Extended Abstracts

VC-Dimensions and ε-Nets

Based on Chapter 10 of Jiří Matoušek's *Lectures on Discrete Geometry* and Chapter 15 of János Pach and Pankaj K. Agarwal's *Combinatorial Geometry*

Extended abstract for first-semester students / beginners

This presentation introduces several central ideas from discrete and combinatorial geometry. The main question is how to control a large family of sets using only a small number of points. This question appears naturally in geometric piercing problems. For example, if we are given many disks in the plane, how many points are needed so that every disk contains at least one chosen point?

The first part of the presentation introduces the language needed to formulate such problems. A set system consists of a ground set X and a family $\mathcal{F}$ of subsets of X. A transversal is a subset of X that intersects every set in $\mathcal{F}$, while a packing is a collection of pairwise disjoint sets from $\mathcal{F}$. The transversal number measures how many points are needed to hit all sets, while the packing number gives a basic lower bound. The presentation also introduces fractional transversals and fractional packings, where instead of choosing points or sets completely, one assigns weights. These fractional versions are connected by linear programming duality.

The central part of the presentation is about VC-dimension. VC-dimension measures how complicated a set family is by asking how many points it can shatter. A set is shattered if the family can cut out every possible subset of it. A key example is the family of intervals on the real line: two points can be shattered, but three cannot. Therefore, intervals on the real line have VC-dimension 2.

The main theorem is the shatter lemma, also called the Sauer–Shelah lemma. It says that if a set system has bounded VC-dimension, then it cannot create too many different patterns on finite samples. Without any restriction, a family could create all 2^n possible subsets of an n-point set. But if the VC-dimension is at most d, then the number of possible patterns is bounded by

$$\sum_{i=0}^{d} \binom{n}{i}.$$

For fixed d, this grows only polynomially in n, rather than exponentially.

Finally, the presentation explains ε-nets. An ε-net is a small sample that hits every large set in the family. In the finite version, a $1/r$-net hits every set containing at least a $1/r$-fraction of all points. The epsilon-net theorem says that, for set systems of bounded VC-dimension, such small representative samples always exist. More precisely, if the VC-dimension is at most d, then there is a $1/r$-net of size at most

$$Cdr \ln r,$$

where C is an absolute constant.

The proof uses random sampling, double sampling, and the shatter lemma. A random sample N is chosen, and the bad event is that N misses some large set. A second random sample M is introduced to show whether a missed set was genuinely large. On the combined sample $A = N \cup M$, such a failure becomes visible as a trace with many points from M and none from N. The shatter lemma then controls the number of possible traces, making the probabilistic argument work.

The presentation ends by connecting these ideas to applications such as public opinion polling, biological sampling, geometric approximation, optimization problems, and learning theory. In all of these settings, the same general idea appears: bounded VC-dimension helps us control large and complicated families of sets using small samples or representative structures.

Extended abstract for experts

This presentation surveys the interaction between transversal theory, fractional relaxations, VC-dimension, the Sauer–Shelah lemma, and ε-nets, following Chapter 10 of Matoušek's *Lectures on Discrete Geometry* and Chapter 15 of Pach and Agarwal's *Combinatorial Geometry*. The talk begins with Gallai-type piercing problems for geometric families, using them as motivation for the abstract study of transversals in set systems and hypergraphs.

For a set system $(X, \mathcal{F})$, a transversal $T \subseteq X$ intersects every $S \in \mathcal{F}$, and the transversal number $\tau(\mathcal{F})$ is compared with the packing number $\nu(\mathcal{F})$. The basic inequality

$$\nu(\mathcal{F}) \leq \tau(\mathcal{F})$$

is immediate, while examples such as lines in general position show that no comparable reverse bound holds in general.

The presentation then introduces fractional transversals and fractional packings. These are formulated as dual linear programs, giving the equality

$$\tau^*(\mathcal{F}) = \nu^*(\mathcal{F})$$

for finite set systems. This motivates Lovász's logarithmic rounding theorem for hypergraphs of maximum degree D:

$$\tau^*(H) \leq \tau(H) \leq (1 + \ln D)\tau^*(H).$$

The proof is presented through the greedy algorithm and a charging argument, where the harmonic bound

$$1 + \frac{1}{2} + \cdots + \frac{1}{D} \leq 1 + \ln D$$

controls the total charge attributable to each vertex.

The central combinatorial object is VC-dimension. A subset $A \subseteq V(H)$ is shattered if every $B \subseteq A$ appears as a trace $E \cap A$ for some hyperedge $E \in E(H)$. The VC-dimension is the maximum size of a shattered subset. The talk illustrates this with intervals on the real line, whose VC-dimension is 2.

The main theorem is the Sauer–Shelah lemma. If H has n vertices and VC-dimension at most d, then

$$|E(H)| \leq \sum_{i=0}^{d} \binom{n}{i}.$$

The proof given in the presentation is the standard deletion-induction argument. One fixes a vertex v, considers traces after deleting v, separates the resulting traces into the auxiliary hypergraphs H_1 and H_2, observes that

$$\text{VC-dim}(H_1) \leq d \quad \text{and} \quad \text{VC-dim}(H_2) \leq d - 1,$$

and applies induction together with Pascal's identity. The decomposition

$$|E(H)| = |E(H_1)| + |E(H_2)|$$

yields the recurrence corresponding to

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

The final part applies this combinatorial bound to ε-nets. In the finite formulation, for a set system $(X, \mathcal{F})$ with VC-dim$(\mathcal{F}) \leq d$ and $r \geq 2$, there exists a $1/r$-net $N \subseteq X$ of size at most

$$Cdr \ln r,$$

meaning that N intersects every $S \in \mathcal{F}$ satisfying

$$|S| \geq \frac{|X|}{r}.$$

The probabilistic proof chooses a random sample N of size $s = Cdr \ln r$, studies the bad event that N misses a large set, and introduces a second sample M to witness largeness. On $A = N \cup M$, a bad set produces a trace containing many points of M and none of N.

The shatter lemma bounds the number of possible traces on A by

$$|\mathcal{F}|_A| \leq \sum_{i=0}^{d} \binom{2s}{i},$$

while each fixed bad trace has exponentially small probability, roughly

$$2^{-s/(2r)}.$$

For sufficiently large C, the exponential term dominates the polynomial trace bound, proving that a $1/r$-net exists.

The presentation concludes by emphasizing applications to representative sampling, geometric approximation, set cover and hitting set problems, and statistical learning theory. Conceptually, the common theme is that bounded VC-dimension limits the number of traces a set system can realize, and this combinatorial control makes small hitting samples and related approximation results possible.