Lesson 9. Epsilon-nets and VC-dimension

M. Pellegrini*

February 17, 2004

1 Disclaimer

These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.

2 Introduction

A nice description of the existence of small epsilon nets is in [Mat02, Chapter 10]. This is applied to set systems with bounded VC-dimension. This is abstract enough but a bit more comprehensible than the configuration spaces with empty triggers used in Mulmuley [Mul93]. Also it is shown how to derive the VC dimension of several geometric range spaces. Applications are to other combinatorial problems (fractional covering numbers etc..) not to algorithmic problems. Epsilon approximations are mentioned but treatment is referred to another book by J. Matoušek *Geometric Discrepancy: a guided tour* Springer Verlag, 1999.

The original paper by Haussler and Welzl [HW87] has an application to simplex range searching in linear space and sub-linear time that by now has been improved by Matoušek's Partition trees. It is still important for historical reasons.

Also many results are proved in Chazelle' s book [Cha00, Chapter 4]. See also Chapter 5 on simplex range searching and point location in a set of hyperplanes via optimal cuttings.

Epsilon-nets seem not to be mentioned in [MR95] (the reason may be that they have not many uses beyond geometry).

Epsilon-approximations instead seem to be more useful, they are used for example in closest pair and diameter algorithms in [FP02] and [Kle97].

*Istituto di Informatica e Telematica del CNR, Via G. Moruzzi 1, 56100-Pisa (Italy). pellegrini@iit.cnr.it

Mulmuley [Mul93] proves the result on the worst case size of conflict lists for the configurations induced by a subset but with the weaker assumption of bounded dimension (instead of bounded degree and valence). Epsilon-nets are a special case of Reduced Configurations spaces with empty triggers, that is the configurations coincide with their conflict list, that is it behaves like a set system (or better a measurable set system).

3 Two proofs

We have a proof in [Mul93, pages 217-219] which is for Configuration spaces of bounded dimension that is cast in the language of configuration spaces. It may be a bit confusing since it uses configuration spaces and the reduced configuration spaces in different parts of the proof. It does not use triggers because the trigger set might be empty (although this fact is not highlighted).

A second proof in [Mat02, Chapter 10] is for epsilon-nets for set systems of bounded VC-dimension.

4 Matoušek's proof

Definition 1 (Epsilon-net) *Let X be a set, let μ be a probability measure defined on X, let F be a collection of μ-measurable subsets of X, and ϵ a real number $0 < \epsilon \leq 1$. A subset $Y \subset X$ is called an ϵ-net for (X, F) if for all $S \in F$, $\mu(S) \geq \epsilon$ implies $Y \cap S \neq \emptyset$.*

Definition 2 (VC-dimension) *Let F be a set system on X. A subset $A \subseteq X$ is shattered by F if $\{(A \cap S) | S \in F\} = 2^A$. We define as the VC-dimension of F the supremum of the sizes of all finite subsets of X shattered by F.*

VC stands for the initials of Vapnik and Chervonenkis who defined this notion in the early '70s.

Theorem 1 (Epsilon-net theorem) *Let (X, F, μ) be a μ-measurable set system such that the VC dimension of F is bounded by d, for $d > 2$. Let $r \geq 2$ be an integer parameter, then there exists an $(1/r)$-net for (X, F, μ) of size $O(dr \log r)$.*

Proof. . Proof of theorem (1). Let $s = Cdr \log r$ and Y a subset of X of size s obtained by s random draws with replacement according to the distribution μ. The goal is to show that Y is a $(1/r)$-net with positive probability.

Since small sets in F play no role we can assume that for every S in F, $\mu(S) > \frac{1}{r}$. The probability that Y misses a fixed S in F is equal to $(1 - \frac{1}{r})^s$ which is at most $e^{-s/r}$.

If we could choose $s = r\ln(|F|+1)$ we get a probability at most $1/(|F|+1)$. Thus the probability of having at lest one missed set is at most $|F|/(|F|+1)$ which is less than 1, and we are done.

However since r might not be polynomially related to $|F|$ to get the stated bound promised by the theorem we need a more sophisticated argument. Let:

$$E_0 = \exists S \in F : |Y \cap S| = 0$$

Since E_0 is the event that Y **is not a $1/r$-net** we want to show that $Prob[E_0]$ is smaller than 1. It suffices to show that an upper bound to $Prob[E_0]$ is less than 1.

Take a second sample M of size s obtained by s random draws with replacement according to the distribution μ.

Set $k = s/2r$, consider the event

$$E_1 = \exists S \in F : |Y \cap S| = 0 \wedge |M \cap S| \geq k$$

Clearly since E_1 is more restrictive than E_0, $Prob[E_1] \leq Prob[E_0]$. Next we show that $Prob[E_1] \geq \frac{1}{2}Prob[E_0]$.

Consider a fixed s-subset Y' of X and the probabilities of events conditioned to $Y = Y'$. Now $Prob[E_1|Y = Y'] = Prob[E_0|Y = Y'] = 0$ when Y' is a $1/r$-net.

When Y' is not an $1/r$-net, consider one of the sets S' in F for which $|Y' \cap S'| = 0$. Then

$$Prob[E_1|Y = Y'] \geq Prob[|M \cap S'| \geq k]$$

this is because $|M \cap S'| \geq k$ implies E_1.

Lemma 1 *It holds:* $Prob[|M \cap S'| \geq k] \geq \frac{1}{2}$.

Therefore,

$$Prob[E_1|Y = Y'] \geq \frac{1}{2}$$

On the other hand by definition:

$$Prob[E_0|Y = Y'] = 1$$

Therefore both when Y' is or is not a $(1/r)$-net it holds:

$$Prob[E_1|Y = Y'] \geq \frac{Prob[E_0|Y = Y']}{2}$$

Thus since Y' being or not being a $(1/r)$-net is a paprtition of the space of events, the relation holds unconditionally:

$$Prob[E_1] \geq \frac{Prob[E_0]}{2}$$

In the second part of the proof we are going to give a (small) upper bound for E_1 which by the above reasoning implies a(small) upper bound on E_0 which is our objective.

Build Random sets Y and M with a different process which results in the same distribution. Pick a subset $A \subset X$, of size $2s$ by s random draws with replacement according to the distribution μ. Apply a random permutations to the elements in A and take the fist s elements as Y and the complement as M.

Next we show that for a fixed value $A = A'$ the probability of E_1 conditioned to $A = A'$ is rather small.

Consider a fixed $S' \in F$.

$$Prob[E_1(S', A')] = Prob[|Y \cap S'| = 0 \wedge |M \cap S'| \geq k|A = A']$$

Now if $|A' \cap S'| < k$ then $Prob[E_1(S', A')] = 0$. Otherwise assume $|A' \cap S'| \geq k$.

$$Prob[E_1(S', A')] \leq Prob[|Y \cap S'| = 0|A = A']$$

This is because of the "law of implication". So now we bound the later. Since Y is a random sample of A' and has to avoid the at least k elements in $S' \cap A'$ we have:

$$Prob[|Y \cap S'| = 0|A = A'] = \frac{\binom{2s-|A' \cap S'|}{s}}{\binom{2s}{s}} \leq \frac{\binom{2s-k}{s}}{\binom{2s}{s}} \leq (1 - \frac{k}{2s})^s \leq e^{-k/2} = r^{-Cd/4}$$

Now we use the Bounded VC-dimension and the following lemma:

Lemma 2 *For a set system (X, F, μ) of VC dimension at most d we have that for all m:*

$$max_{Y \subseteq X, |Y|=m} |\{S \cap Y : S \in F\}| = \Phi_d(m) = \sum_{i=0}^{d} \binom{m}{i}$$

The left hand side as a function of m is called the primal shatter function.

We can approximate $\Phi_d(m) \leq \left(\frac{em}{d}\right)^d$.

Coming back to the proof we have a bound on the probability that for a fixed set S' and A' E_1 holds, Using the union bound and the fact that the number of distinct sets $S \cap A'$ is bounded by $\Phi_d(2s)$ we get;

$$Prob[E_1|A = A'] \leq \Phi_d(2s) r^{-Cd/4} \leq \left(\frac{e2s}{d}\right)^d r^{-Cd/4} < \frac{1}{2}$$

For a suitable constant C. Note that the expression of the bound does not depend on A'.

Now we use the fact that for disjoint events $A = A'$:

$$Prob[E_1] = \sum_{A' \subset X, |A'| = 2s} Prob(E_1|A = A') Prob(A = A') < \frac{1}{2} \sum_{A' \in F} Prob(A = A') = \frac{1}{2}$$

So finally we have $Prob[E_0] \leq 2Prob[E_1] < 1$.

5 Proof of two lemmas

5.1 Proof of lemma 1

To recap. We have a fixed set S' (chosen in dependence of the fixed set Y'). We have a random set M of size s. We set the parameter $k = s/2r$. The aim it to prove that $Prob[|M \cap S'| \geq k] \geq \frac{1}{2}$, under the assumption $s/r = 2k > 8$, that is $k > 4$.

We can assume that $\mu(S') = 1/r$ since a set with larger measure increases the probability of a large intersection with a randomly chosen subset. Thus $E[|M \cap S'|] = s/r = 2k$.

Let X_i be 1 if the i-th element of M is in S', and 0 otherwise, clearly $X = \sum_i X_i$ is the size of $M \cap S'$. And $E[X] = \sum_i E[X_i] = s(1/r) = s/r = 2k$. We apply now Chebyshev's inequality for which we need a bound on the variance: $Var(X) = \sum_i Var(X_i) = \sum_i (1/r - 1/r^2) < s/r = 2k$. By Chebyshev:

$$Prob[|X - E[E]| > t] \leq Var[X]/t^2$$

Clearly:

$$Prob[X < (1/2)E[X]] \leq Prob[|X - E[E]| > (1/2)E[X]]$$

Setting $t = (1/2)E[X] = k$ we get:

$$Prob[X < k] \leq (2/k) \leq 1/2$$

Thus the opposite event holds with probability greater than 1/2:

$$Prob[X > k] \geq 1/2$$

(end of proof).

5.2 Proof of lemma 2

This is called the *shatter function lemma* and relates the VC dimension of a set system with the maximum number of distinct intersections of a set with that set system.

Consider the function $\Phi_d(m)$, for d fixed and $m \to \infty$, it can be approximate by $O(m^d)$. A more precise estimate is $\Phi_d(m) \leq (em/d)^d$ which is valid for every value of m and d.

The proof is an induction on d and n.

Consider a set system (X, F) with $X = n$ and $VCdim(F) = d$.

Fix an element $x \in X$. Consider the set system F_1 obtained by eliminating x from any set in F, that is $F_1 = \{S \setminus \{x\} : S \in F\}$. Two sets in F give a single set in F_1 if and only if they differ by x. So we consider also the following set system $F_2 = \{S \in F : x \notin S, \{x\} \cup S \in F\}$.

Clearly $|F| = |F_1| + |F_2|$.

1. F_1 is defined on a ground set of $n-1$ elements and its VC-dimension cannot increase with respect to that of F, so $VCdim(F_1) \leq d$.

2. However F_2 has a smaller VC dimension $VCdim(F_2) \leq d-1$. This is because, if $A \subset X \setminus \{x\}$ is shattered by F_2, then $A \cup \{x\}$ must be shattered by F.

$$\pi_F(n) \leq \pi_{F_1}(n-1) + \pi_{F_2}(n-1)$$

By induction we can bound the righthand side with

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

which by the properties of binomial coefficients is equal to $\Phi_d(n)$.

We can group together the terms of the two summations as:

$$\binom{n-1}{0} + [\binom{n-1}{1} + \binom{n-1}{0}] + \ldots [\binom{n-1}{d} + \binom{n-1}{d-1}]$$

Since in general $\binom{a-1}{b} + \binom{a-1}{b-1} = \binom{a}{b}$ we get exactly $\Phi_d(n)$.

(end of proof).

6 Observation on composition of spaces with bounded VC dimension

The notion of VC dimension is important and useful because

1. In many concrete situations we can estimate the VC-dimension of a set system by rather simple arguments.
2. There are rules of composition for spaces of bounded VC dimension
3. Results (existence of epsilon-nets or epsilon-approximations) can be proved for spaces of bounded VC dimension.

From [Mat02, Chapter 10]:

Theorem 2 *Let X be a set of points in R^d, the VC-dimension of the set system induced by the class of closed halfspaces is $d+1$.*

$d+1$ distinct points in R^d for a (possibly degenerate) simplex and every subset appears as a face of the simplex. On the other hand Radon's lemma, $d+2$ points cannot be shattered, i.e.there are bi-partitions of the set of points such that the two convex hulls intersect.

Theorem 3 *Let X be a set of points in R^d, the VC-dimension of the set system induced by the class of real polynomials on d variables of degree at most D is at most $\binom{d+D}{d}$.*

Theorem 4 *Let (X, S) be a set system with $VCdim(S) = d$. Let F be a set theoretic formula using the operators union, intersection and difference, with k variables taking values in S. Then the induced set system has VC dimension $O(kd \log k)$.*

7 Epsilon-approximations

Definition 3 (Epsilon-net) *Let X be a set, let μ be a probability measure defined on X, let F be a collection of μ-measurable subsets of X, and ϵ a real number $0 < \epsilon \leq 1$. A subset $Y \subset X$ is called an ϵ-approximation for (X, F) if for all $S \in F$:*

$$\left| \mu(S) - \frac{|S \cap Y|}{|Y|} \right| \leq \epsilon.$$

Theorem 5 (Epsilon-approximation theorem) *Let (X, F, μ) be a μ-measurable set system such that the VC dimension of F is bounded by d, for $d > 2$. Let $r \geq 2$ be an integer parameter, then there exists an $(1/r)$-approximation for (X, F, μ) of size $O(dr^2 \log r)$.*

Epsilon-approximation intersect each set in the set system in a number of points proportional (up to a small error) to the measure of that set.

References

[Cha00] B. Chazelle. *The Discrepancy method: randomness and complexity*. Cambridge University press, Cambridge, 2000.

[FP02] D. Finocchiaro and M. Pellegrini. On computing the diameter of a point set in high dimensional euclidean space. *Theoretical Computer Science*, 287(2):501–514, 2002.

[HW87] David Haussler and Emo Welzl. Epsilon-nets and simplex range queries. *Discrete and Computational Geometry*, 2:127–151, 1987.

[Kle97] Jon M. Kleinberg. Two algorithms for nearest-neighbor search in high dimensions. In *Proceedings ACM STOC*, pages 599–608, 1997.

[Mat02] J. Matoušek. *Lectures on Discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer erlag, New York, 2002.

[MR95] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

[Mul93] K. Mulmuley. *Computational Geometry, an Introduction through Randomized Algorithms*. Prentice Hall, 1993.