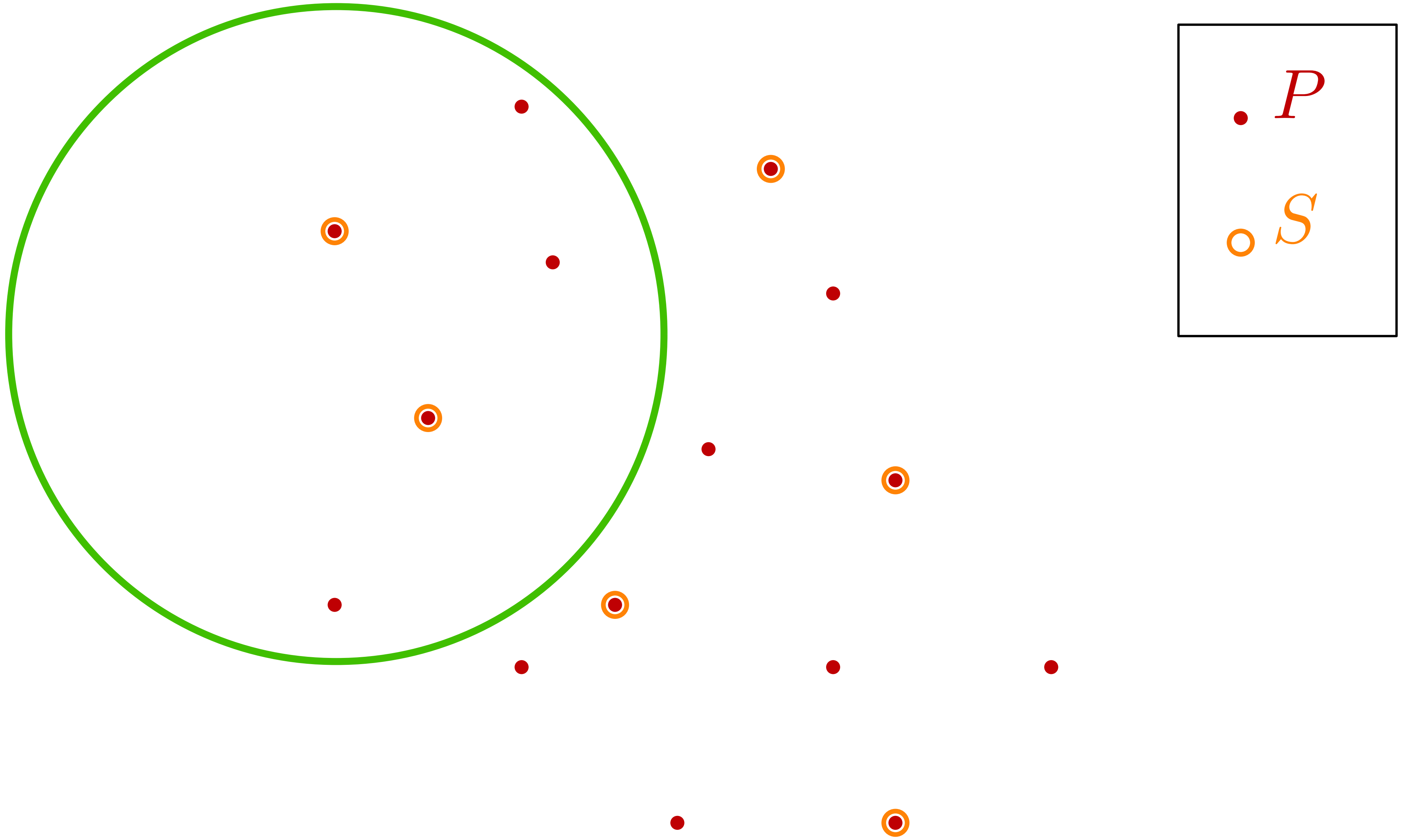
$$|\mu(r) - \hat{\mu}(r)| = \ldots ?$$

A 0.0

B 0.1

C 0.2

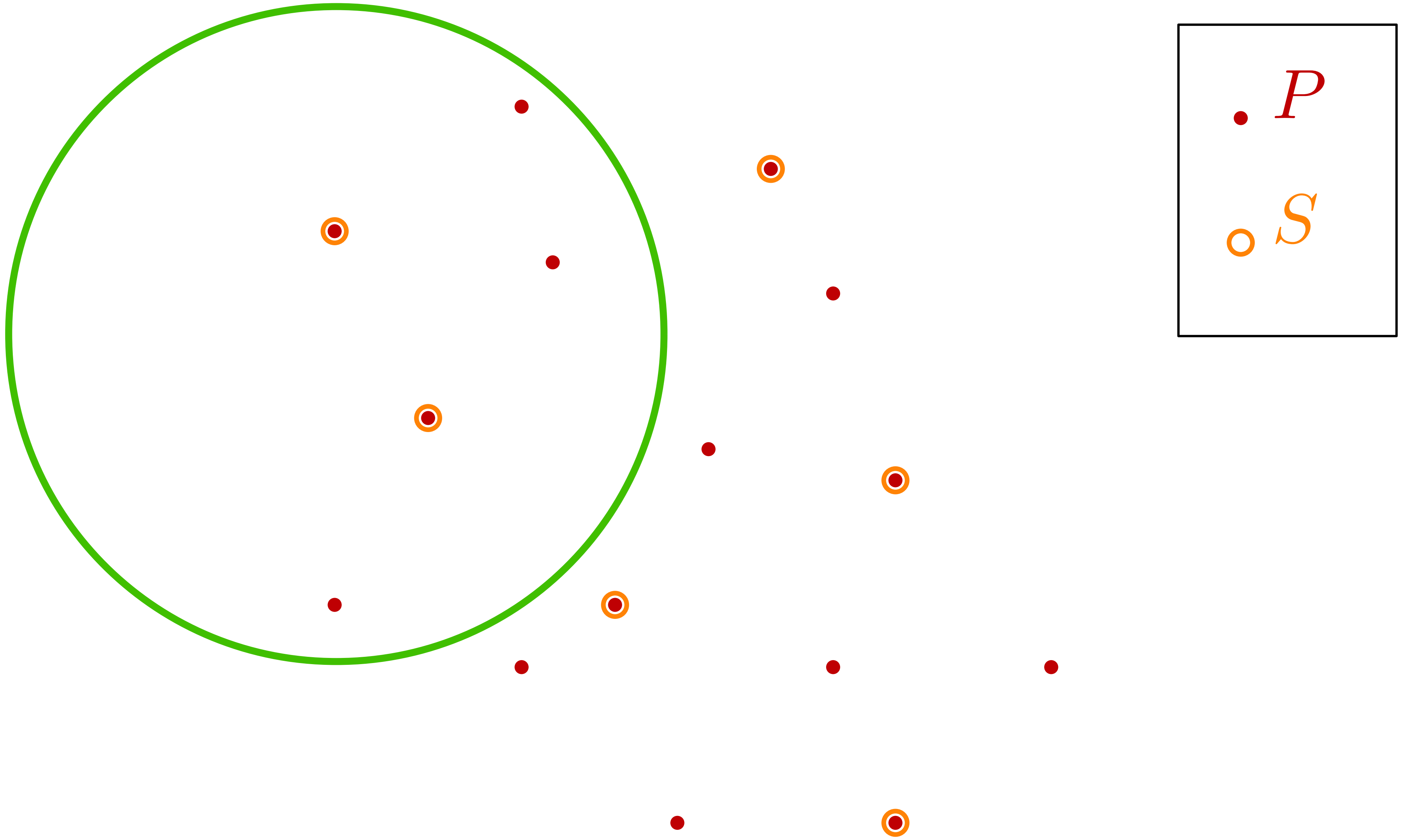
D none of the above



Quiz

$$|\mu(r) - \hat{\mu}(r)| = \ldots ?$$

A	0.0	$\frac{2}{6} = \frac{5}{15}$
B	0.1	
C	0.2	
D	none of the above	



ε-sample theorem

Let $\varphi, \varepsilon > 0$ be parameters and $(X, \mathcal{R})$ be a range space with finite X and VC-dimension δ. A sample of size

$$O\left(\frac{1}{\varepsilon^2}\left(\delta + \log \varphi^-1\right)\right)$$

is an ε-sample for $(X, \mathcal{R})$ with probability $\geq 1 - \varphi$

(we skip the proof)

Example from motivation

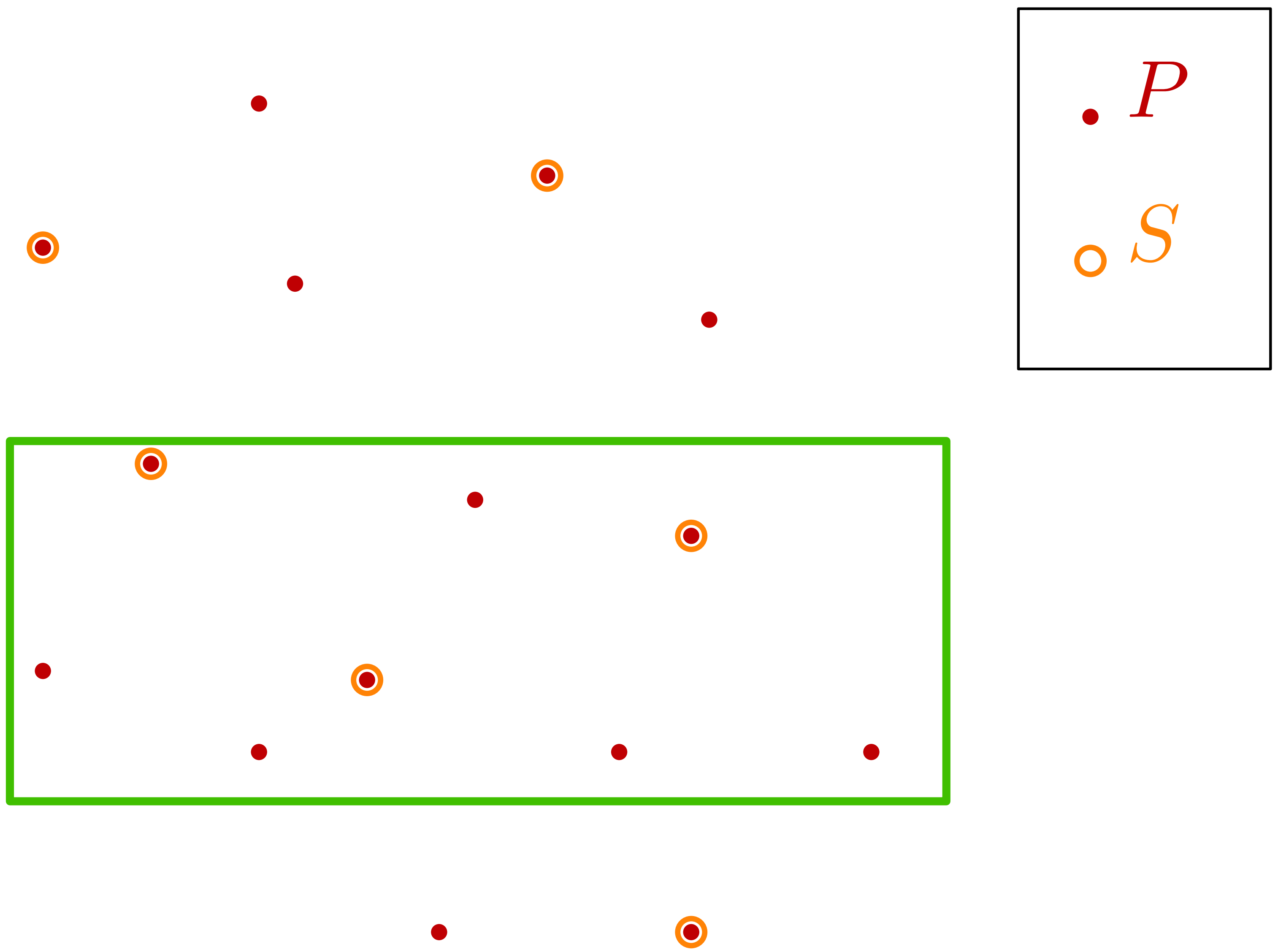
Given P,

how many points do we need to sample ($S \subset P$), such that

2. for any query rectangle r

$$\left| \frac{|r \cap P|}{|P|} - \frac{|r \cap S|}{|S|} \right| \leq 0.25 \qquad ?$$

with probability 0.999



Example from motivation

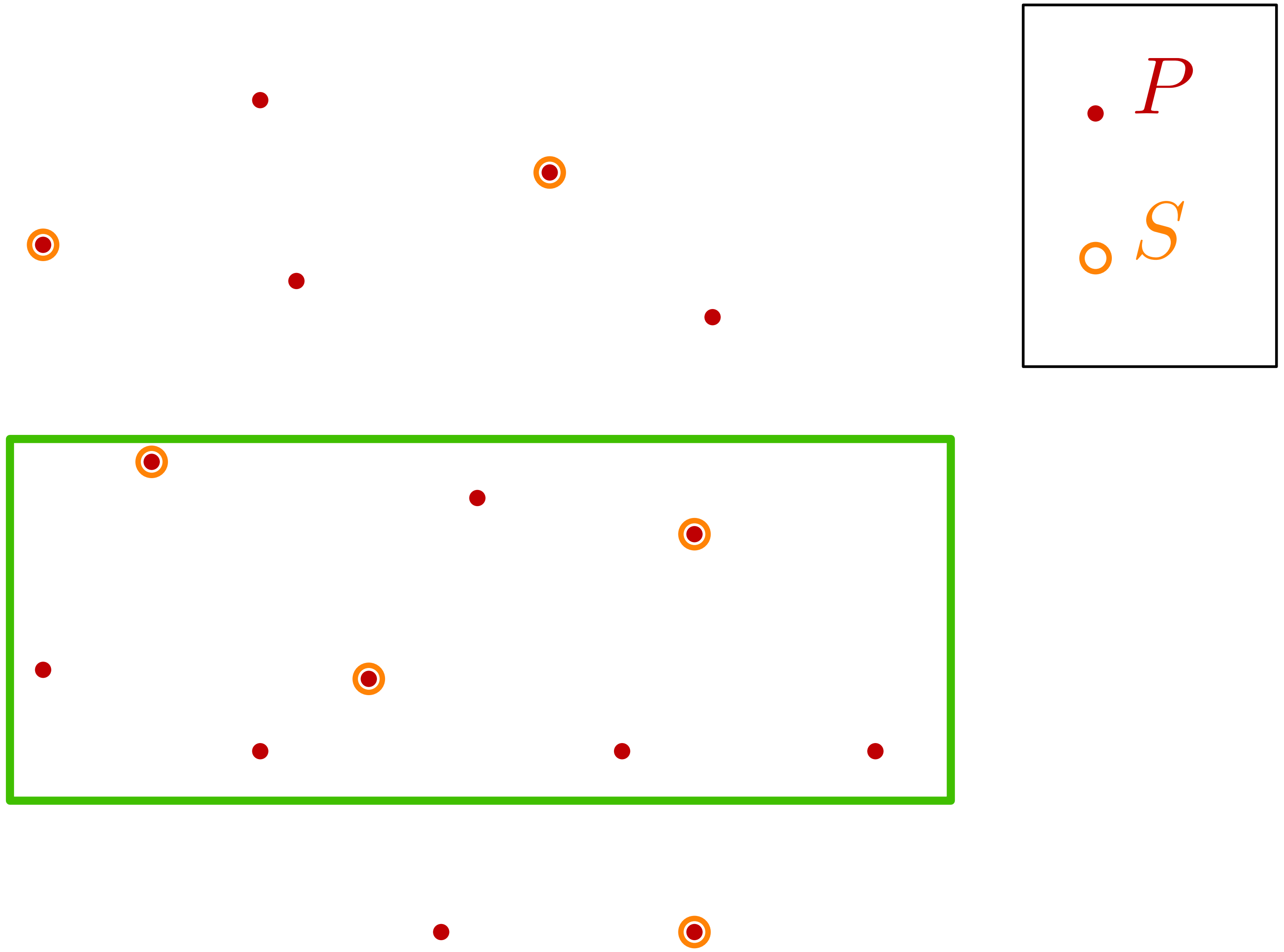
Given P,

how many points do we need to sample ($S \subset P$), such that

2. for any query rectangle r

$$\left| \frac{|r \cap P|}{|P|} - \frac{|r \cap S|}{|S|} \right| \leq 0.25 = \varepsilon \text{ ?}$$

with probability 0.999



Example from motivation

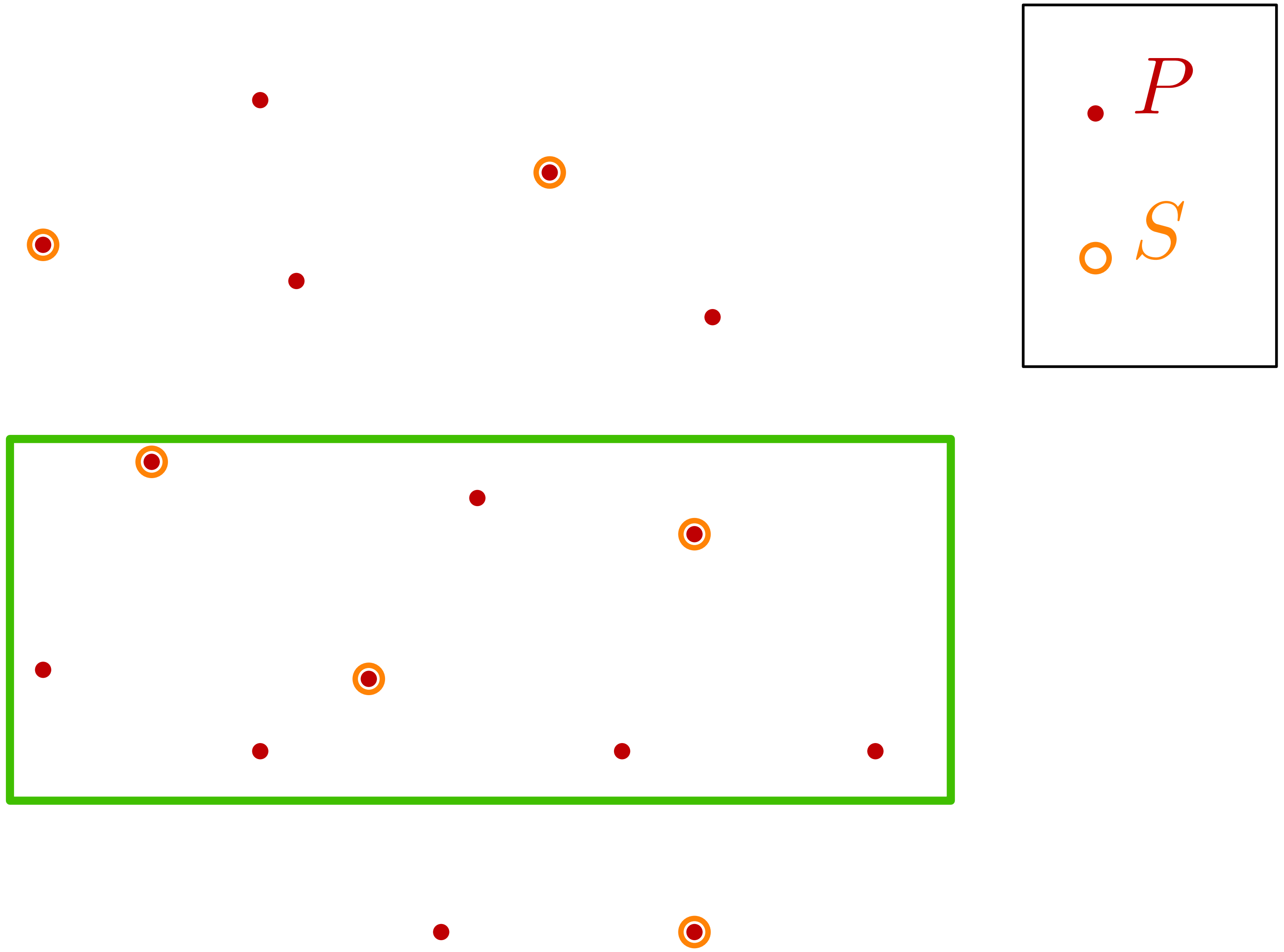
Given P,

how many points do we need to sample ($S \subset P$), such that

2. for any query rectangle r

$$\left| \frac{|r \cap P|}{|P|} - \frac{|r \cap S|}{|S|} \right| \leq 0.25 = \varepsilon \, ?$$

with probability 0.999 $= 1 - \varphi$



Example from motivation

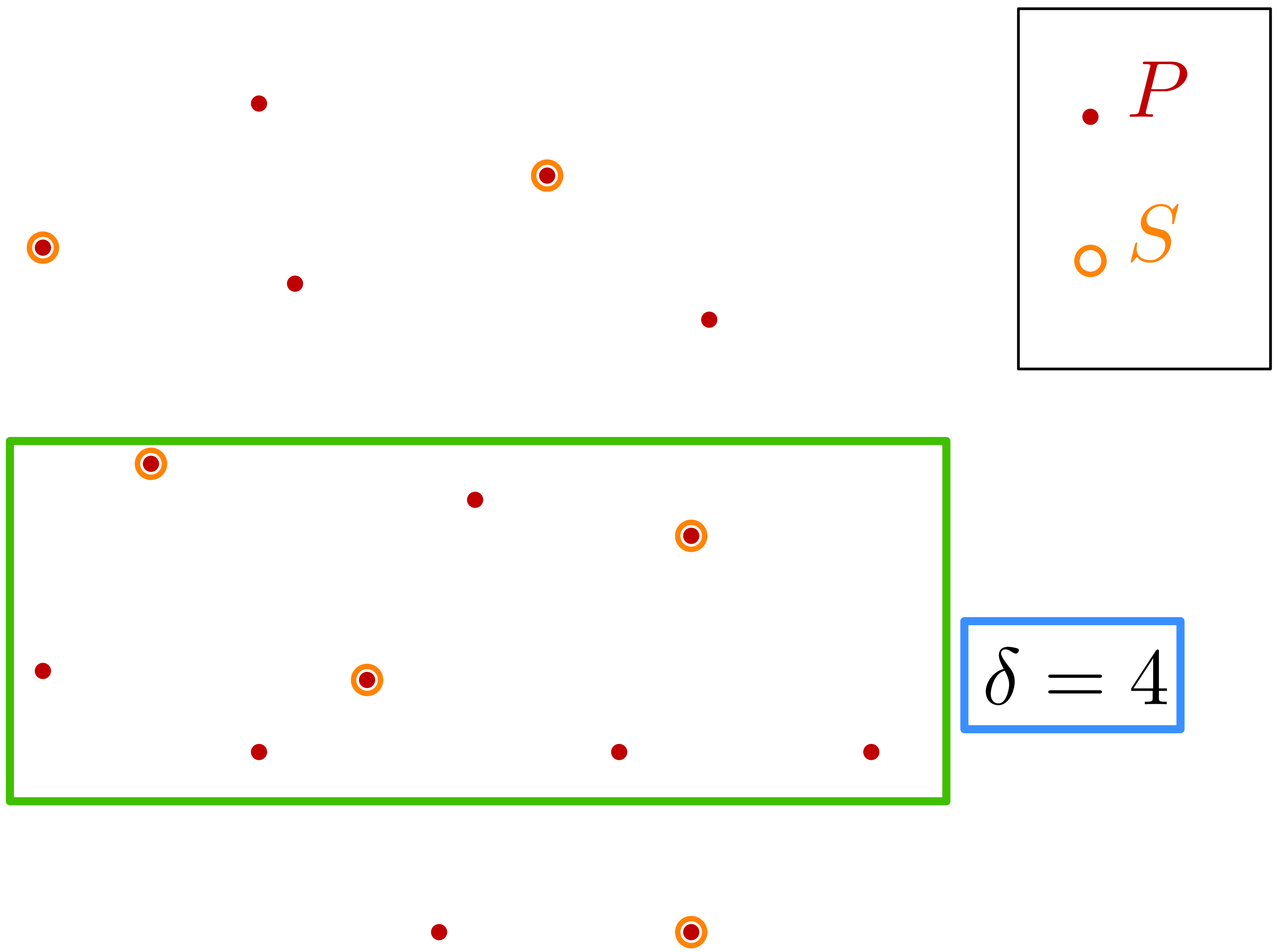
Given P,

how many points do we need to sample ($S \subset P$), such that

2. for any query rectangle r

$$\left| \frac{|r \cap P|}{|P|} - \frac{|r \cap S|}{|S|} \right| \leq 0.25 = \varepsilon \ ?$$

with probability $0.999 = 1 - \varphi$



Example from motivation

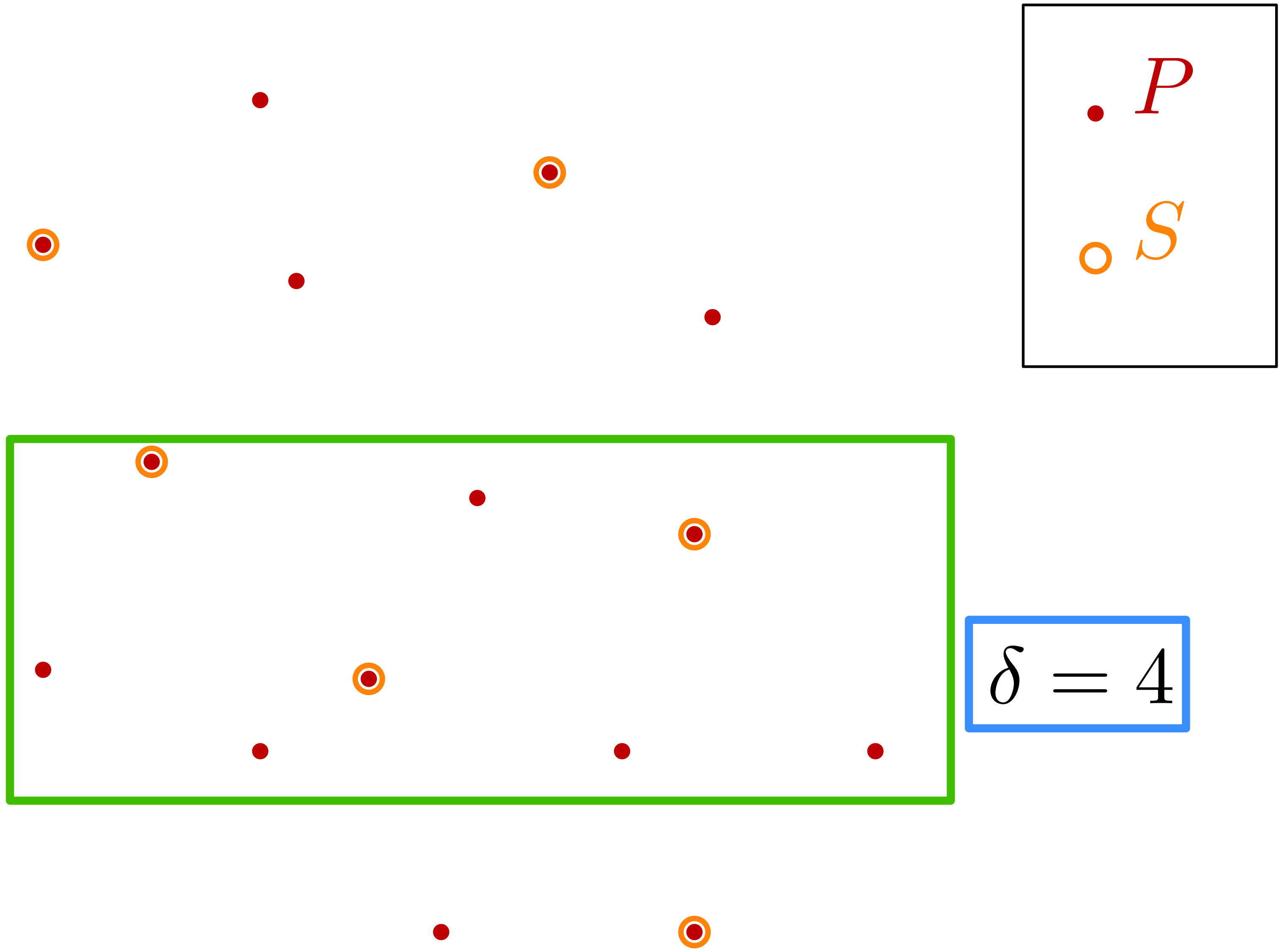
Given P,

how many points do we need to sample ($S \subset P$), such that

2. for any query rectangle r

$$\left| \frac{|r \cap P|}{|P|} - \frac{|r \cap S|}{|S|} \right| \leq 0.25 = \varepsilon \; ?$$

with probability $0.999 = 1 - \varphi$



answer: $O\left(\frac{1}{\varepsilon^2}\left(4 + \log \phi^{-}1\right)\right)$,
in particular $O(1)$ for given ε, φ independent of n

ε-nets

ε-nets

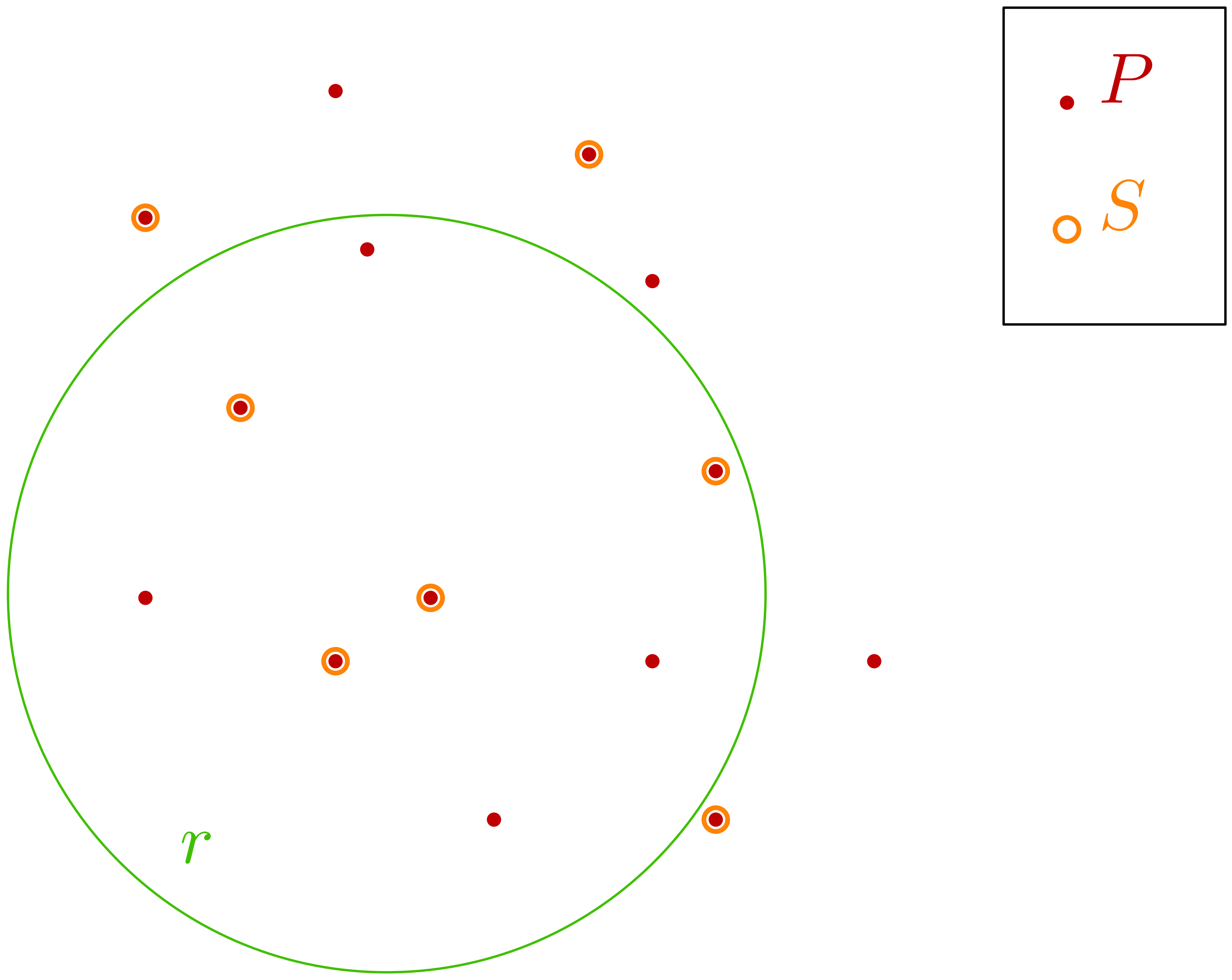
ε-sample S:

for all $r \in \mathcal{R}$ and any

$0 \leq \varepsilon \leq 1$

if $\mu(r) \geq \varepsilon$ and

$|\mu(r) - \hat{\mu}(r)| \leq \varepsilon$ then $\hat{\mu}(r) > 0$



ε-nets

ε-sample S:

for all $r \in \mathcal{R}$ and any

$0 \leq \varepsilon \leq 1$

if $\mu(r) \geq \varepsilon$ and

$|\mu(r) - \hat{\mu}(r)| \leq \varepsilon$ then $\hat{\mu}(r) > 0$

weaker notion:

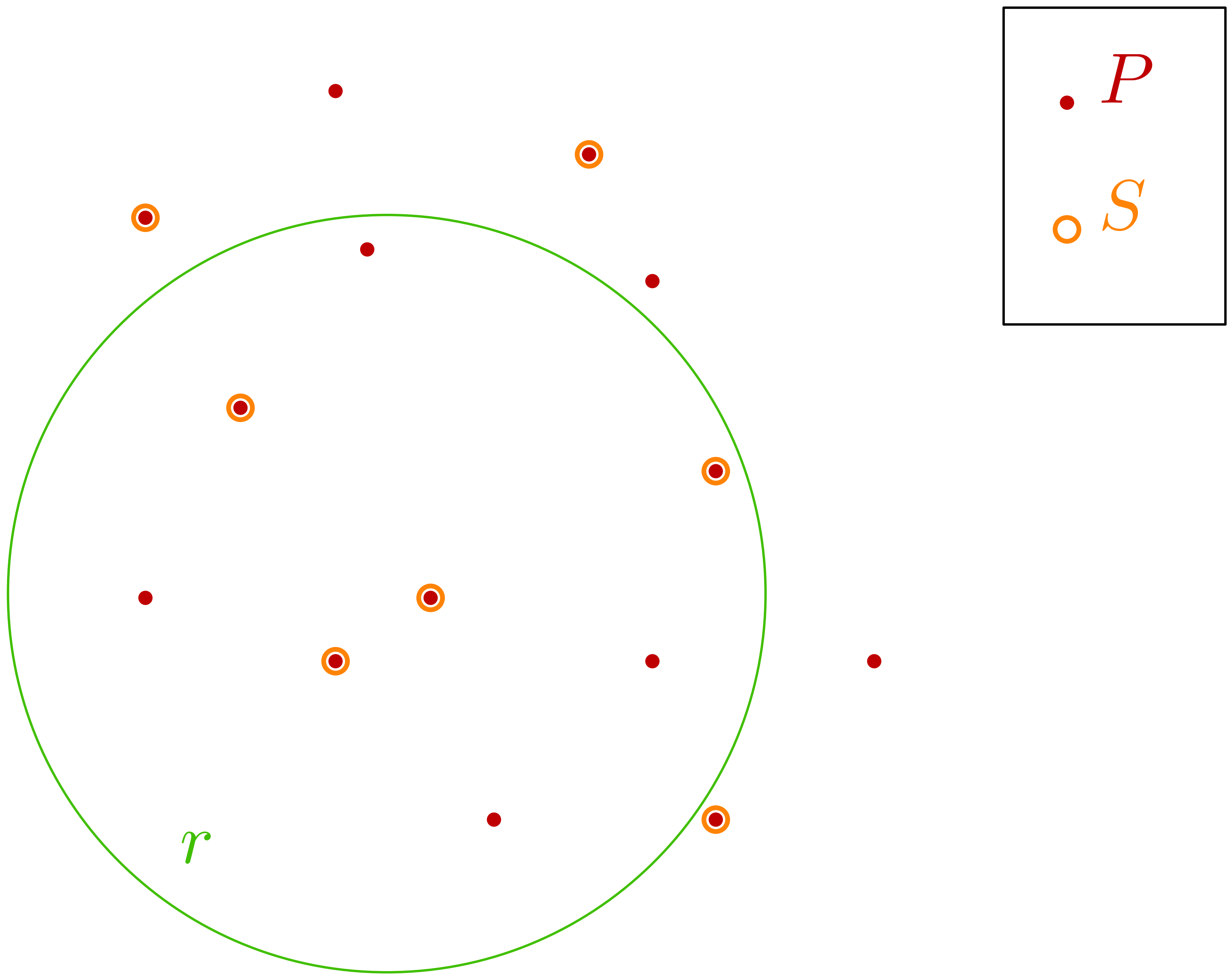
ε-net S:

for all $r \in \mathcal{R}$ and any

$0 \leq \varepsilon \leq 1$

if $\mu(r) \geq \varepsilon$ then r contains

at least one point of S



ε-nets

ε-sample S:

for all $r \in \mathcal{R}$ and any

$0 \leq \varepsilon \leq 1$

if $\mu(r) \geq \varepsilon$ and

$|\mu(r) - \hat{\mu}(r)| \leq \varepsilon$ then $\hat{\mu}(r) > 0$

weaker notion:

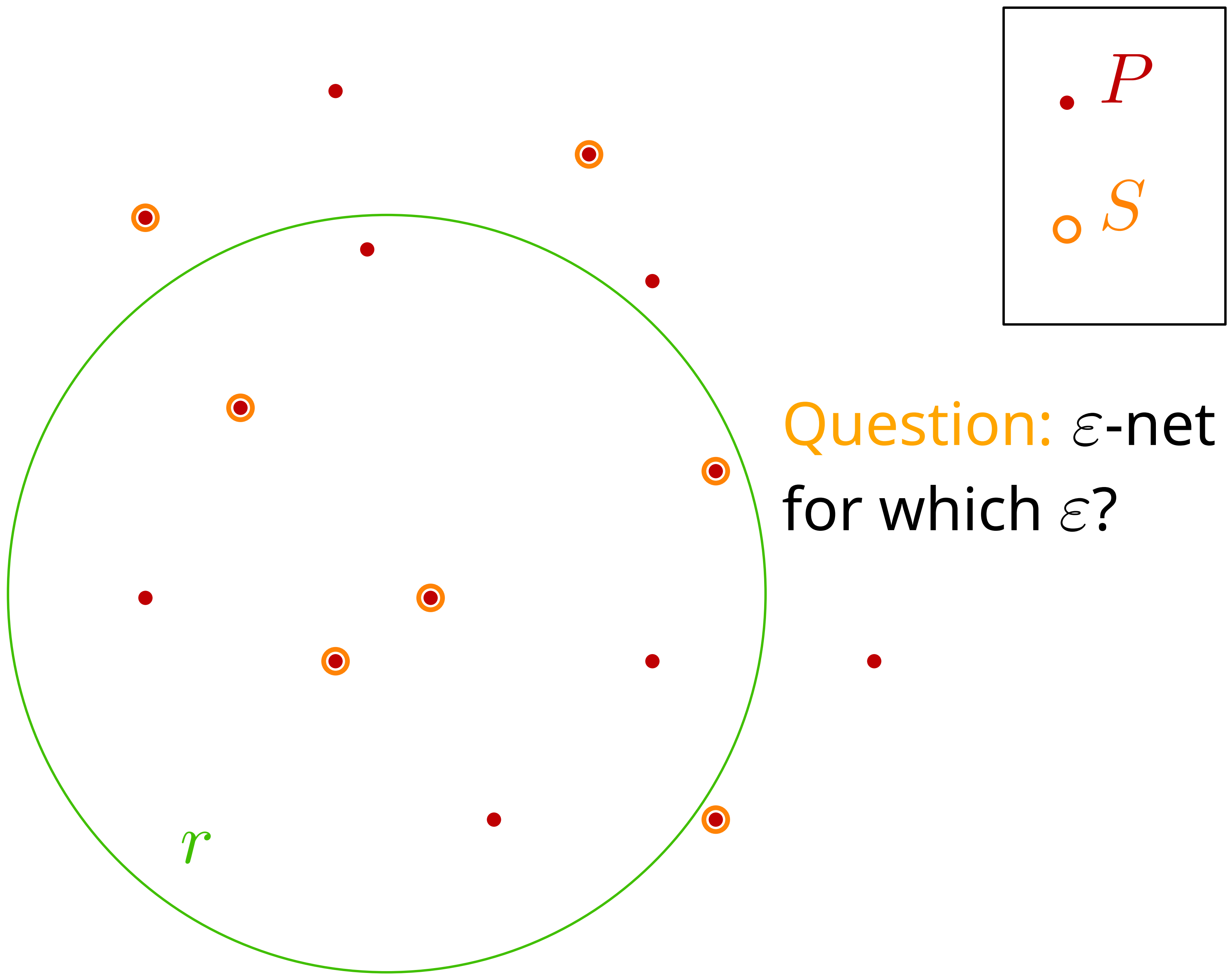
ε-net S:

for all $r \in \mathcal{R}$ and any

$0 \leq \varepsilon \leq 1$

if $\mu(r) \geq \varepsilon$ then r contains

at least one point of S



Question: ε-net for which ε?

ε-Net Theorem

Let $\varphi, \varepsilon > 0$ be parameters and $(X, \mathcal{R})$ be a range space with finite X and VC-dimension δ. A sample obtained by m random draws from X with

$$m \geq \max\left(\frac{4}{\varepsilon}\log\frac{4}{\varphi}, \frac{8\delta}{\varepsilon}\log\frac{16}{\varepsilon}\right)$$

is an ε-net for $(X, \mathcal{R})$ with probability $\geq 1 - \varphi$

(we skip the proof, but there is a proof sketch in book)

ε-Net Theorem

Let $\varphi, \varepsilon > 0$ be parameters and $(X, \mathcal{R})$ be a range space with finite X and VC-dimension δ. A sample obtained by m random draws from X with

$$m \geq \max\left(\frac{4}{\varepsilon}\log\frac{4}{\varphi}, \frac{8\delta}{\varepsilon}\log\frac{16}{\varepsilon}\right)$$

is an ε-net for $(X, \mathcal{R})$ with probability $\geq 1 - \varphi$

(we skip the proof, but there is a proof sketch in book)

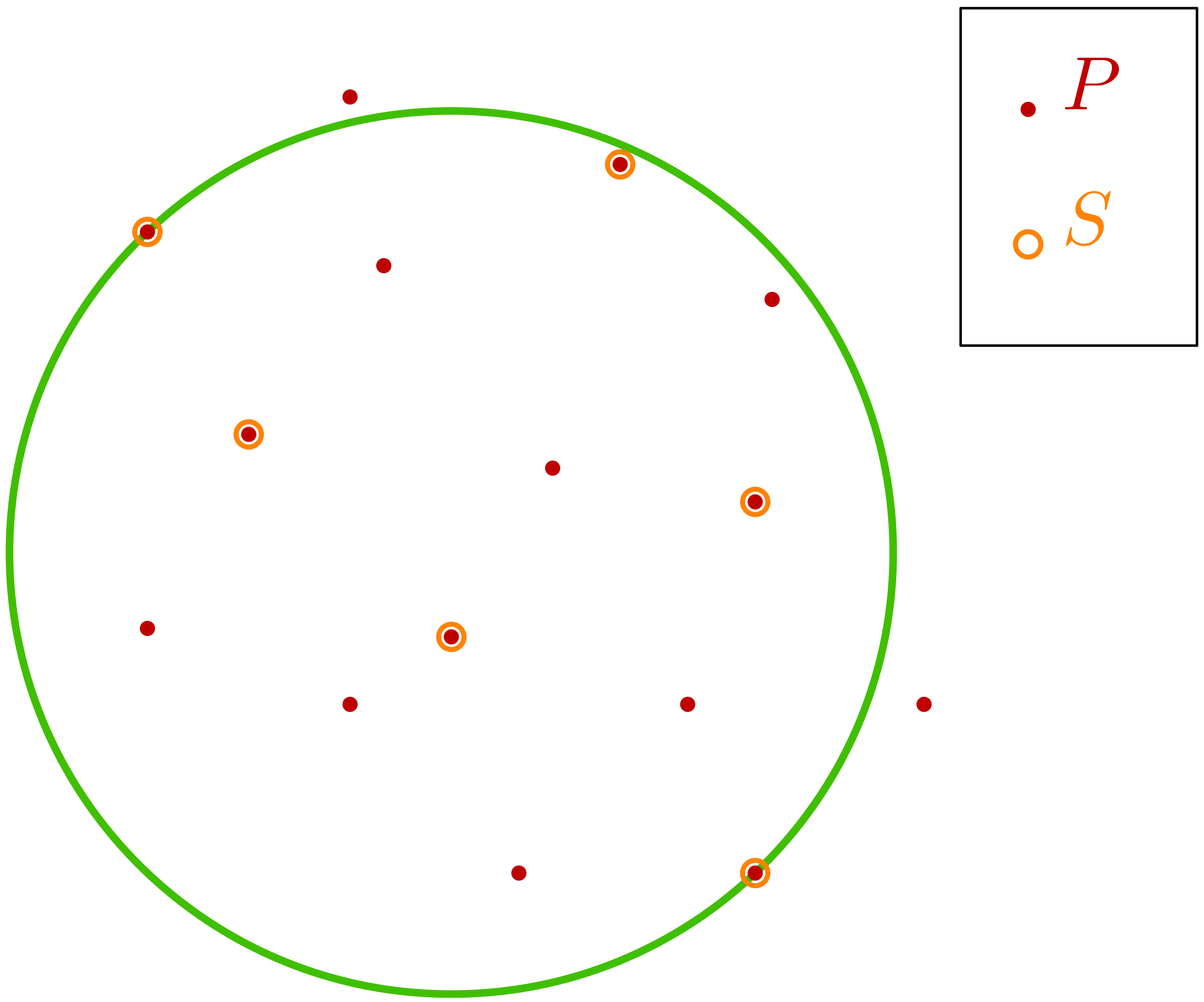
in short: ε-sample $O\left(\frac{\delta}{\varepsilon^2}\right)$ vs ε-net $O\left(\frac{\delta}{\varepsilon}\log\frac{1}{\varepsilon}\right)$

Motivation: sampling for approximation

Given P, how many points do we need to sample ($S \subset P$), such that

the smallest enclosing disk contains 90% of the points in P?

with probability 0.999

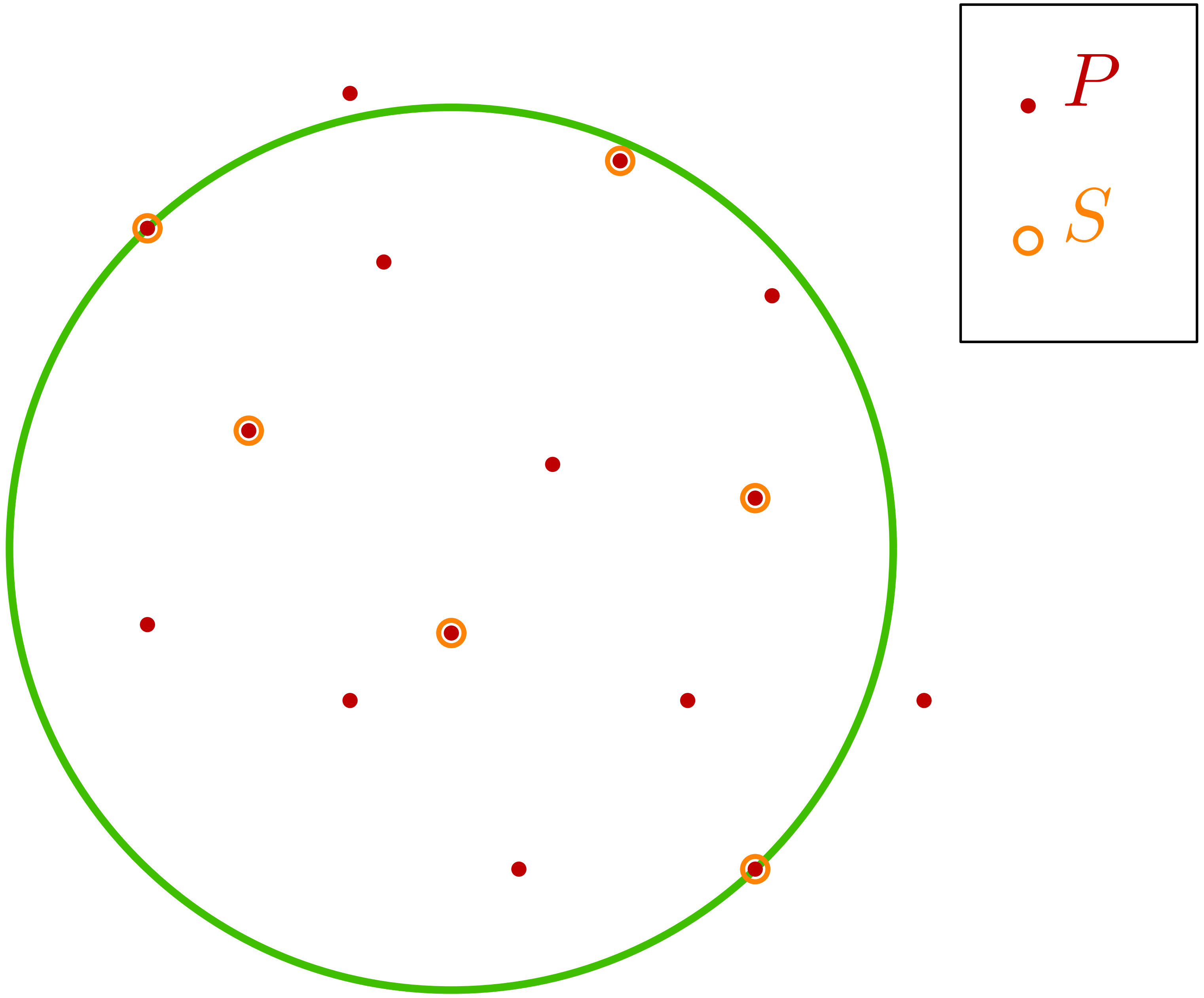


Motivation: sampling for approximation

Given P, how many points do we need to sample ($S \subset P$), such that

the smallest enclosing disk contains 90% of the points in P?

with probability 0.999 $= 1 - \varphi$

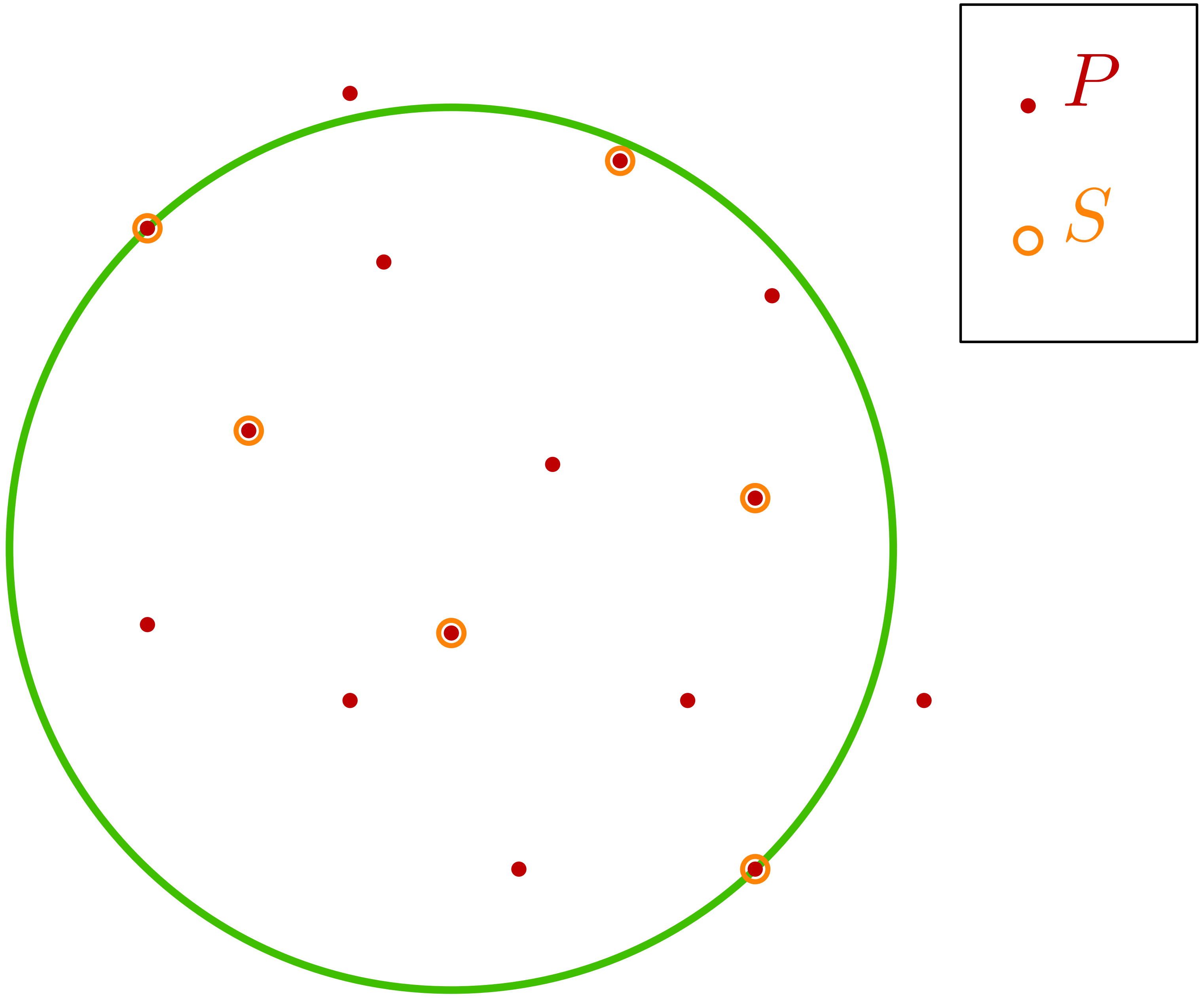


Motivation: sampling for approximation

Given P, how many points do we need to sample ($S \subset P$), such that

the smallest enclosing disk contains 90% of the points in P? $\varepsilon = 0.1$

with probability 0.999 $= 1 - \varphi$



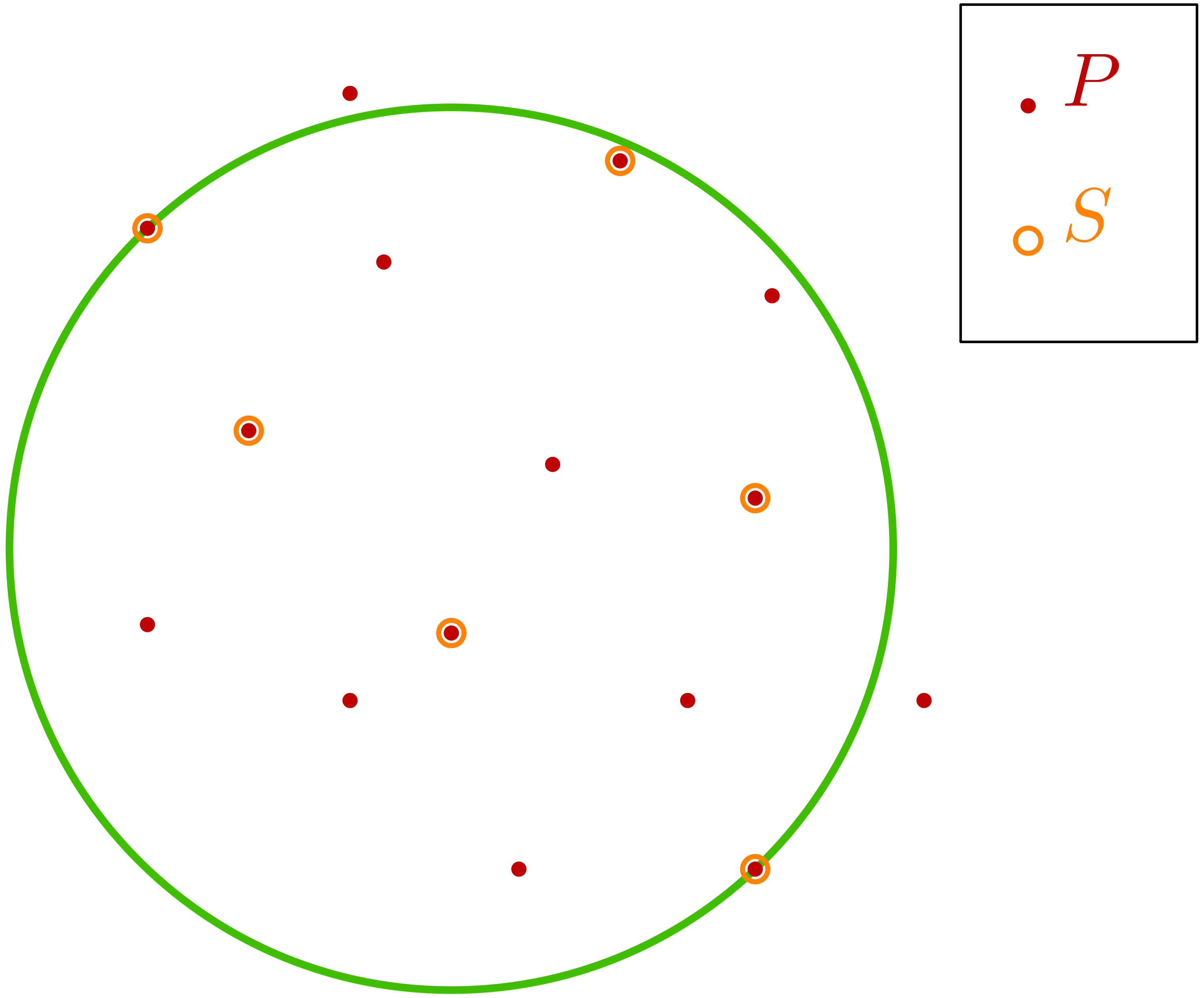
Motivation: sampling for approximation

Given P, how many points do we need to sample ($S \subset P$), such that

the smallest enclosing disk contains 90% of the points in P? $\varepsilon = 0.1$

with probability 0.999 $= 1 - \varphi$

Question: Which range space?



Motivation: sampling for approximation

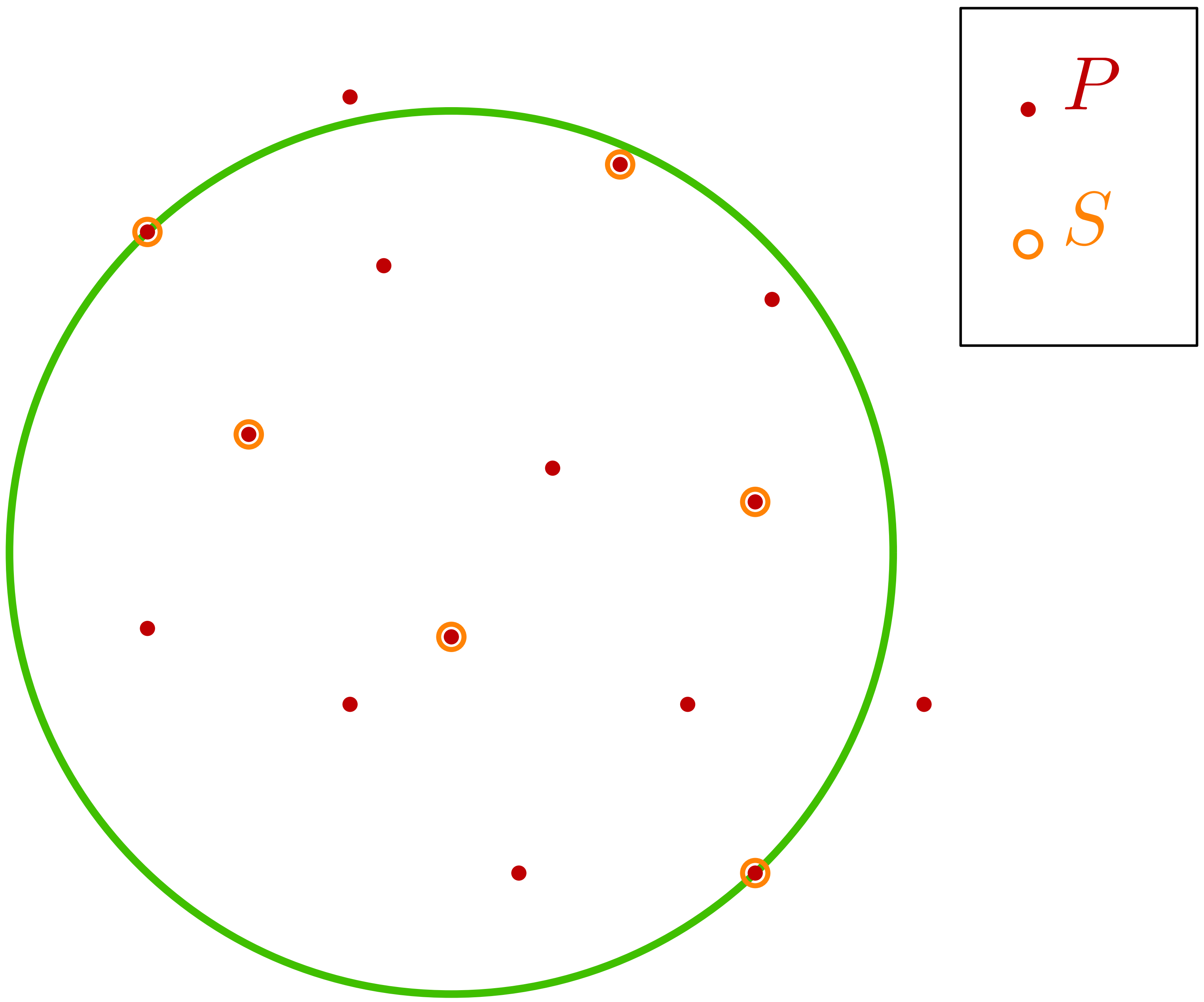
Given P, how many points do we need to sample ($S \subset P$), such that

the smallest enclosing disk contains 90% of the points in P? $\varepsilon = 0.1$

with probability 0.999 $= 1 - \varphi$

Question: Which range space?

If 10% of P outside a circle, then there should be a point of S outside the circle



Motivation: sampling for approximation

Given P, how many points do we need to sample ($S \subset P$), such that

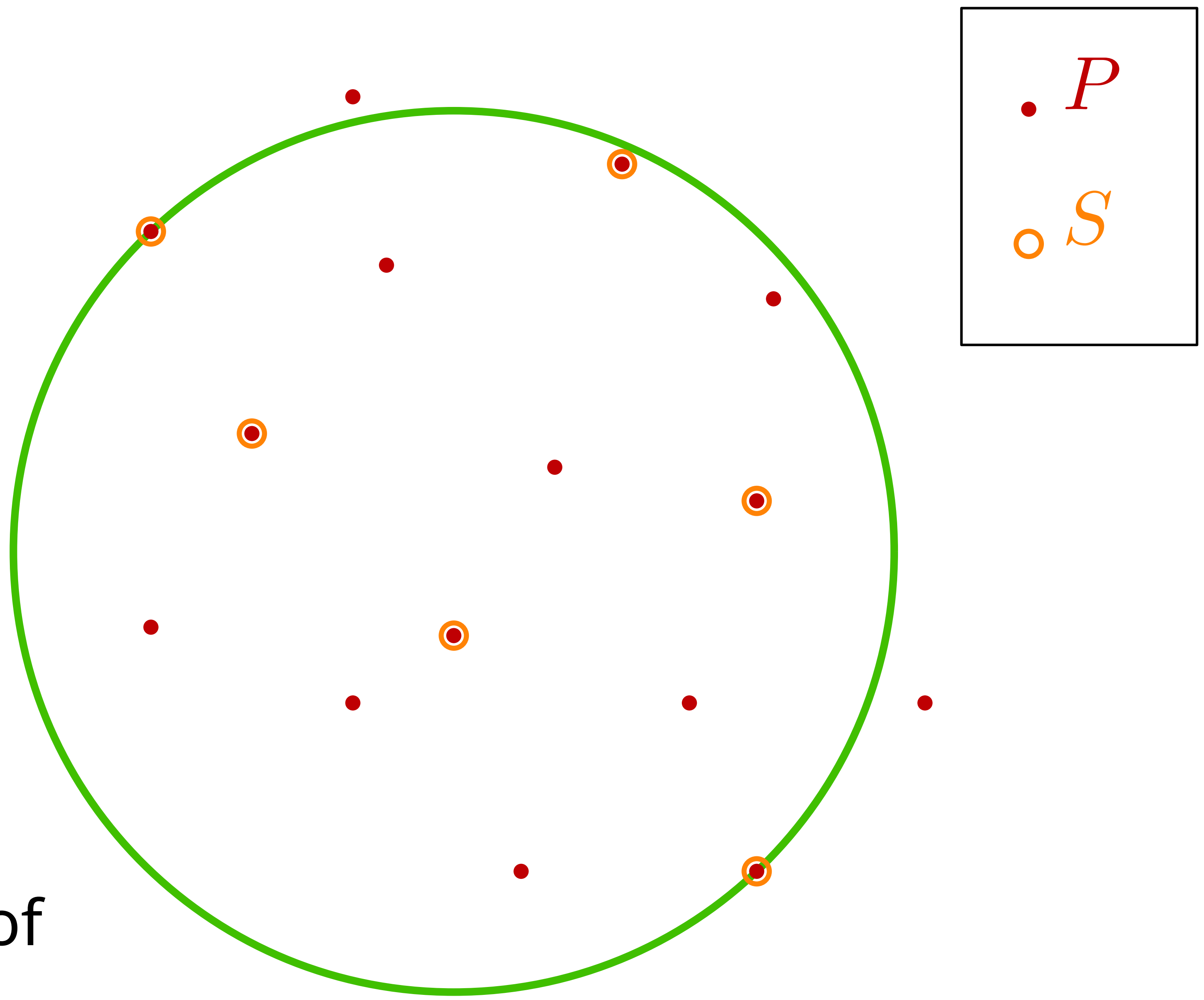
the smallest enclosing disk contains 90% of the points in P? $\varepsilon = 0.1$

with probability 0.999 $= 1 - \varphi$

Question: Which range space?

If 10% of P outside a circle, then there should be a point of S outside the circle

range space: $(\mathbb{R}^2, \mathcal{D}^c)$, with $\mathcal{D}^c$ the set of complements of disks.

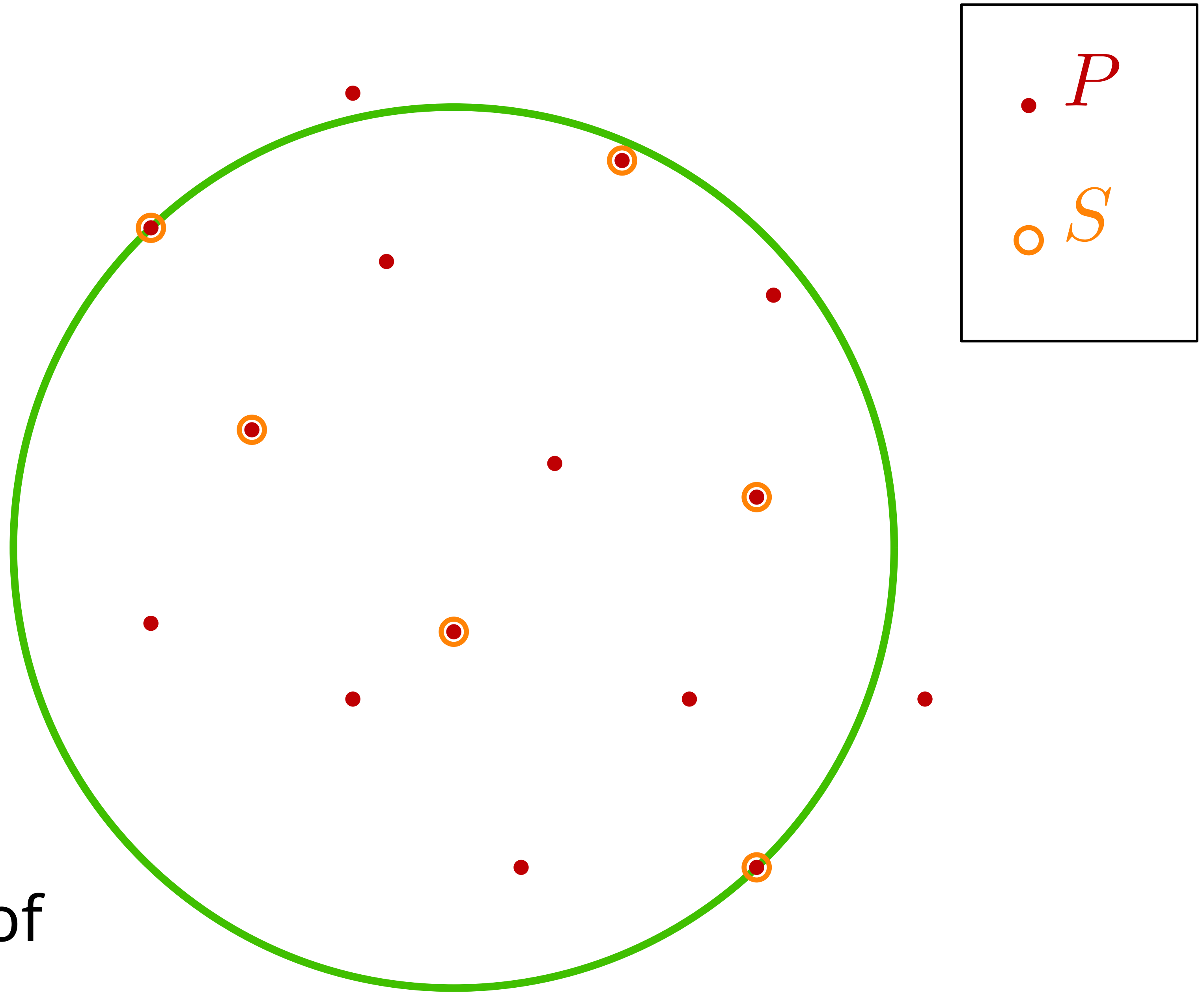


Motivation: sampling for approximation

Given P, how many points do we need to sample ($S \subset P$), such that

the smallest enclosing disk contains 90% of the points in P? $\varepsilon = 0.1$

with probability 0.999 $= 1 - \varphi$



Question: Which range space?

If 10% of P outside a circle, then there should be a point of S outside the circle

range space: $(\mathbb{R}^2, \mathcal{D}^c)$, with $\mathcal{D}^c$ the set of complements of disks.

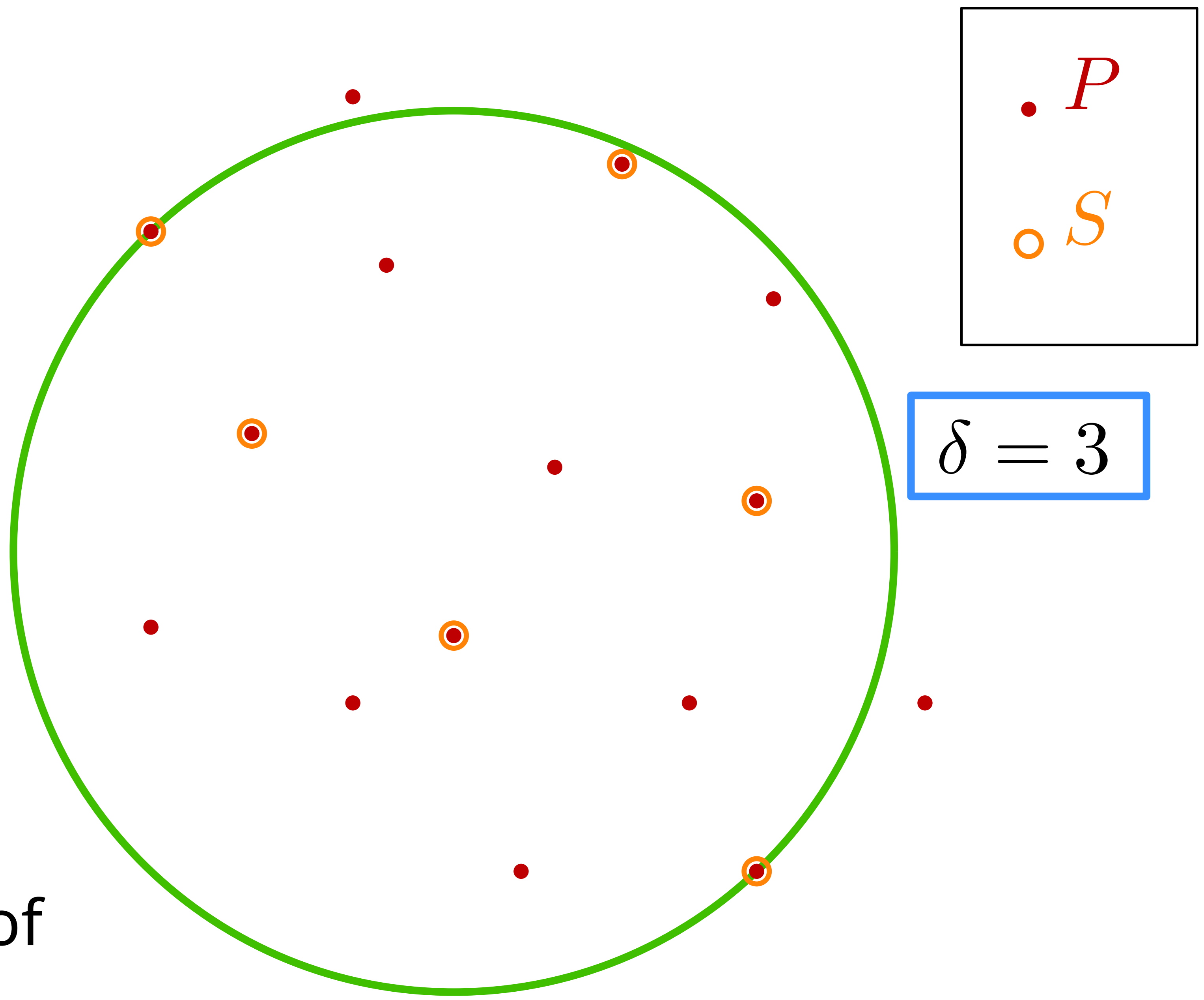
range space and its complement have same VC-dimension

Motivation: sampling for approximation

Given P, how many points do we need to sample ($S \subset P$), such that

the smallest enclosing disk contains 90% of the points in P? $\varepsilon = 0.1$

with probability 0.999 $= 1 - \varphi$



$\delta = 3$

Question: Which range space?

If 10% of P outside a circle, then there should be a point of S outside the circle

range space: $(\mathbb{R}^2, \mathcal{D}^c)$, with $\mathcal{D}^c$ the set of complements of disks.

range space and its complement have same VC-dimension

Motivation: sampling for approximation

Given P, how many points do we need to sample ($S \subset P$), such that

the smallest enclosing disk contains 90% of the points in P? $\varepsilon = 0.1$

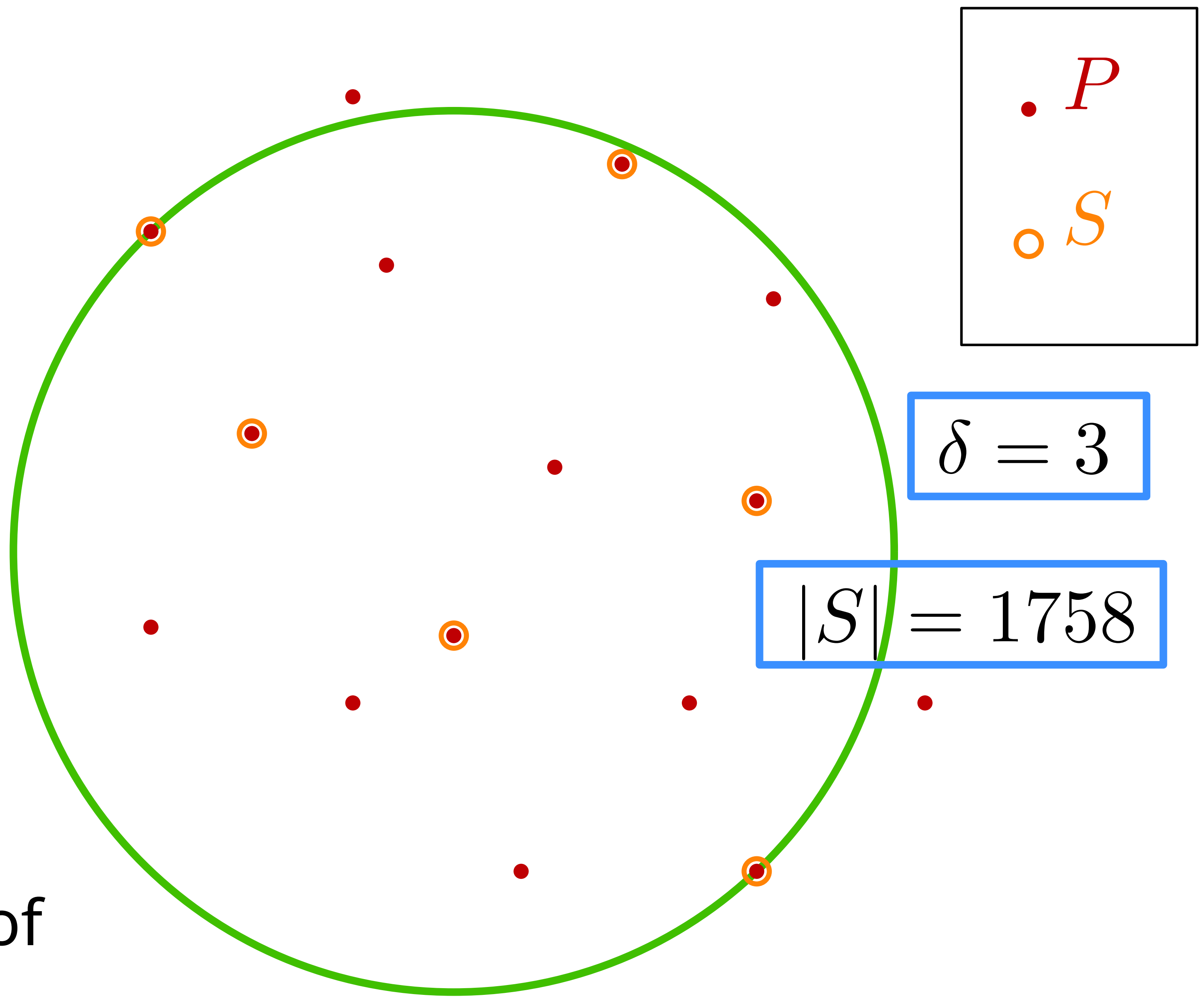
with probability 0.999 $= 1 - \varphi$

Question: Which range space?

If 10% of P outside a circle, then there should be a point of S outside the circle

range space: $(\mathbb{R}^2, \mathcal{D}^c)$, with $\mathcal{D}^c$ the set of complements of disks.

range space and its complement have same VC-dimension



ε-sample theorem, revisited

ε-sample (and -net) theorem use random sample.

ε-sample theorem, revisited

ε-sample (and -net) theorem use random sample.

It is also possible to construct an ε-sample of size $O(\frac{\log |\mathcal{R}|}{\varepsilon^2})$ deterministically.

ε-sample theorem, revisited

ε-sample (and -net) theorem use random sample.

It is also possible to construct an ε-sample of size $O(\frac{\log|\mathcal{R}|}{\varepsilon^2})$ deterministically.

Question: How large is $\log|\mathcal{R}|$?

Sauer's Lemma

bounding $|\mathcal{R}|$

Number of small subsets

Given $0 \leq d \leq n$, define $\Phi_d(n)$ to be the number of subsets of size at most d over a set of size n.

Number of small subsets

Given $0 \le d \le n$, define $\Phi_d(n)$ to be the number of subsets of size at most d over a set of size n.

$$\Phi_d(n) = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d} = \sum_{i=0}^{d} \binom{n}{i}$$

Number of small subsets

Given $0 \leq d \leq n$, define $\Phi_d(n)$ to be the number of subsets of size at most d over a set of size n.

$$\Phi_d(n) = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d} = \sum_{i=0}^{d} \binom{n}{i} \leq n^d$$

Number of small subsets

Given $0 \leq d \leq n$, define $\Phi_d(n)$ to be the number of subsets of size at most d over a set of size n.

$$\Phi_d(n) = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d} = \sum_{i=0}^{d} \binom{n}{i} \leq n^d$$

This function satisfies the following recurrence

$$\Phi_d(n) = \Phi_d(n-1) + \Phi_{d-1}(n-1)$$

Number of small subsets

Given $0 \leq d \leq n$, define $\Phi_d(n)$ to be the number of subsets of size at most d over a set of size n.

$$\Phi_d(n) = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d} = \sum_{i=0}^{d} \binom{n}{i} \leq n^d$$

This function satisfies the following recurrence

$$\Phi_d(n) = \Phi_d(n-1) + \Phi_{d-1}(n-1)$$

Intuition: Take element x: subsets don't contain x or do

Sauer's Lemma

If $(X, \mathcal{R})$ is a range space with VC-dimension d and $|X| = n$, then $|\mathcal{R}| \leq \Phi_d(n)$.

Sauer's Lemma

If $(X, \mathcal{R})$ is a range space with VC-dimension d and $|X| = n$, then $|\mathcal{R}| \leq \Phi_d(n)$.

proof

Induction on d and n

Sauer's Lemma

If $(X, \mathcal{R})$ is a range space with VC-dimension d and $|X| = n$, then $|\mathcal{R}| \leq \Phi_d(n)$.

proof

Induction on d and n

Base: $d = 0$ and $n = 0$ trivially true

Sauer's Lemma

If $(X, \mathcal{R})$ is a range space with VC-dimension d and $|X| = n$, then $|\mathcal{R}| \leq \Phi_d(n)$.

proof

Step:

$\mathcal{R}_x = \{Q \setminus \{x\} : Q \cup \{x\} \in \mathcal{R}$ and $Q \setminus \{x\} \in \mathcal{R}\}$

$\mathcal{R} \setminus x = \{Q \setminus \{x\} : Q \in \mathcal{R}\}$

Sauer's Lemma

If $(X, \mathcal{R})$ is a range space with VC-dimension d and $|X| = n$, then $|\mathcal{R}| \leq \Phi_d(n)$.

proof

Step:

$\mathcal{R}_x = \{Q \setminus \{x\} : Q \cup \{x\} \in \mathcal{R}$ and $Q \setminus \{x\} \in \mathcal{R}\}$

$\mathcal{R} \setminus x = \{Q \setminus \{x\} : Q \in \mathcal{R}\}$

claim: $|\mathcal{R}| = |\mathcal{R}_x| + |\mathcal{R} \setminus x|$

Sauer's Lemma

If $(X, \mathcal{R})$ is a range space with VC-dimension d and $|X| = n$, then $|\mathcal{R}| \leq \Phi_d(n)$.

proof

Step:

$\mathcal{R}_x = \{Q \setminus \{x\} : Q \cup \{x\} \in \mathcal{R}$ and $Q \setminus \{x\} \in \mathcal{R}\}$

$\mathcal{R} \setminus x = \{Q \setminus \{x\} : Q \in \mathcal{R}\}$

claim: $|\mathcal{R}| = |\mathcal{R}_x| + |\mathcal{R} \setminus x|$

Charge each range of $\mathcal{R}$ to corresponding range in $\mathcal{R} \setminus \{x\}$

Sauer's Lemma

If $(X, \mathcal{R})$ is a range space with VC-dimension d and $|X| = n$, then $|\mathcal{R}| \leq \Phi_d(n)$.

proof

Step:

$\mathcal{R}_x = \{Q \setminus \{x\} : Q \cup \{x\} \in \mathcal{R}$ and $Q \setminus \{x\} \in \mathcal{R}\}$

$\mathcal{R} \setminus x = \{Q \setminus \{x\} : Q \in \mathcal{R}\}$

claim: $|\mathcal{R}| = |\mathcal{R}_x| + |\mathcal{R} \setminus x|$

Charge each range of $\mathcal{R}$ to corresponding range in $\mathcal{R} \setminus \{x\}$

Range r with $r \cup \{x\} \in \mathcal{R}$ and $r \setminus \{x\} \in \mathcal{R}$ charged twice ?!

Sauer's Lemma

If $(X, \mathcal{R})$ is a range space with VC-dimension d and $|X| = n$, then $|\mathcal{R}| \leq \Phi_d(n)$.

proof

Step:

$\mathcal{R}_x = \{Q \setminus \{x\} : Q \cup \{x\} \in \mathcal{R}$ and $Q \setminus \{x\} \in \mathcal{R}\}$

$\mathcal{R} \setminus x = \{Q \setminus \{x\} : Q \in \mathcal{R}\}$

claim: $|\mathcal{R}| = |\mathcal{R}_x| + |\mathcal{R} \setminus x|$

Charge each range of $\mathcal{R}$ to corresponding range in $\mathcal{R} \setminus \{x\}$

Range r with $r \cup \{x\} \in \mathcal{R}$ and $r \setminus \{x\} \in \mathcal{R}$ charged twice ?!

These are exactly elements in $\mathcal{R}_x$!

Sauer's Lemma

If $(X, \mathcal{R})$ is a range space with VC-dimension d and $|X| = n$, then $|\mathcal{R}| \leq \Phi_d(n)$.

proof

$\mathcal{R}_x = \{Q \setminus \{x\}$: $Q \cup \{x\} \in \mathcal{R}$ and $Q \setminus \{x\} \in \mathcal{R}\}$

$|\mathcal{R}| = |\mathcal{R}_x| + |\mathcal{R} \setminus x|$

claim: $(X \setminus \{x\}, \mathcal{R}_x)$ has VC-dimension at most $d - 1$

Sauer's Lemma

If $(X, \mathcal{R})$ is a range space with VC-dimension d and $|X| = n$, then $|\mathcal{R}| \leq \Phi_d(n)$.

proof

$\mathcal{R}_x = \{Q \setminus \{x\} : Q \cup \{x\} \in \mathcal{R}$ and $Q \setminus \{x\} \in \mathcal{R}\}$

$|\mathcal{R}| = |\mathcal{R}_x| + |\mathcal{R} \setminus x|$

claim: $(X \setminus \{x\}, \mathcal{R}_x)$ has VC-dimension at most $d - 1$

If $B \subset X \setminus \{x\}$ is shattered by $\mathcal{R}_x$, then

$B \cup \{x\}$ is shattered in $\mathcal{R}$

Sauer's Lemma

If $(X, \mathcal{R})$ is a range space with VC-dimension d and $|X| = n$, then $|\mathcal{R}| \leq \Phi_d(n)$.

proof

$\mathcal{R}_x = \{Q \setminus \{x\} : Q \cup \{x\} \in \mathcal{R}$ and $Q \setminus \{x\} \in \mathcal{R}\}$

$|\mathcal{R}| = |\mathcal{R}_x| + |\mathcal{R} \setminus x|$

claim: $(X \setminus \{x\}, \mathcal{R}_x)$ has VC-dimension at most $d - 1$

If $B \subset X \setminus \{x\}$ is shattered by $\mathcal{R}_x$, then

$B \cup \{x\}$ is shattered in $\mathcal{R}$

$(X \setminus \{x\}, \mathcal{R}_x)$ has smaller VC-dim. than $(X, \mathcal{R})$

Sauer's Lemma

If $(X, \mathcal{R})$ is a range space with VC-dimension d and $|X| = n$, then $|\mathcal{R}| \leq \Phi_d(n)$.

proof

$\mathcal{R}_x = \{Q \setminus \{x\} : Q \cup \{x\} \in \mathcal{R}$ and $Q \setminus \{x\} \in \mathcal{R}\}$
$|\mathcal{R}| = |\mathcal{R}_x| + |\mathcal{R} \setminus x|$

claim: $(X \setminus \{x\}, \mathcal{R}_x)$ has VC-dimension at most $d - 1$

Thus, by induction hypothesis:

$$|\mathcal{R}| \leq \Phi_{d-1}(n-1) + \Phi_d(n-1)$$

Sauer's Lemma

If $(X, \mathcal{R})$ is a range space with VC-dimension d and $|X| = n$, then $|\mathcal{R}| \leq \Phi_d(n)$.

proof

$\mathcal{R}_x = \{Q \setminus \{x\} : Q \cup \{x\} \in \mathcal{R}$ and $Q \setminus \{x\} \in \mathcal{R}\}$

$|\mathcal{R}| = |\mathcal{R}_x| + |\mathcal{R} \setminus x|$

claim: $(X \setminus \{x\}, \mathcal{R}_x)$ has VC-dimension at most $d - 1$

Thus, by induction hypothesis:

$$|\mathcal{R}| \leq \Phi_{d-1}(n-1) + \Phi_d(n-1) = \Phi_d(n)$$

Quiz

Which bound on $O\left(\frac{\log|\mathcal{R}|}{\varepsilon^2}\right)$ does the previous lemma give for $(X, \mathcal{R})$ with $n = |X|$ and VC-dimension δ?

A $O\left(\frac{\delta}{\varepsilon^2}\right)$

B $O\left(\frac{\delta \log n}{\varepsilon^2}\right)$

C $O\left(\frac{\delta n}{\varepsilon^2}\right)$

Quiz

Which bound on $O\left(\frac{\log|\mathcal{R}|}{\varepsilon^2}\right)$ does the previous lemma give for $(X, \mathcal{R})$ with $n = |X|$ and VC-dimension δ?

A $\quad O\left(\frac{\delta}{\varepsilon^2}\right)$

B $\quad O\left(\frac{\delta \log n}{\varepsilon^2}\right)$

C $\quad O\left(\frac{\delta n}{\varepsilon^2}\right)$

VC-dim $\delta \quad \Rightarrow \quad |\mathcal{R}| \leq n^\delta$

Quiz

Which bound on $O\left(\frac{\log|\mathcal{R}|}{\varepsilon^2}\right)$ does the previous lemma give for $(X, \mathcal{R})$ with $n = |X|$ and VC-dimension δ?

A $\quad O\left(\frac{\delta}{\varepsilon^2}\right)$

B $\quad O\left(\frac{\delta \log n}{\varepsilon^2}\right)$

C $\quad O\left(\frac{\delta n}{\varepsilon^2}\right)$

VC-dim $\delta \quad \Rightarrow \quad |\mathcal{R}| \leq n^\delta$

What does $|\mathcal{R}| = O(n^d)$ imply about the VC-dimension?

Shattering dimension

Shattering Dimension

Given a range space $S = (X, \mathcal{R})$, its shatter function $\pi_S(m)$ is the maximum number of sets that might be created by S when restricted to subsets of size m. Formally,

$$\pi_S(m) = \max_{\substack{B \subset X \\ |B| = m}} |R_{|B}|$$

Shattering Dimension

Given a range space $S = (X, \mathcal{R})$, its shatter function $\pi_S(m)$ is the maximum number of sets that might be created by S when restricted to subsets of size m. Formally,

$$\pi_S(m) = \max_{\substack{B \subset X \\ |B| = m}} |R_{|B}|$$

The shattering dimension of S is the smallest d such that $\pi_S(m) = O(m^d)$, for all m

Shattering Dimension

Given a range space $S = (X, \mathcal{R})$, its shatter function $\pi_S(m)$ is the maximum number of sets that might be created by S when restricted to subsets of size m. Formally,

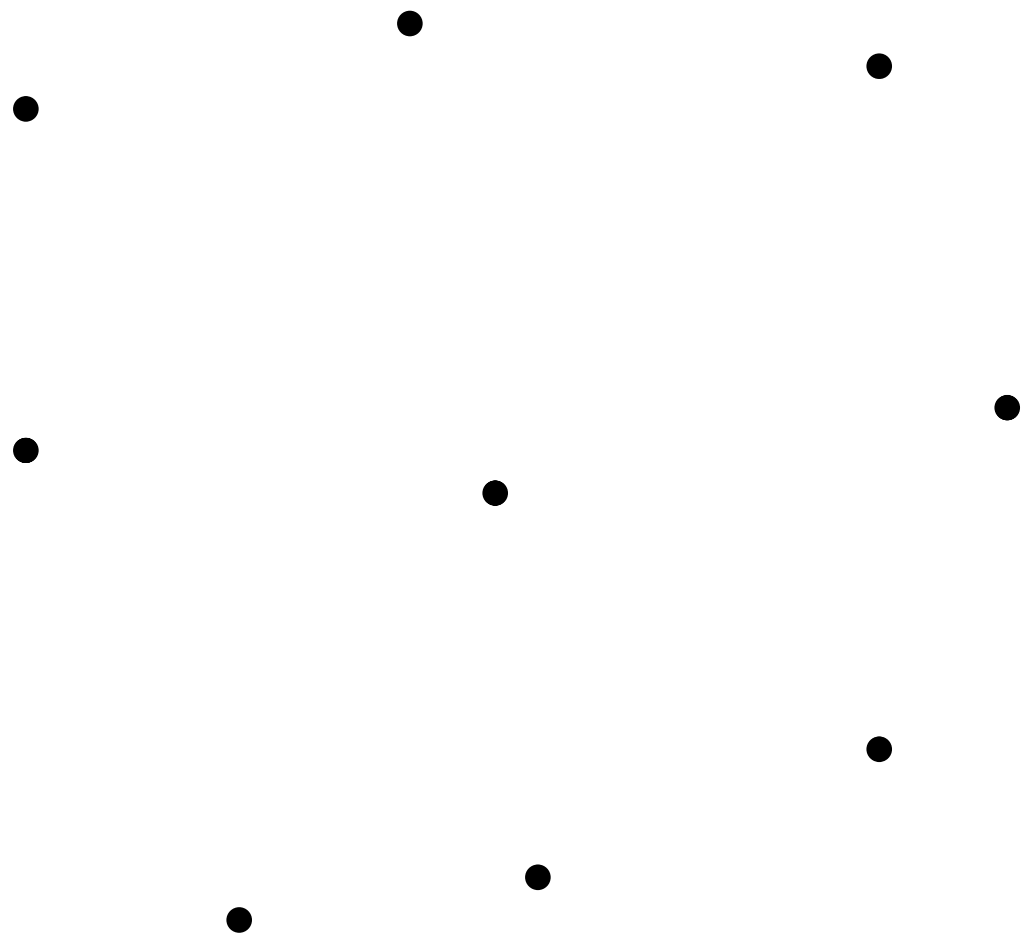
$$\pi_S(m) = \max_{\substack{B \subset X \\ |B| = m}} |R_{|B}|$$

The shattering dimension of S is the smallest d such that $\pi_S(m) = O(m^d)$, for all m

Sauer's lemma: shattering dimension $\leq$ VC-dimension

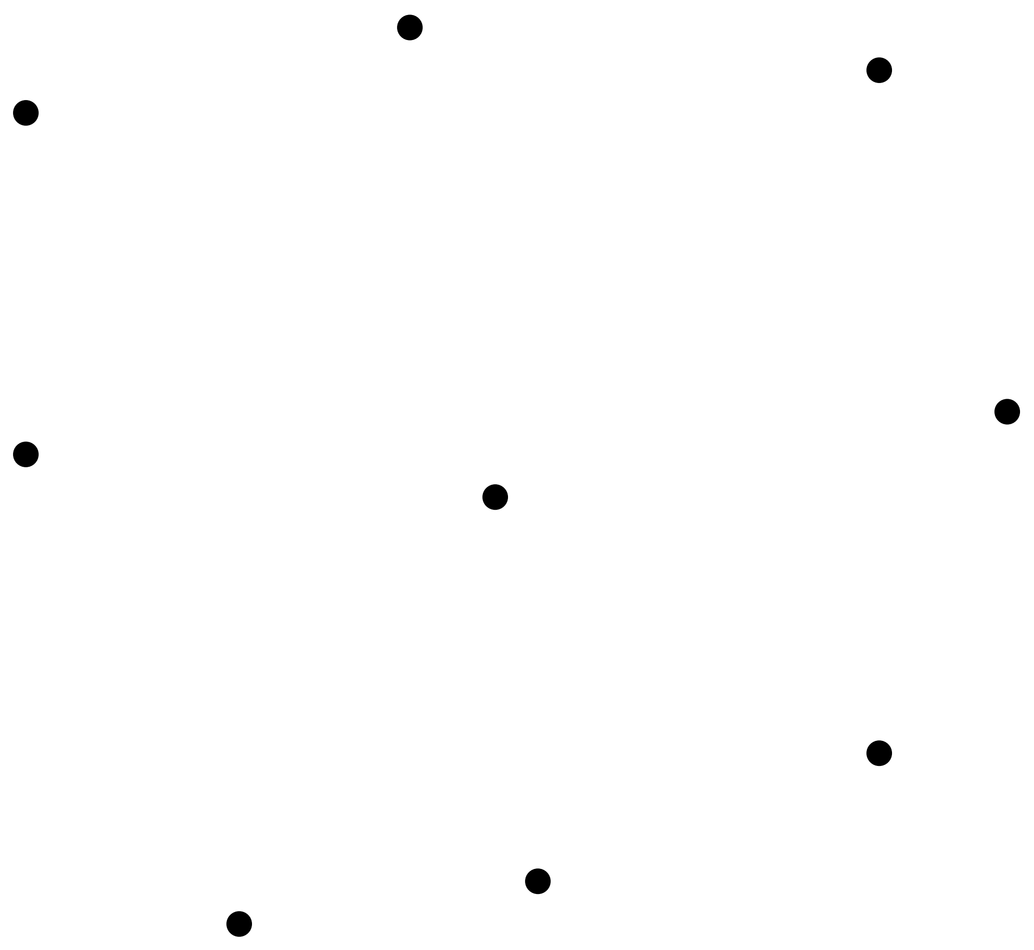
Examples of Shattering Dimension

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D} =$ set of disks



Examples of Shattering Dimension

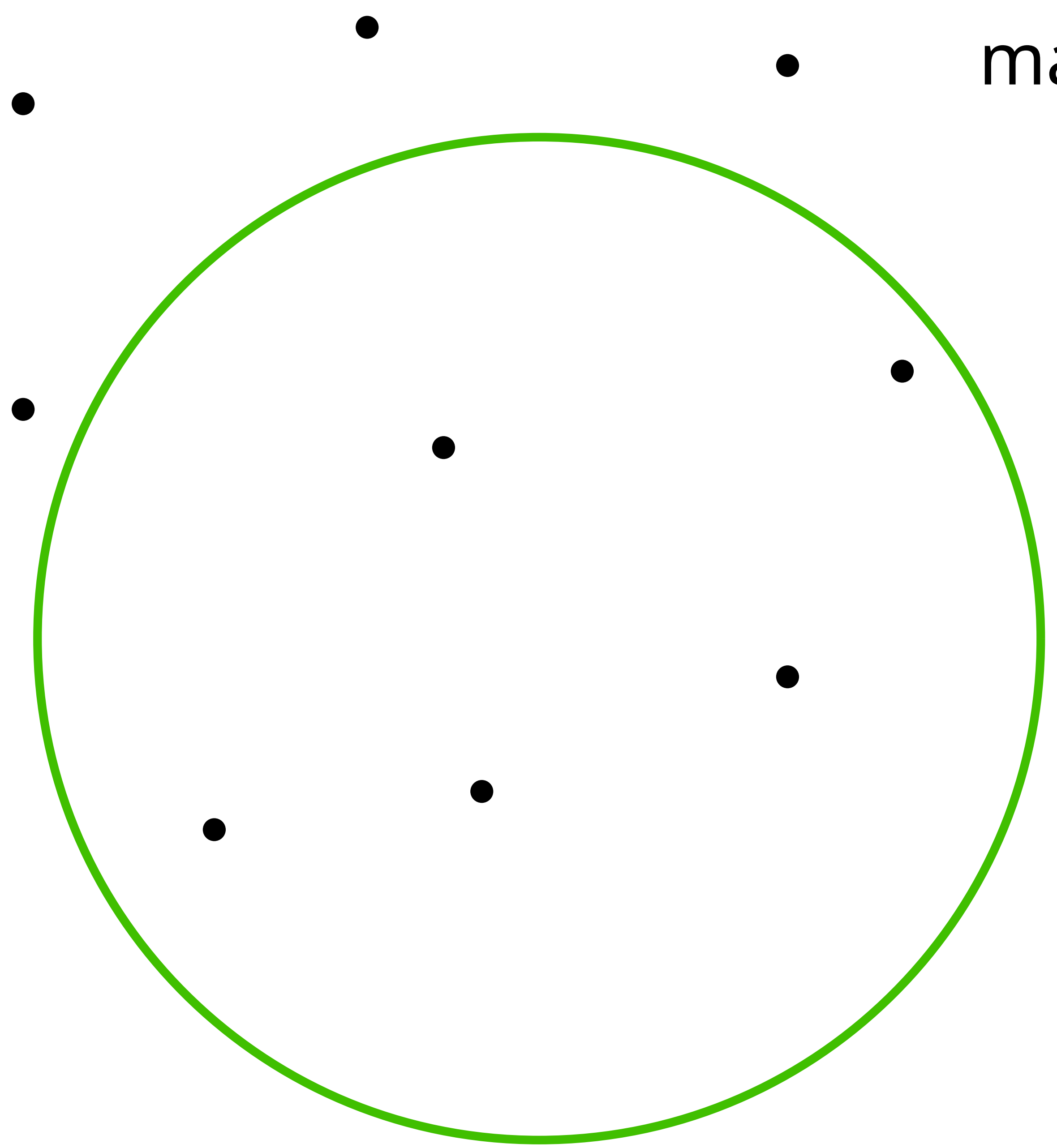
range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D} =$ set of disks



Question: When restricted to n points, how many ranges are there?

Examples of Shattering Dimension

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D}$ = set of disks

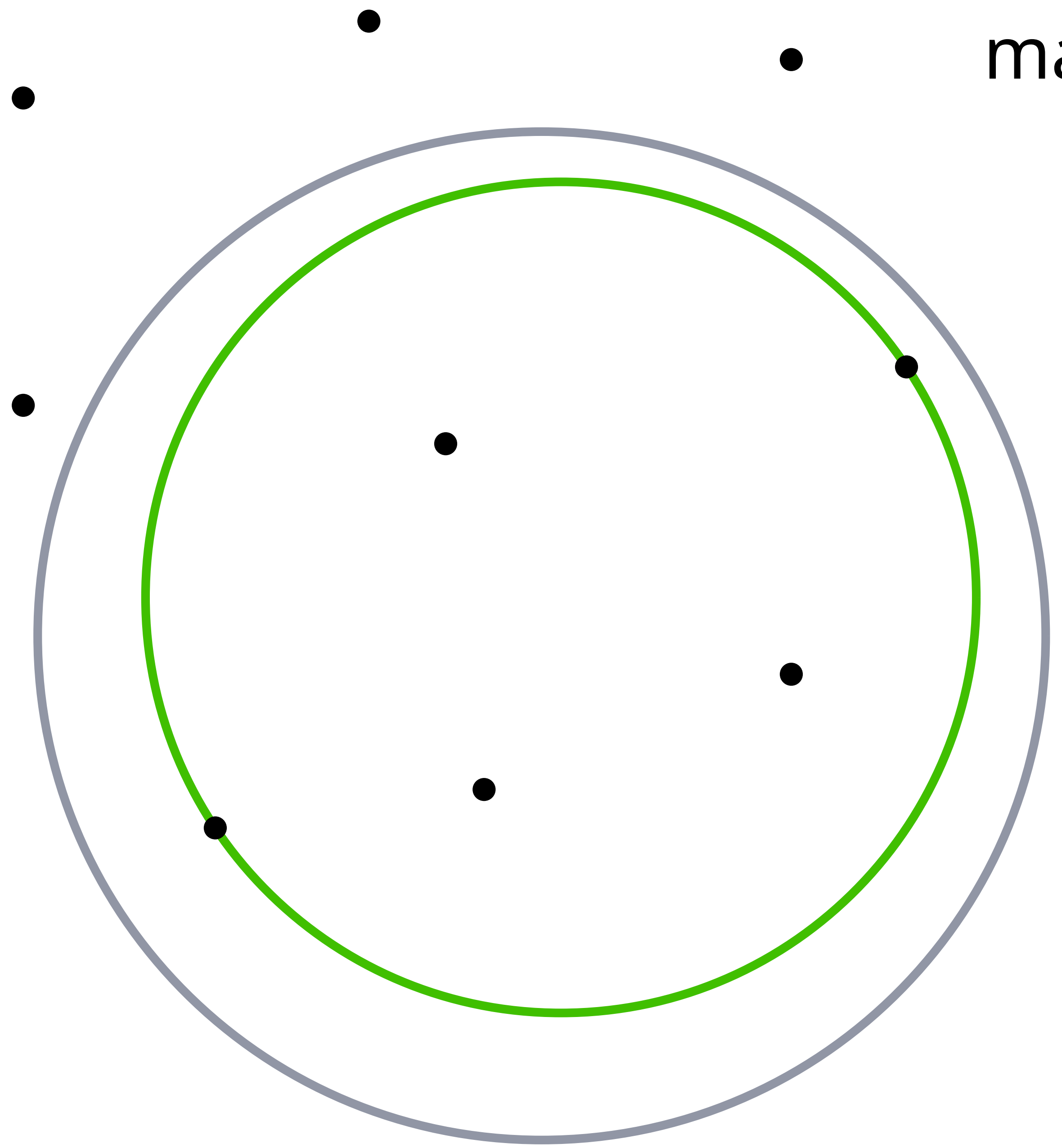


Question: When restricted to n points, how many ranges are there?

consider a disk

Examples of Shattering Dimension

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D}$ = set of disks



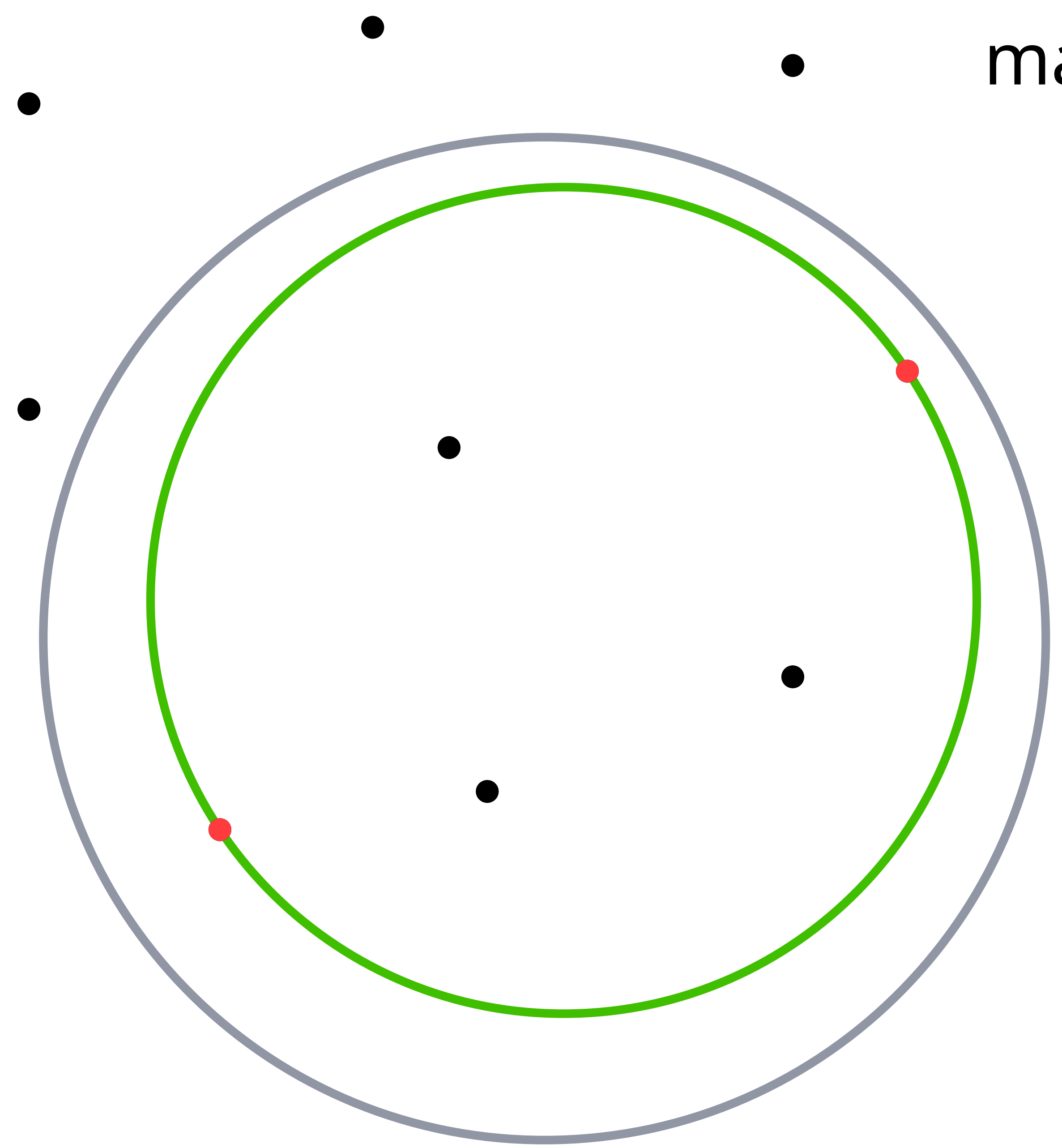
Question: When restricted to n points, how many ranges are there?

consider a disk

canonical disk for points in range: smallest enclosing disk

Examples of Shattering Dimension

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D} =$ set of disks



Question: When restricted to n points, how many ranges are there?

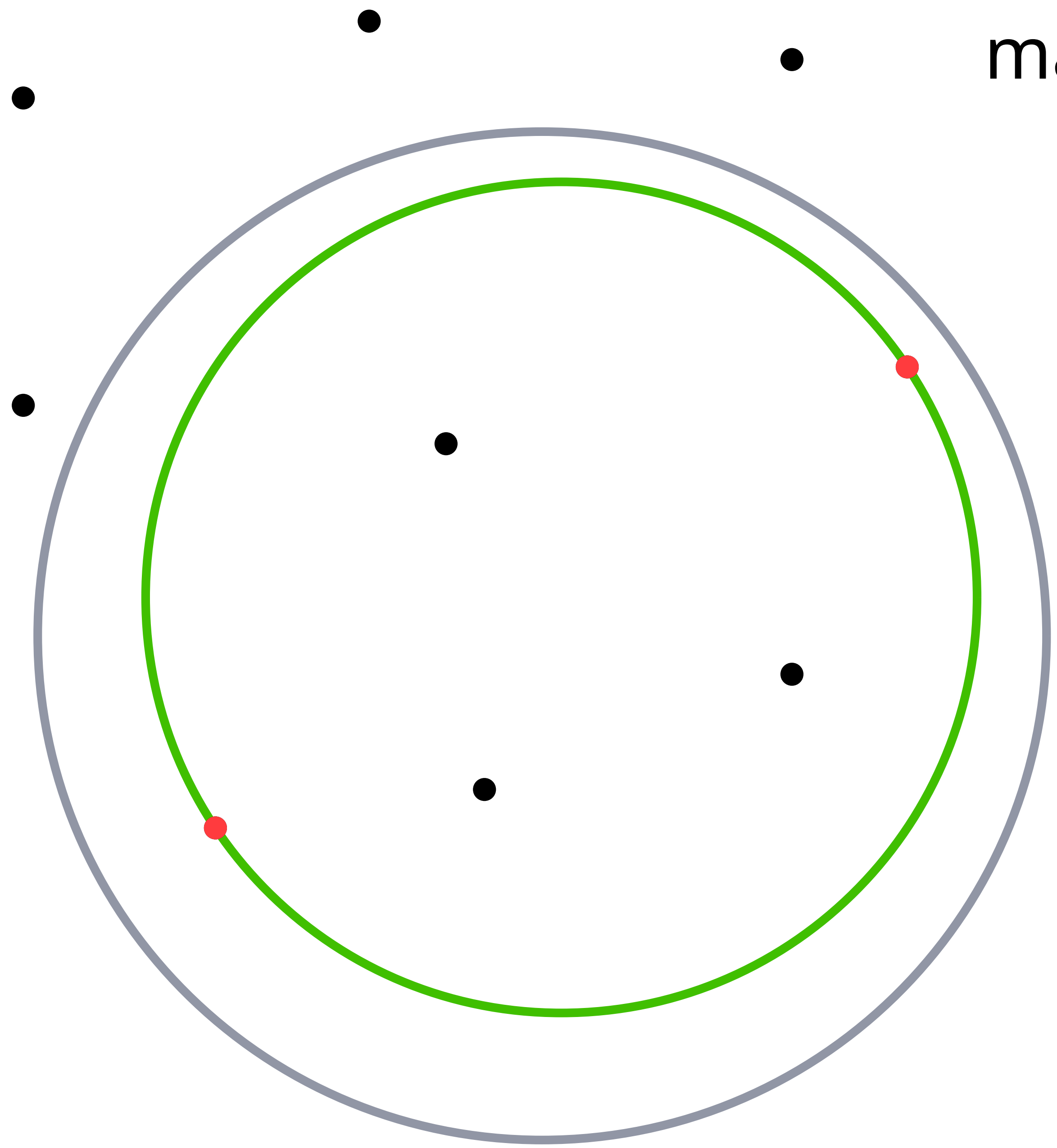
consider a disk

canonical disk for points in range:
smallest enclosing disk

uniquely determined by ≤ 3 points on boundary

Examples of Shattering Dimension

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D} =$ set of disks



Question: When restricted to n points, how many ranges are there?

consider a disk

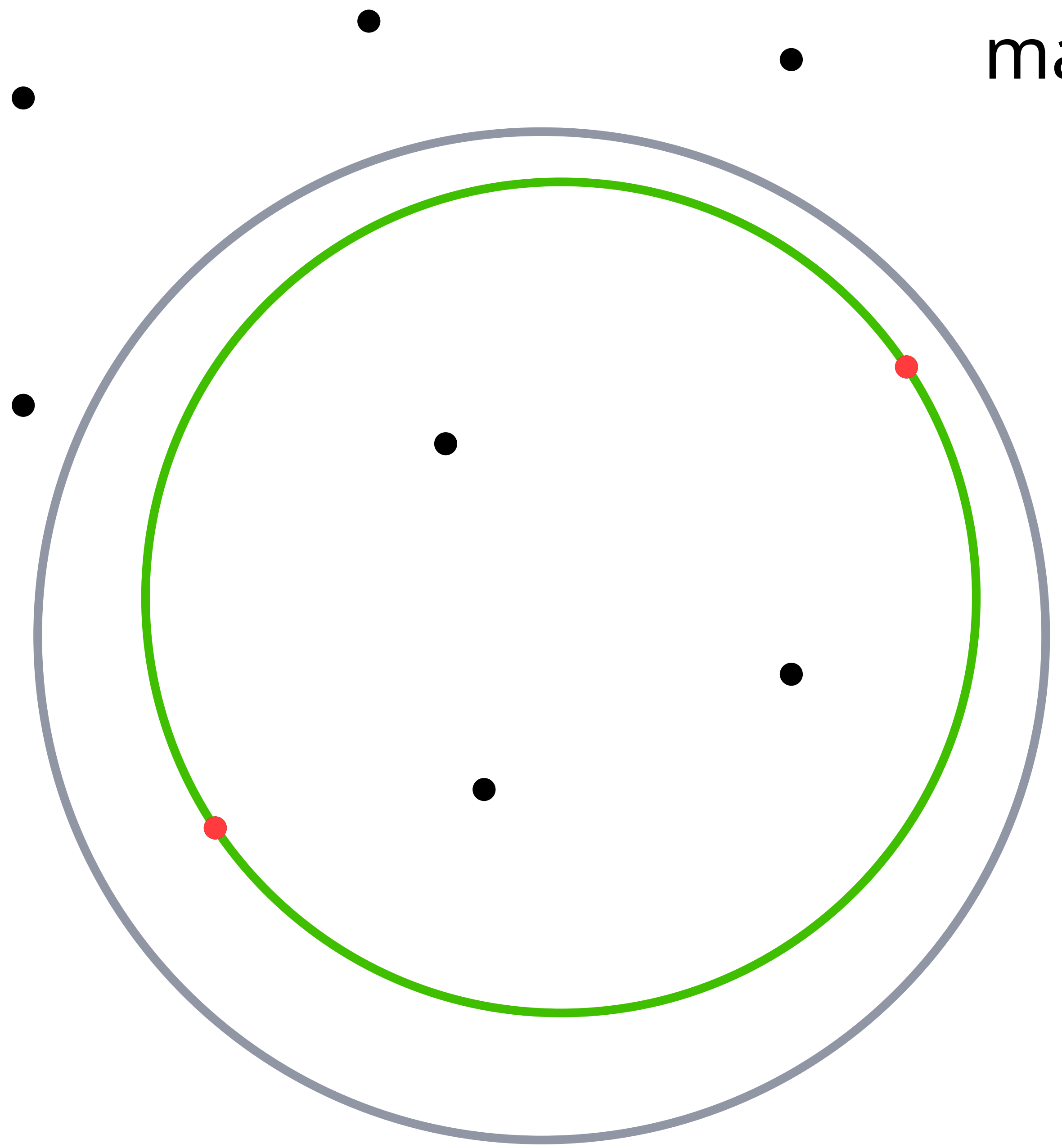
canonical disk for points in range: smallest enclosing disk

uniquely determined by ≤ 3 points on boundary

$|\mathcal{R}_{|P}| \leq \Phi_d(n) \leq n^3$

Examples of Shattering Dimension

range space $(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D} =$ set of disks



Question: When restricted to n points, how many ranges are there?

consider a disk

canonical disk for points in range:
smallest enclosing disk

uniquely determined by ≤ 3 points on boundary

$|\mathcal{R}_{|P}| \leq \Phi_d(n) \leq n^3$

shattering dim ≤ 3

Shattering dimension of geometric range spaces

shattering dimension $\approx$ how many points determine a range

Shattering dimension of geometric range spaces

shattering dimension $\approx$ how many points determine a range

range space	shattering dimension
$(\mathbb{R}, \mathcal{I})$, with $\mathcal{I}$ $=$ set of closed intervals ?	
$(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D}$ $=$ set of disks	3
$(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ $=$ set of axis-aligned rectangles	
$(\mathbb{R}^2, \mathcal{GR})$, with $\mathcal{GR}$ $=$ set of arbitrary oriented rectangles	
$(\mathbb{R}^2, \mathcal{C})$, with $\mathcal{C}$ $=$ set of closed convex sets	–

Shattering dimension of geometric range spaces

shattering dimension $\approx$ how many points determine a range

range space	shattering dimension
$(\mathbb{R}, \mathcal{I})$, with $\mathcal{I} =$ set of closed intervals	2
$(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D} =$ set of disks	3
$(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR} =$ set of axis-aligned rectangles ?	
$(\mathbb{R}^2, \mathcal{GR})$, with $\mathcal{GR} =$ set of arbitrary oriented rectangles	
$(\mathbb{R}^2, \mathcal{C})$, with $\mathcal{C} =$ set of closed convex sets	–

Shattering dimension of geometric range spaces

shattering dimension $\approx$ how many points determine a range

range space	shattering dimension
$(\mathbb{R}, \mathcal{I})$, with $\mathcal{I}$ = set of closed intervals	2
$(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D}$ = set of disks	3
$(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR}$ = set of axis-aligned rectangles	4
$(\mathbb{R}^2, \mathcal{GR})$, with $\mathcal{GR}$ = set of arbitrary oriented rectangles	
$(\mathbb{R}^2, \mathcal{C})$, with $\mathcal{C}$ = set of closed convex sets	–

Shattering dimension of geometric range spaces

shattering dimension $\approx$ how many points determine a range

range space	shattering dimension
$(\mathbb{R}, \mathcal{I})$, with $\mathcal{I} =$ set of closed intervals	2
$(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D} =$ set of disks	3
$(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR} =$ set of axis-aligned rectangles	4
$(\mathbb{R}^2, \mathcal{GR})$, with $\mathcal{GR} =$ set of arbitrary oriented rectangles	5
$(\mathbb{R}^2, \mathcal{C})$, with $\mathcal{C} =$ set of closed convex sets	–

Shattering dimension of geometric range spaces

shattering dimension $\approx$ how many points determine a range

range space	shattering dimension
$(\mathbb{R}, \mathcal{I})$, with $\mathcal{I} =$ set of closed intervals	2
$(\mathbb{R}^2, \mathcal{D})$, with $\mathcal{D} =$ set of disks	3
$(\mathbb{R}^2, \mathcal{AR})$, with $\mathcal{AR} =$ set of axis-aligned rectangles	4
$(\mathbb{R}^2, \mathcal{GR})$, with $\mathcal{GR} =$ set of arbitrary oriented rectangles	5
$(\mathbb{R}^2, \mathcal{C})$, with $\mathcal{C} =$ set of closed convex sets	–

Can be easier to compute than VC-dimension

Shattering dimension vs VC-dimension

VC-dimension δ

shattering dimension d

Shattering dimension vs VC-dimension

VC-dimension δ

shattering dimension d

Sauer's lemma: $d \leq \delta$

Shattering dimension vs VC-dimension

VC-dimension δ

shattering dimension d

Sauer's lemma: $d \leq \delta$

claim: $\delta \leq O(d \log d)$

Shattering dimension vs VC-dimension

VC-dimension δ

shattering dimension d

Sauer's lemma: $d \leq \delta$

claim: $\delta \leq O(d \log d)$

Consider largest shattered $N \subset X$: $\quad \delta = |N|$

Shattering dimension vs VC-dimension

VC-dimension δ

shattering dimension d

Sauer's lemma: $d \leq \delta$

claim: $\delta \leq O(d \log d)$

Consider largest shattered $N \subset X$: $\quad \delta = |N|$

$2^\delta = \mathcal{R}_{|N} \leq c\delta^d$

Shattering dimension vs VC-dimension

VC-dimension δ

shattering dimension d

Sauer's lemma: $d \leq \delta$

claim: $\delta \leq O(d \log d)$

Consider largest shattered $N \subset X$: $\quad \delta = |N|$

$2^\delta = \mathcal{R}_{|N} \leq c\delta^d$

$\delta \leq \log(c) + d \log \delta$

Shattering dimension vs VC-dimension

VC-dimension δ

shattering dimension d

Sauer's lemma: $d \leq \delta$

claim: $\delta \leq O(d \log d)$

Consider largest shattered $N \subset X$: $\quad \delta = |N|$

$2^\delta = \mathcal{R}_{|N} \leq c\delta^d$

$\delta \leq \log(c) + d \log \delta$

$\log \delta \leq \log(\log(c) + d \log \delta) = O(\log(d \log \delta)) = O(\log d + \log \log \delta)$

Shattering dimension vs VC-dimension

VC-dimension δ

shattering dimension d

Sauer's lemma: $d \leq \delta$

claim: $\delta \leq O(d \log d)$

Consider largest shattered $N \subset X$: $\quad \delta = |N|$

$2^\delta = \mathcal{R}_{|N} \leq c\delta^d$

$\delta \leq \log(c) + d \log \delta$

$\log \delta \leq \log(\log(c) + d \log \delta) = O(\log(d \log \delta)) = O(\log d + \log \log \delta)$

$\delta \leq O(d \log \delta) = O(d \log d)$

Summary

range space $(X, \mathcal{R})$

VC-dimension δ

examples of geometric range spaces

ε-sample of size $O\left(\frac{\delta + \log \varphi^{-}1}{\varepsilon^2}\right)$

ε-net of size $O\left(\frac{\delta \log \varepsilon^{-1} + \log \varphi^{-}1}{\varepsilon}\right)$

applications for geometric approximation

shattering dimension d

$d \leq \delta \leq d \log d$