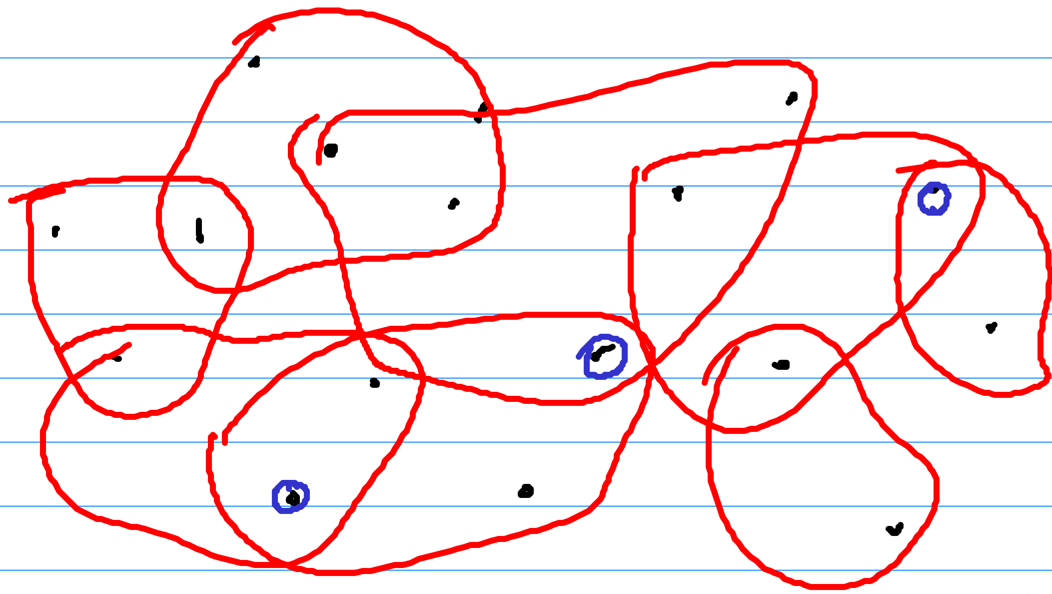
Computational Geometry Lecture 35

Topic : ε-nets & V.C. dimension

Vapnik-Chevronenkis

n: population size



Special Interest groups (with common interests)

Only the "large" groups matter

$\geq \varepsilon \cdot n$

$0 < \varepsilon < 1$

We want to choose a subset of size r s.t.

(i) We must have at least one representative from each of the "large" subset.

(ii) If m is the size of a subset ($m > \epsilon n$) then the number of representative $r(m)$ should be such that

$$\frac{r(m)}{r} \sim \frac{m}{n}$$

How small can r be to guarantee conditions (i) & (ii)

Choose a vertex independently with probability $p = \frac{1}{\epsilon n} \ln n$ to be in the sample

What is the prob that a "large subset" is not hit?

$$(1-p)^{\epsilon n} = \left(1 - \frac{1}{\epsilon n}\ln n\right)^{\epsilon n \cdot c}$$

$$= \left[\underbrace{\left(1 - \frac{1}{\frac{\epsilon n}{\log n}}\right)^{\frac{\epsilon n}{\log n}}}_{\leq \frac{1}{e}} \right]^{\log n} \leq \frac{1}{n}$$

The probability that we miss even one among all the large subsets $\leq \frac{1}{n} \cdot$ # large subs

If this is less than 1 then <u>some</u> sample (of size $c \cdot \frac{\ln n}{\varepsilon}$) will hit all large subsets

$\Rightarrow$ there must be at least one such sample that hits all "large subsets"

When can we do away with "$\ln n$" factor

Range space

$S = (X, \mathbb{R})$

X : set of points (may be infinite)

$\mathbb{R}$: family of subsets of X.

Eg.

(1) X : set of points in the real line

$\mathbb{R}$: closed intervals on real line.

(2) X : set of points on the plane

$\mathbb{R}$: set of all closed disks

If $Y \subset X$, then

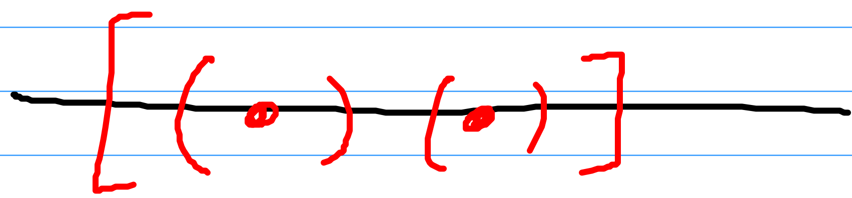
$$S_{|Y} = (Y, \mathbb{R}_{|Y})$$

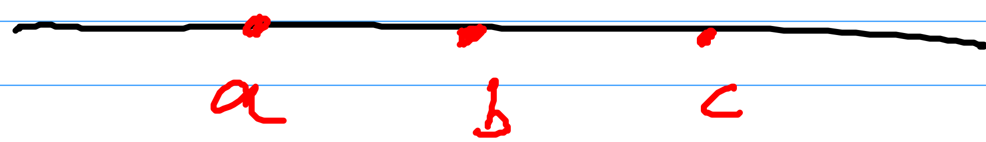
$$\mathbb{R}_{|Y} = \{ Y \cap r \mid r \in \mathbb{R} \}$$

In particular, Y can be a finite set of points

We say that a finite subset of points Y is <u>shattered</u> if

$$|R_{|Y}| = |2^Y|$$

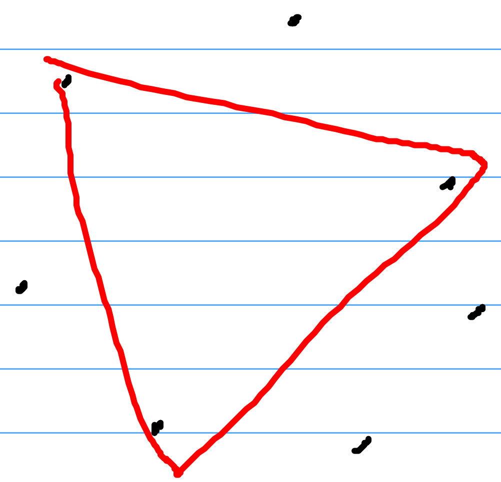




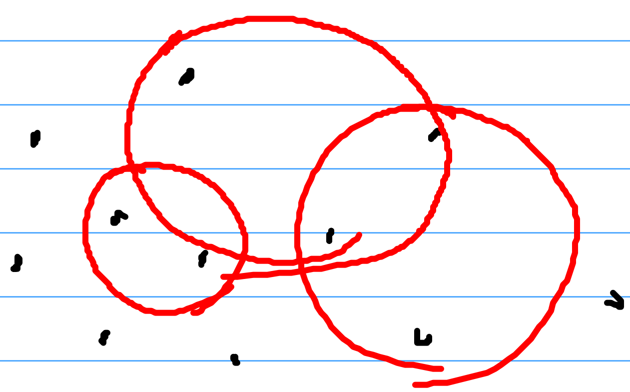
The VC dimension of a range space $S = (X, R)$ is δ iff for all finite Y, $|Y| > \delta$, Y cannot be shattered, and there exists a subset of size δ that can be shattered.

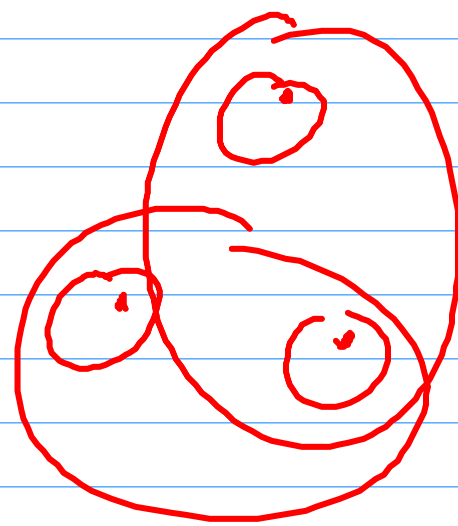
<u>Comment</u>: Some Range spaces may have $\delta = \infty$

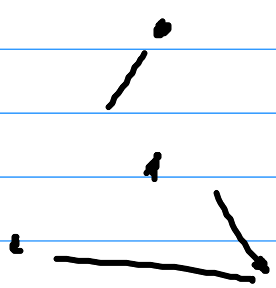
$$S_{conv} = \{\mathbb{R}^2, \text{ convex subsets}\}$$

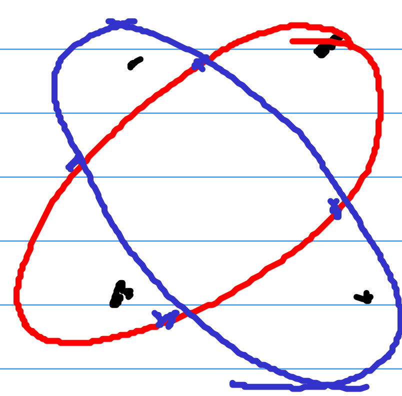


$$S' = \{\mathbb{R}^2, \text{ disks}\}$$









$S'' = \{ \mathbb{R}^2, \text{half-planes} \}$

A Range space with finite VC dimension has $\underbrace{\varepsilon\text{-net}}$

covering all subsets of size $\geq \varepsilon \cdot n$

of size that is only a function of ε.